SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-52

Detecting Structural Breaks using Hidden Markov Models

Christos Ntantamis

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Detecting Structural Breaks using Hidden Markov Models

Christos Ntantamis *

CREATES, Aarhus University

August 31, 2010

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 its performance is superior in the assessing the number of breaks and their respective locations. Finally, the methodology is applied for the detection of breaks in the monetary policy of United States, the different functional form being variants of the Taylor (1993) rule.

JEL Classification Numbers: C13, C22, C52

Keywords: Structural change, Hidden Markov Model, Regime Switching, Bayesian Segmentation, Monetary Policy.

*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, and to participants in the Italian Congress of Econometrics and Empirical Economics 2009, the Canadian Economics Association Annual Meeting 2009, and the International Conference on Computational and Financial Econometrics 2009, for their suggestions. 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. Ift 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) in the engineering literature (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 involve a structural break, thus there is no reason to assume that the process will reverse to its original specification, the HMM approach, despite being a misspecified model, it still can identify the location of this break as the location of 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 conditionally 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 have been generated according to different distributions across different 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. Since the parameters of Hidden Markov Models are not identified, unless restrictions are put in the parameters, 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 in the literature 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 as it takes into account the classification of the data points to the different regimes in order to penalize the lack of fit, and it can be easily computed; it adds an extra term to the Bayesian Information Criterion (BIC) that is readily available when 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. Regardless, the Bai&Perron (1998) method is still used for comparison purposes to demonstrate the effects of using this method to identify structural breaks, when the change occurs in the model and not in the parameters.

The empirical application presented in this paper involves the identification of different regimes in the monetary policy in United States. It has been documented that the monetary policy of the United States has changed over time (Qin and Enders 2008). Thus, it is of interest to examine whether such changes are also related to changes in the functional form of the monetary policy rules. The candidate functional forms to be used in the IMI methodology will be assumed to be variants of the Taylor (1993) rule, which assumes that the short-term nominal interest rate reacts linearly to deviations of inflation and output from their target levels in order to establish the equilibrium level of the real interest rate. The results suggest that there has been a change in the way monetary policy has been conducted, with a single break occurring in the middle 80s.

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. The application of the IMI methodology in the monetary policy of USA is presented in Section 4, and Section 5 concludes the paper.

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

[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 $X_{\neg t} \triangleq X_{1:T} \setminus X_t = \{X_1, X_2, \ldots, X_{t-1}, X_{t+1}, X_{t+2}, \ldots, X_T\}$.

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):

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

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

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

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

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

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 π which is a vector of length equal to the number of the different states. Then, $\pi_i = P(X_1 = i)$, where π_i is the i^{th} element of π. 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 (α and β 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 estimation 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

[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.

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

It is of interest to provide a more detailed description of the ICE-GEMI algorithm for the case that data are conditionally i.i.d., that is (Y_t) is i.i.d. given the value of (X_t). This will also facilitate the exposition of the case of (Y_t) being conditionally dependent on explanatory variables in the next section.

The following assumptions must hold for ICE-GEMI algorithm to ensure convergence:

1. An estimator of the parameter vector ζ 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 θ^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 θ_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: ζ^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 ζ^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 $y_{1:T}$, $y_i^q = (y_s)_{s \in S_i^q}$, 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: $D(y_i^q) \in \left\{ f_{\theta_i^1}, \ldots, f_{\theta_i^M} \right\}$.

(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 ζ^{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].

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) are conditionally i.i.d.

1. The parameter ζ 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 (α) and backward(β) probabilities

[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. The detailed results from the simulations run in this paper, although not reported, support their conjecture.

according to the Baum-Welch algorithm [5]:

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

and

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{l=1}^{n} \alpha_t(l)\beta_t(l)} \tag{12}$$

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 $\gamma_1(i)$.
- For each new element $t = 1, \ldots, T-1$, 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.

Note that 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.

The above discussion argues heuristically why the algorithm converges to the true parameters. The theoretical conditions for the algorithm convergence are stated in Theorem 1 in Pieczynski (2007). Regardless, the necessary hypotheses are easy to verify for relatively simple models, such as estimation of the means of a mixture model, with known mixing probabilities and variances. Trying to verify the hypothesis for the case of a Hidden Markov Model becomes extremely difficult due to the uncertainty generated by simulating at each iteration the hidden path. On the other hand, Delmas (1997) can provide a shortcut: if the exponential family is assumed, then the ICE algorithm and the Expectation Maximization (EM) algorithm are

[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.

equivalent. Since results for the converge of the EM algorithm are available (Wu 1983), one could apply the ICE algorithm using the exponential family, e.g. assuming a normal distribution for the observations, and thus ensure, theoretically, that parameter estimates will converge to the true values.

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. That is, instead of applying the ICE-GEMI method for the i.i.d case to $((X_t, Y_t))$, it will be applied for the pair of $((X_t, \epsilon_t))$[6]. 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:

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

[6] The innovations ϵ_t are now playing the role of the observed variables.

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

The iterative procedure can be depicted in the following figure:

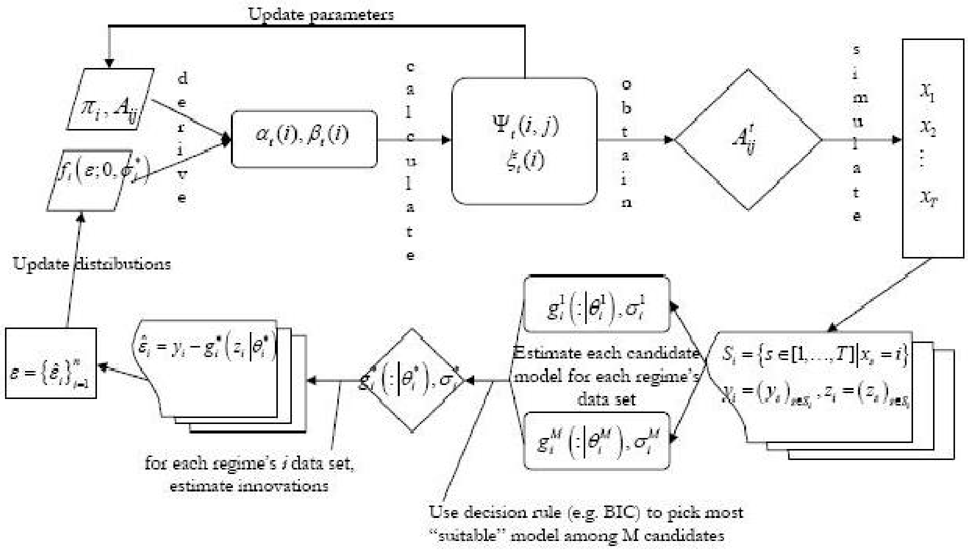

Figure 1: ICE-GEMI algorithm

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: 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, in a normal mixture model used to detect the presence of grouping in some data, spurious components may be detected in an attempt to model a skewed distribution for one of the groups (Bilmes 2006).

What about the convergence in this case? As it was discussed before, a shortcut is provided by assuming that the observations' distributions belong to the exponential family. This implies that the innovations should be conditionally i.i.d. normal. This assumption will be maintained for the implementation of the ICE-GEMI algorithm in this paper.

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 employment of nonparametric model validation, and in particular density based specification tests (Aït-Sahalia, Fan and Peng 2009, Aït-Sahalia, Fan and Jiang 2009). Even though this approach is appealing, it adds computational burden to the model. 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, which 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 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.

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

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

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

not the case in the economic literature. Unless there are theoretical grounds (which is not straightforward), 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 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.

Another 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).

In this paper, a version of a penalized log likelihood criterion (ICL) will be preferred because of its lower computationally 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.

2.4.1 The Integrated Classification Likelihood Criterion

In this section, a more detailed discussion of the ICL criterion will be provided. To do so, the notion of complete-data likelihood $L_c(\Psi)$ needs to be introduced. This is clearer within the Expectation Maximization (EM) framework (Dempster et al. 1977); 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:

$$\log L(\Psi) = \log L_c(\Psi) - \log k(\Psi) \qquad (18)$$

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) followed 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 (π), with all its parameters set equal to ν, 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

3.1.1 Large sample sizes

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:

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

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, emphasis needs to be put into the difference between identifying regimes and detecting breaks. As it was discussed in previous sections, the ICE method is

[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.

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.

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 should choose the optimal number of regimes through the ICL-BIC criterion, and b) it should identify 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 comparison of results obtained by the two approaches is presented in Tables 1,2, and 3 whereas the detailed results for the ICL-BIC criterion are presented in Table 4. 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 1). 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 2). 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 1). The ICL-BIC criterion presents its minimum average value when the ICE algorithm is performed for the correct number of regimes (Table 4). 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 supportive for the IMI methodology.

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

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 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 2). 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, Table 3 report those for the case of 3 Regimes and 2 Breaks[12]. These estimates 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.

3.1.2 Small sample sizes

The sample sizes considered in the comparison discussed before are more appropriate if a financial market application is in mind where daily (or weekly) data are available. Regardless, many interesting economic models are relevant to macro data that are available in the best case at a monthly frequency; monetary policy rules that determine interest rates on a monthly level are such an example. In order to assess the ability of the IMI methodology of detecting break in such a scenario, a subset of simulations were considered with more appropriate sample sizes. The Data Generating Processes from the previous section were used, i.e. linear models with changing parameters across regimes. In particular,the following cases were examined:

1. DGP 3, 3 regimes with 2 breaks: The sample size was set equal to 600, with breaks occuring at points 200 and 400.

2. DGP 1, 2 regimes with 1 break: The sample size was set equal to 300, with the breaks occurring at middle-point (150).

3. DGP 1, 2 regimes with 1 break: The sample size was set equal to 400, with breaks occurring at a quarter-of-the-sample distance from the end-points, i.e. breaks occurring either at point 100 or at

[11] The "correct" stands for a regime switching location estimated to be in the range of ± 10 time intervals from the true value. Detailed results are available from the author upon request.

[12] Estimation results are similar in the other cases as well.

point 300.

The results are tabulated in Tables 5 to 7. The main focus is on the performance of the IMI methodology in terms of estimating the correct number of regimes, and if so, the ability to correctly identify the location of the break points. Table 5 presents the descriptive statistics of the values of the ICL-BIC criterion for the cases considered. The correct number of regimes presents the lower value for either the mean or the median value of the information criterion. Examining how many times the ICL-BIC picked the correct number of regimes (Table 6), it can be noticed that the performance deteriorates with the smaller sample size. For the case of 3 regimes with 2 breaks, the correct number of regimes is identified for 86% of the total number of simulation compared to 93% when the sample size was larger. The same result holds for all the cases of 2 regimes with 1 break: the corrent number of regimes (2) is identified for about 80+% of the simulations compared to 100% for the larger sample size. Nevertheless, the results remain comparable to the results obtained using the Bai and Perron (1998) method for the larger sample sizes. If the attention is turned to the actual location of the identified break points, that is the location of the regime switches, (see Table 7) it can be noticed that 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 the switches falls in the range of ±5 time intervals from the true value at 73.3 to 91.4 % of the replications, still comparable to the results obtained for the Bai and Perron method for the larger sample size [13]. Overall, it can be infered that the IMI methodology compared adequately even for the smaller sample sizes.

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, Linear Model vs Nonlinear Model

$$y_t = \begin{cases} 1 + 0.5x_{1t} - 0.4x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.2^2\right), \quad 0 \le t \le T/2 \\ 1 + 0.5x_{1t} + 0.2x_{2t} - 0.3x_{1t}^2 + 0.1x_{2t}^2 - 0.5x_{1t}x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.2^2\right), \quad T/2 + 1 \le t \le T \end{cases} \tag{28}$$

2. Case II: Models 1,3 Linear Model vs Autoregressive of Order 1 (AR(1)) Model

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.5x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.2^2\right), \quad 0 \le t \le T/2 \\ 0.2 + 0.8y_{t-1} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad T/2 + 1 \le t \le T \end{cases} \tag{29}$$

3. Case III: Models 1,4 Linear Model vs Semi-Logarithmic Model

$$y_t = \begin{cases} 1 + 1x_{1t} + 0.2x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.2^2\right), \quad 0 \le t \le T/2 \\ exp\left(1 + 0.5x_{1t} - 0.2x_{2t} + \varepsilon_t\right) & \varepsilon_t \sim N\left(0, 0.2^2\right), \quad T/2 + 1 \le t \le T \end{cases} \tag{30}$$

[13] The discussion regarding the quality of the parameter estimates is omitted.As it was noted before, if the correct number of regimes and the correct location of the regime switching point are obtained, then the parameter estimates will also be very good.

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 four 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 non-linear regression model that includes the squares and the cross-product of the two variables X_1, X_2; 3) an Autoregressive Model of order 1 (AR(1)); and 4) a semi-logarithmic model. These are also the models that are considered in the IMI algorithm[14]. Two sample sizes will be considered $T = \{300, 1000\}$ with the break occurring in the middle of the sample.

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. Even though the Bai and Perron (1998) methodology should not be used in this setting, as the simulation design is such that there exist different models and not just different parameters across the regimes, making any comparison by default unfair, it can still reveal interesting results. If the researcher is not suspecting the existence of different models and still uses this standard method, how frequently she will be wrong?

The simulation results are presented in Tables 8 and 9. In summary, the following hold for the IMI methodology in terms of the identification of the number of regimes (Table 8): a) the best performance occurs when a switch from a linear to nonlinear model occurs, and the worst when a switch from a linear to a semi-logarithmic model takes place, b) it tends to overfit, i.e. it identifies a larger number than the true one, compared to the underfitting observed when the change is only at the parameters of a linear model, c) the results improve as the sample size increases. Overall, the IMI methodology's worst result is for Case II, with a small sample size, when the correct number of regimes is identified at only 62% of the replications, whereas the best result is for Case I, with a large sample size, when the correct number of regimes is identified at 99 % of the replications. Even when the ICL-BIC criterion overestimates the number of the regimes, the IMI method can still obtain the correct number of regime switches and their corresponding locations. For example, in Case III, when 3 regimes are picked, instead of 2, which occurs at 30% of the replications, the reconstructed, by the MPM approach, hidden state sequence identifies the correct 2 regimes and the correct switch location at approximately 70% of the cases. Even though these cases are not considered to satisfy the evaluation criteria as stated before, this result suggests that extra attention should be put to the reconstructed hidden sequence.

The Bai and Perron (1998) methodology is not performing well, as expected. This is more evident by the fact that the results are not improving much with the increase of the sample size. Overall, the method tends to point to a larger number of breaks, compared to the true number, and it fails dramatically for Case III, i.e. a switch from the linear model to an AR(1) one (when the sample size is large, only at 16.5% of the replications the true break number is correclty picked). Thus, a research should be cautious in using this methodology, when there may exist reasons to point to different model across the different segments.

[14]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.

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 always fall in the range of ± 5 time intervals from the true value (100% of these cases). Finally, in terms of the parameter estimates, not reported in the paper, they are quite accurate in the cases that the IMI methodology picks the correct number of regimes. This is expected as when the number of regimes and the location of regime switches are accurately estimated, the candidate models are consistently estimated.

Overall, the IMI method performs adequately in estimating the correct number of regimes and the respective locations of breaks.

4 Empirical Application: Monetary Policy Rules in United States

4.1 Candidate models based on Taylor Rule

The appeal of Taylors (1993) work was that he proposed a simple rule for conducting monetary policy in United States. The short-term nominal interest rate reacts linearly to deviations of inflation and output from their target levels in order to establish the equilibrium level of the real interest rate. The original specification of the rule took the form:

$$i_t^* = r^* + \pi_t + \beta_\pi (\pi_t - \pi^*) + \beta_y y_t \tag{31}$$

where, i_t^* is the target short-term nominal interest rate for quarter t, r^* is the equilibrium real interest rate, π_t is the inflation rate, π^* is the inflation target, and y_t is the output gap measured by the deviation of real output from its potential level. The original specification set r^* and π^* equal to 2, and the parameters β_π and β_y equal to one half.

Over the past years, a large amount of literature has emerged in discussing different variants of the rule, and the type of data that need to be used for its estimation. Regarding the different variants of the Taylor rule, we can consider two major variations. The first one considers that the Federal Reserve smooths the interest rate to help avoid policy reversals that could be damaging in its credibility. Clarida et al., (1998) and Clarida et al., (2000) generalized the Taylor rule by including lagged values of the federal fund rates. The second variant, advocated by the work of Orphanides (2003) showed that a forward-looking rule, that is the inflation and output gap are replaced in the Taylor rule by their forecasts, performs better than a backward-looking rule. Regarding the data used in the estimation, a number of recent policy evaluation studies suggested that information that was actually available to the policy makers when they were making their interest-setting decisions should be used (Orphanides 2001, 2002,and 2003). These real-time data are not subject to the subsequent revisions in the values of inflation and output present in *ex post* data.

More recent theoretical and empirical work has consider the policy reaction function to be nonlinear. Orphanides and Wilcox (2002) search for a specification of the policymakers loss function that would lead to an asymmetric response to disinflation: when inflation is moderately above or below its long run objective, the Federal Reserve should not take any action by changing the Fed Funds rate. On the other hand, should inflation steps outside the inaction band, the Fed should adjust the Fed Funds rate aggressively. Qin and Enders (2008) compare the in-sample and the out-of-sample properties of a number of linear and nonlinear Taylor rules to select the most appropriate form.

In this paper, the nonlinearity is assumed to represent itself in 2 different ways. First, the monetary policy rule is allowed to change over time and not over different values of the inflation. Second, the IMI methodology allows for different, potential nonlinear forms, of the rule across the different regimes in the time domain.

Following Qin and Enders (2008), the candidate models for the application of IMI are the following:

1. Model 1: Backward Looking Taylor Rule

$$i_t = \alpha_0 + \alpha_1 \pi_{t-1} + \alpha_2 y_{t-1} + \epsilon_t \tag{32}$$

2. Model 2: Backward Looking non-linear Taylor Rule

$$i_t = \alpha_0 + \alpha_1 \pi_{t-1} + \alpha_2 y_{t-1} + +\alpha_3 \pi_{t-1}^2 + \alpha_4 y_{t-1}^2 + \alpha_5 \pi_{t-1} y_{t-1} + \epsilon_t \tag{33}$$

3. Model 3: Backward Looking Taylor Rule, interest rate smoothing

$$i_t = \alpha_0 + \rho i_{t-1} + \alpha_1 \pi_{t-1} + \alpha_2 y_{t-1} + \epsilon_t \tag{34}$$

Model 4: Forward Looking Taylor Rule, within quarter-rule

$$i_t = \alpha_0 + \rho i_{t-1} + \alpha_1 E_t \pi_t + \alpha_2 E_t y_t + \alpha_3 y_{t-1} + \epsilon_t \tag{35}$$

Model 5: Forward Looking Taylor Rule

$$i_t = \alpha_0 + \rho i_{t-1} + \alpha_1 E_t \pi_{t+1} + \alpha_2 E_t y_t + \alpha_3 y_{t-1} + \epsilon_t \tag{36}$$

Model 6: AR(2) model

$$i_t = \alpha_0 + \rho_1 i_{t-1} + \rho_2 i_{t-2} + \epsilon_t \tag{37}$$

Model 7: ARIMA(1,1,0) model

$$\Delta i_t = \alpha_0 + \rho_1 \Delta i_{t-1} + \epsilon_t \tag{38}$$

Some discussion is required for the aforementioned models. The first three models are variants of the backward looking Taylor rule, i.e. the policy reacts to the previous quarters inflation and output gap. Model 1 is the classic backward-looking Taylor rule, Model 2 considers the backward-looking Taylor rule taking a nonlinear form with respect to the output gap and the inflation, and Model 3 assumes interest rate smoothing behavior on behalf of the Federal Reserve. In Model 4, the policy reacts to within-quarter forecasts of both inflation and the output gap, that is to forecasts for the current values of inflation and output gap at time t available with the quarter t. Model 5 suggests that the policy reacts to 1-quarter-ahead forecasts of the inflation and the within-quarter forecasts of the output gap (Orphanides 2004) uses this specification as the forward-looking rule. Finally, Models 6 and 7 are two time series models: Model 6 is a univariate AR(2) model, and Model 7 is an ARIMA(1,1,0), the latter being support by the fact that interest rates have been found to act as unit-root processes.

The IMI methodology is applied to the following four cases:

1. Case I: Models 1-7
2. Case II: Models 1-5
3. Case III: Backward Looking Taylor Rule (Model 1)
4. Case IV: Forward Looking Taylor Rule (Model 5)

In Cases III and IV, the same function form is assumed across all regimes and only parameters are allowed to change. While Case I considers all models as potential candidates for explaining the monetary policy within each regime, in Case II the Time Series models are excluded so that to focus only on the variants of the Taylor rule.

4.2 The data

Orphanides (2000, 2003) has pointed out that estimation of the Taylor rule should be performed by using real-time data, as the inflation and output in *ex post* data are subjected to constant revisions. Real-time data are constructed primarily based on the forecasts and information from the Federal Reserve Greenbook, a document prepared by Federal Reserve Board staff for the Federal Open Market Committee before every regularly scheduled meeting, and they include projections of real output growth and inflation. Such data series are now readily available in the website of the Federal Reserve Bank of Philadelphia, with a time lag of 5 years. Currently, the data set includes the projections from November 1965 to November 2004 for real output and inflation, and values for the output gap from November 1987 to November 1984.

The data used in this paper span the period 1967:Q1 till 2004:Q4. The data contain the real-time lags and the real-time forecasts of inflation rate and the output gap. The inflation rate is defined as the difference of the logarithm of the GDP deflator, and the output gap is defined as the difference of the logarithm of actual GDP and the logarithm of the potential GDP. The inflation forecasts are available for up to five quarters ahead in most cases for the entire sample, while only within-quarter forecasts are available for the output gap, and that are not available till 1987. Thus, the values of the output gap (lagged y_{t-1}, and within-quarter forecasts $E_t y_t$) had to be reconstructed from the historical real output data available in the Philadelphia Fed website. Using the Orphanides and van Norden (2002) approach, the real-time measure of output gap was constructed by applying the Hodrick-Prescott filter in successive vintages of the data, providing the lagged values of the output gap available at time t, y_{t-1} , as they were available at time t. In order to obtain the within-quarter estimate of the output gap $E_t y_t$, an extra step was required [15]. For each quarter t, the real output vintage series was augmented with the projection for the value of the real output at quarter t, as it was available at time t. Using this updated series, the Orphanides and van Norden (2002) approach was used once more, and a series of within-quarter forecasts for the output gap was constructed. Finally, following previous literature, the federal funds rate is constructed as a quarterly average.

[15]For each quarter t, the vintage series contains the values of the real output up to time $t-1$.

4.3 Results

The results of the estimation process are tabulated in Tables 10 to 13 [16] They can be summarized as follows:

(I) Case I: The ICL-BIC (Table 10) suggests the existence of two different regimes in the monetary policy rules, with a single break occurring in the second quarter of 1985. During the first period, 1967:Q1 to 1985:Q2, the BIC suggests that the best model to describe the Federal Fund Rate was Model 7, that is the ARIMA(1,1,0) model. During the second period, 1985:Q3 to 2004:Q4, the BIC suggests that the best model is Model 5, i.e. the forward-looking Taylor Rule. The parameter estimates for Model 5 indicate that the monetary policy put more weight in inflation compared to the combine effect of the output gap (Table 12) [17]. One further remark is the vast improvement in terms of BIC values (Table 11) for models that incorporate interest rate smoothing. For example, the BIC of Model 1 is almost double of the BIC value for Model 3. Of course, this can also attributed to the high persistence of the interest rate, as it is also evident by the BIC values for the simple time series models.

(II) Case II: Once more the ICL-BIC suggests the existence of two different regimes, with a single break occurring again in the second quarter of 1985. The BIC picked the same model for both periods, Model 5. The forward-looking Taylor rule describes the monetary policy in both regimes. Regardless, the parameter estimates point a change in the policy. In particular, the rule puts an almost equal weight for the inflation (0.253) and the output gap (an approximate overall value of 0.21) during the first period, whereas during the second period more weight is place on inflation.

(III) Case III: When only the backward-looking Taylor rule (Model 1) is considered, so the interest is in the change of the coefficient values across time and not on the function forms, the ICL-BIC suggests the existence of three different regimes. The transitions between these regimes are rather frequent, especially before 1980, but the following remarks can be made. The first regime, which put an almost 4 times higher weight on inflation (1.195) compared to the output gap (0.332) dominates the US monetary policy during the decade of 1980. This is not a surprising result as it coincides with most of the Volcker period, when the Federal Reserve pursued a more aggressive policy in targeting inflation. The second regime is more prominent during the decade of 90s. Compare to the first regime the difference in the monetary policy weights between inflation and output gap is reduced, to less than double (0.796 and 0.425 for inflation and output gap respectively). Finally, the third regime appears to be more important during the 1970s and after the third quarter of 2001. The difference in the inflation and output gap weights is similar to the one in the case of the second regime. Regardless, the constant term of the Taylor rule is almost half to the one of the second regime.

Overall, when all candidate models are considered, the IMI procedure points to a break in the middle of 1980s, with a forward-looking Taylor rule putting more weight in inflation describing the monetary policy in the second part.

[16] Case IV is not reported as it yielded the exact same results as Case II, which can be fully explained by a closer look into Table 11 that reports the BIC for each case, each regime, and each candidate model.

[17] Given that the values of $E_t y_t$ and y_{t-1} are close, the overall coefficient for the output gap is approximately 0.2 compared to 0.422 for inflation.

5 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. Most importantly, it can simulate 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 method was also examined in smaller sample sizes, more appropriate for macroeconomic application. The IMI procedure performed adequately in this settin as well. 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.

Finally, the proposed methodology has been applied in the identification of different regimes/policies in United States data. Considering candidate functional forms to be variants of the Taylor (1993) rule, the results suggest that there has been a change in the way monetary policy has been conducted, with a single break occurring in the middle 80s. In particular, when all candidate models are considered, the IMI procedure points to a break in the middle of 1980s, with a forward-looking Taylor rule, which puts more weight in the inflation compared to the output gap, describing the monetary policy in the second part.

References

[1] Aït-Sahalia., Y., J. Fan, and J. Jiang, (2009). Nonparametric Tests of the Markov Hypothesis in Continuous-time Models, *The Annals of Statistics*, forthcoming.

[2] Aït-Sahalia., Y., J. Fan, and H. Peng, (2009). Nonparametric Transition-based Tests for Jump Diffusions, *Journal of the American Statistical Association*, 104, 1102-1116.

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

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

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

[6] 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.

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

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

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

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

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

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] Clarida R., J. Gali and M. Gertler, (1998). Monetary Policy Rules in Practice: Some International Evidence, *European Economic Review*, 42, 1033-1067.

[19] Clarida R., J. Gali and M. Gertler, (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *Quarterly Journal of Economics*, 115, 147-180.

[20] 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.

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

[22] 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.

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

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

[25] 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.

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

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

[28] 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.

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

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

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

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

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

[34] Orphanides A., (2000). Activist Stabilization Policy and Inflation: The Taylor rule in the 1970s, *Finance and Economics Discussion Series 2003-13*, Federal Reserve Board, Washington, DC.

[35] Orphanides A., (2001). Monetary Policy Rules based on Real-time Data, *American Economic Review*, 91, 964-985.

[36] Orphanides A., (2002). Monetary Policy Rules and the Great Inflation, *American Economic Review*, 92, 115-120.

[37] Orphanides A., (2003). Historical Monetary Policy Analysis and the Taylor Rule, *Journal of Monetary Economics*, 50, 983-1022.

[38] Orphanides A., (2004). Monetary Policy Rules, Macroeconomic Stability, and Inflation: A view from the Trenches, *Journal of Monetary, Credit, and Banking*, 36, 151-175.

[39] Orphanides, A., and S. van Norden, (2002). The Unreliability of Output Gap Estimates in Real Time, *Review of Economics and Statistics*, 84, 569-583.

[40] Orphanides, A. and D. Wilcox, (2002). An Opportunistic Approach to Disinflation, *International Finance*, 5, 47-71.

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

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

[43] 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.

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

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

[46] Pieczynski W., (2007). Convergence of the Iterative Conditional Estimation and Application to Mixture Proportion Identification, *IEEE Statistical Signal Processing Workshop*, Madison, Wisconsin, USA.

[47] Qin, T., and W. Enders, (2008). In-sample and Out-of-sample Properties of Linear and Nonlinear Taylor Rules, *Journal of Macroeconomics*, 30, 428-443.

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

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

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

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

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

[53] Wu, C.F.J., (1983). On the Convergence Properties of the EM Algorithm, *Annals of Statistics*, 11, 95-103.

[54] Taylor, J.B., (1993). Discretion versus Policy Rules in Practice, *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

A Tables

A.1 Simulation Results: Different parameters across regimes

A.1.1 Large Sample Size 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

Parameter Estimates. Case III: 3 Regimes - 2 Breaks									
	IMI			**B&P: Sequential**			**B&P: Repartition**		
Regime 1	**ct**	**X1**	**X2**	**ct**	**X1**	**X2**	**ct**	**X1**	**X2**
True Values	**1**	**0.7**	**-0.25**	**1**	**0.7**	**-0.25**	**1**	**0.7**	**-0.25**
mean	0.999	0.700	-0.250	1.001	0.701	-0.256	1.001	0.701	-0.256
median	1.006	0.703	-0.259	0.996	0.702	-0.248	0.996	0.702	-0.248
stdev	0.083	0.038	0.114	0.088	0.028	0.120	0.088	0.028	0.120
min	0.797	0.434	-0.694	0.822	0.620	-0.547	0.822	0.620	-0.547
max	1.284	0.774	-0.002	1.256	0.765	-0.009	1.256	0.765	-0.009
Regime 2	**ct**	**X1**	**X2**	**ct**	**X1**	**X2**	**ct**	**X1**	**X2**
True Values	**1**	**0.7**	**0.5**	**1**	**0.7**	**0.5**	**1**	**0.7**	**0.5**
mean	1.002	0.702	0.494	1.010	0.700	0.484	1.010	0.700	0.489
median	1.003	0.701	0.495	1.006	0.699	0.495	1.008	0.700	0.497
stdev	0.087	0.028	0.116	0.092	0.026	0.120	0.092	0.026	0.121
min	0.769	0.632	0.158	0.787	0.642	0.172	0.789	0.639	0.172
max	1.245	0.796	0.830	1.269	0.763	0.784	1.271	0.763	0.782
Regime 3	**ct**	**X1**	**X2**	**ct**	**X1**	**X2**	**ct**	**X1**	**X2**
True Values	**1**	**0.2**	**0.5**	**1**	**0.2**	**0.5**	**1**	**0.2**	**0.5**
mean	1.009	0.214	0.464	0.993	0.202	0.506	0.992	0.204	0.502
median	1.008	0.199	0.480	0.985	0.204	0.509	0.983	0.205	0.505
stdev	0.108	0.085	0.183	0.103	0.029	0.153	0.102	0.029	0.151
min	0.660	0.092	-0.275	0.764	0.131	0.070	0.751	0.135	0.070
max	1.278	0.707	0.889	1.335	0.268	0.875	1.335	0.268	0.875

Table 3: IMI vs B&P summary results, Part 3

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 4: IMI: ICL-BIC Criterion

A.1.2 Small Sample Size Results

ICL-BIC Criterion					
	3 Regimes -2 Breaks				
no Regimes	**1**	**2**	**3***	**4**	**5**
mean	1182.843	1132.523	1102.645	1201.914	1295.73
median	1182.719	1128.302	1098.721	1175.584	1263.166
stdev	36.72484	44.23593	44.73931	83.7347	94.99763
min	1073.419	1011.761	1009.369	1079.962	1148.85
max	1286.017	1262.458	1251.009	1525.626	1684.817
	2 Regimes -1 Break: Break in 1/2 sample				
no Regimes	**1**	**2***	**3**	**4**	**5**
mean	529.9614	511.3238	597.78	666.6026	737.396
median	530.4737	511.5959	574.7925	641.9447	714.6363
stdev	22.3064	26.83388	64.74904	66.65845	69.4693
min	473.9721	446.4248	497.7623	560.5054	640.2616
max	581.5548	582.7748	831.0557	920.0295	1029.025
	2 Regimes -1 Break: Break in 1/4 sample				
no Regimes	**1**	**2***	**3**	**4**	**5**
mean	680.2815	660.1716	893.0383	988.0106	1082.347
median	678.904	659.4563	914.9034	1005.61	1099.892
stdev	28.71104	28.55525	138.9608	160.1993	160.3868
min	602.2491	575.9798	657.0156	725.8269	824.1438
max	749.7623	732.4259	1154.217	1373.733	1422.178
	2 Regimes -1 Break: Break in 3/4 sample				
no Regimes	**1**	**2***	**3**	**4**	**5**
mean	675.8609	656.6956	876.5606	974.2372	1044.22
median	678.2106	659.5003	896.2958	989.8087	1059.285
stdev	26.57679	29.17681	148.1347	166.3048	169.9986
min	604.7817	581.6823	650.806	705.0874	789.3558
max	739.0759	728.567	1266.913	1396.797	1387.94
	*true value				

Table 5: IMI: ICL-BIC Criterion

No of Regimes Identified							
3 Regimes - 2 Breaks		**2 Regimes - 1 Breaks: 1/2**		**2 Regimes - 1 Break: 1/4**		**2 Regimes - 1 Break: 3/4**	
No	**% of Sim**	**No**	**% of Sim**	**No**	**% of Sim**	**No**	**% of Sim**
1	0.0	**1**	17.2	**1**	14.5	**1**	13.8
2	13.8	**2***	82.3	**2***	85.5	**2***	86.2
3*	86.2	**3**	0.0	**3**	0.0	**3**	0.0
4	0.0	**4**	0.0	**4**	0.0	**4**	0.0
5	0.0	**5**	0.0	**5**	0.0	**5**	0.0
*: true number of regimes							

Table 6: IMI: No of Regimes Identified

Break Locations					
	3 Reg.-2 Br.		**2 Reg.-1 Br.: 1/2**	**2 Reg.-1 Br.: 1/4**	**2 Reg.-1 Br.: 3/4**
True Locations	**200**	**400**	**150**	**100**	**300**
Mean	200.7	401.2	151.1	100.8	300.6
Median	201	401	151	101	301
Stdev	3.9	2.9	3.9	5.8	4.3
Min	183	392	135	80	290
Max	214	417	176	118	319
b-5<%<b+5	81.5	91.4	85.1	73.3	80

Table 7: IMI: Break Locations

A.2 Simulation Results: Different models across regimes

No Regimes Identified (% of Sim)						
Sample Size=300, Break Location=150						
	IMI			**BP**		
Models	**1,2**	**1,3**	**1,4**	**1,2**	**1,3**	**1,4**
1	0	0	0	0	0	0
2*	84	62	68	49.5	7.5	53.5
3	13	26.5	28.5	30.5	26.5	28.5
4	3	11	3	17.5	41.5	17
5	0	0.5	0.5	2.5	24	1
6	0	0	0	0	0.5	0
Sample Size=1000, Break Location=500						
	IMI			**BP**		
Models	**1,2**	**1,3**	**1,4**	**1,2**	**1,3**	**1,4**
1	0	0	0	0	0	0
2*	99	84.5	70	52.5	16.5	60
3	1	11	28.5	28.5	22.5	26.5
4	0	2.5	1.5	15	38.5	11
5	0	1	0	4	22	2.5
6	0	1	0	0	0.5	0
	* true number of regimes					

Table 8: IMI vs B&P, No Regimes Identified

Break Locations						
	Sample Size=300, Break Location=150					
	IMI			**BP**		
Models	**1,2**	**1,3**	**1,4**	**1,2**	**1,3**	**1,4**
mean	150.97	150.98	150.99	151.32	154.67	150.36
median	151	151	151	150	152	150
stdev	0.712	0.861	0.106	2.559	6.079	0.994
min	149	148	150	149	148	150
max	154	154	151	165	168	157
b-5<%<b+5	100	100	100	91.92	66.67	99.07
	Sample Size=1000, Break Location=500					
	IMI			**BP**		
Models	**1,2**	**1,3**	**1,4**	**1,2**	**1,3**	**1,4**
mean	500.94	501.22	500.99	501.32	503.30	500.36
median	501	501	501	500	501	500
stdev	0.838	0.871	0.143	3.002	6.039	0.786
min	496	497	500	498	499	499
max	503	505	502	520	532	504
b-5<%<b+5	100	100	100	93.33	75.76	100

Table 9: IMI vs B&P, Break Locations

A.3 Empirical Application

ICL-BIC						
	Potential Number of Regimes					
Case	**1**	**2**	**3**	**4**	**5**	**6**
I	496.55	399.27	449.51	509.80	580.21	660.51
II	460.45	379.94	430.18	490.46	560.84	641.22
III	694.81	630.64	624.71	741.22	788.95	848.42

Table 10: Taylor Rules for United States, ICL-BIC

BIC: Determine the Model within Each Regime								
		Candidate Models						
	Regimes	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Case I	**1**	265.65	278.52	124.12	95.80	73.63	124.12	119.75
	2	373.21	382.28	279.94	279.93	279.45	279.94	275.65
Case II	**1**	265.65	278.52	124.12	95.80	73.63	-	-
	2	373.21	382.28	279.94	279.93	279.45	-	-

Table 11: Taylor Rules for United States, BIC of Candidate Models

Parameter Estimates								
	Case I				**Case II**			
Regime		**1**		**2**		**1**		**2**
Model		**5**		**7**		**5**		**5**
	α_0	0.045	α_0	0.032	α_0	0.045	α_0	-0.119
		(0.113)		(0.187)		(0.113)		(0.512)
	ρ	0.787	ρ_1	0.108	ρ	0.787	ρ	0.861
		(0.037)		(0.118)		(0.037)		(0.060)
	α_1	0.422	σ	1.593	α_1	0.422	α_1	0.253
		(0.072)				(0.072)		(0.095)
	α_2	0.627			α_2	0.627	α_2	-0.445
		(0.101)				(0.101)		(0.348)
	α_3	-0.436			α_3	-0.436	α_3	0.658
		(0.110)				(0.110)		(0.354)
	σ	0.337			σ	0.337	σ	1.423
	Case III							
Regime		**1**		**2**		**3**		
Model		**1**		**1**		**1**		
	α_0	3.609	α_0	3.159	α_0	1.536		
		(0.438)		(0.155)		(0.208)		
	α_1	1.195	α_1	0.796	α_1	0.638		
		(0.084)		(0.038)		(0.043)		
	α_2	0.332	α_2	0.425	α_2	0.294		
		(0.104)		(0.044)		(0.051)		
	σ	1.462	σ	0.560	σ	0.718		

Table 12: Taylor Rules for United States, Parameter Estimates

Regime Changes Locations		
Date	**Regime Before**	**Regime After**
Case I		
1985:Q3	2	1
Case II		
1985:Q3	2	1
Case III		
1969:Q2	2	1
1971:Q1	1	2
1971:Q2	2	3
1973:Q2	3	2
1973:Q4	2	1
1975:Q1	1	2
1975:Q2	2	3
1979:Q1	3	2
1980:Q1	2	1
1980:Q4	1	2
1981:Q1	2	1
1992:Q2	1	2
1993:Q2	2	3
1994:Q4	3	2
2001:Q3	2	3

Table 13: Taylor Rules for United States, Regimes Changes Locations

Research Papers 2010

2010-38: Thomas Q. Pedersen: Predictable return distributions

2010-39: Rasmus Tangsgaard Varneskov: The Role of Dynamic Specification in Forecasting Volatility in the Presence of Jumps and Noisy High-Frequency Data

2010-40: Antonis Papapantoleon and David Skovmand: Picard Approximation of Stochastic Differential Equations and Application to Libor Models

2010-41: Ole E. Barndorff–Nielsen, Fred Espen Benth and Almut E. D. Veraart: Modelling electricity forward markets by ambit fields

2010-42: Robinson Kruse and Philipp Sibbertsen: Long memory and changing persistence

2010-43: Dennis Kristensen: Semi-Nonparametric Estimation and Misspecification Testing of Diffusion Models

2010-44: Jeroen V.K. Rombouts and Lars Stentoft: Option Pricing with Asymmetric Heteroskedastic Normal Mixture Models

2010-45: Rasmus Tangsgaard Varneskov and Valeri Voev: The Role of Realized Ex-post Covariance Measures and Dynamic Model Choice on the Quality of Covariance Forecasts

2010-46: Christian Bach and Stig Vinther Møller: Habit-based Asset Pricing with Limited Participation Consumption

2010-47: Christian M. Dahl, Hans Christian Kongsted and Anders Sørensen: ICT and Productivity Growth in the 1990's: Panel Data Evidence on Europe

2010-48: Christian M. Dahl and Emma M. Iglesias: Asymptotic normality of the QMLE in the level-effect ARCH model

2010-49: Christian D. Dick, Maik Schmeling and Andreas Schrimpf: Macro Expectations, Aggregate Uncertainty, and Expected Term Premia

2010-50: Bent Jesper Christensen and Petra Posedel: The Risk-Return Tradeoff and Leverage Effect in a Stochastic Volatility-in-Mean Model

2010-51: Christos Ntantamis: A Duration Hidden Markov Model for the Identification of Regimes in Stock Market Returns

2010-52: Christos Ntantamis: Detecting Structural Breaks using Hidden Markov Models