Clustering Macroeconomic Variables

Chiara Perricone*

University of Rome "Tor Vergata"

September 28, 2012

Abstract

Many papers have highlighted that some macroeconomic time series present structural instability, for example NIPA decompositions of real GDP, money, credit, interest rates and stock prices series as analyzed in Stock and Watson (2003). The causes of these remarkable changes in the reduced form properties of the macroeconomy is a debated argument. In literature this issue is handled with three main econometric methodologies: structural breaks, regime-switching and time-varying parameters (TVP). Nevertheless all these approaches need some ex ante structure in order to model the change.
Based on the Recurrent Chinese Restaurant Process, we have specified a model for an autoregressive process and estimated it via particle filter assuming a conjugate prior, i.e. we have applied the idea of evolutionary cluster to the study of the instability in output and inflation for US after War World II. This procedure displays some advantages, in particular does not require a strong ex ante structure in order to neither detect the breaks nor manage the parameters' evolutions, avoiding the main drawbacks of the standard methodologies. The application of the cluster procedure to GDP growth and inflation rate for US from 1957 to 2011 shows a good ability in fit the data, producing a clusterization of the time series that could be interpreted in terms of economic history and recovering the key data features without making restrictive assumptions.
Considering the open debate on the source of the Great Moderation, i.e. Bad Policy and/or Bad Luck, under the caveat that until now we have not studied a VAR or a structural form, this approach leads to conclusions that support the findings of Cogley and Sargent (2001, 2005): the presence of both Bad Policy and Bad Luck. In fact there are evidences of changes over time in both volatility and autoregressive coefficients for the AR specification presents, even if the latter are less marked.

JEL: C18 C22 C51 E17

*chiara.perricone@gmail.com

1 Introduction

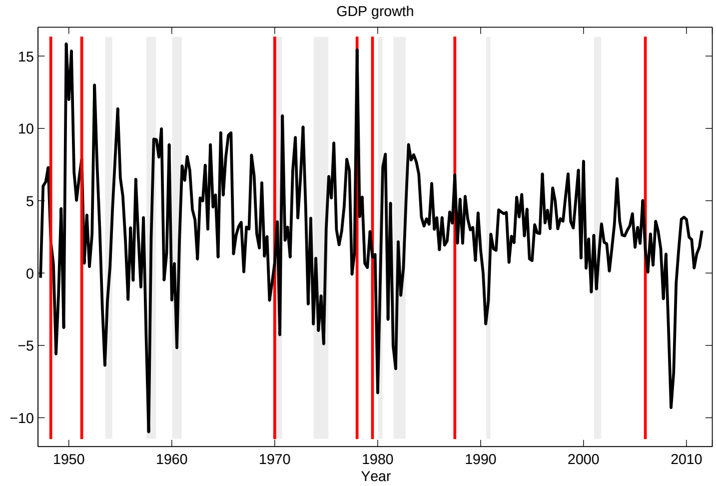
Many papers have highlighted that some macroeconomic time series present structural instability, the main example is Stock and Watson (2003), where the authors consider data on 168 quarterly U.S. macroeconomic time series, from 1959 to 2001, studying both reduced and structural form. In particular we are going to consider the annualized quarterly series for output and inflation for the period $1957Q1 - 2011Q3$ as in figure 1. The shaded areas represent the NBER recessions and the vertical red lines mark the appointment dates of the Federal Reserve chairmen[1]. The instability, i.e the heterogeneity in both level and volatility of the time series over the studied period, is highlighted by the summary statistics in table 1: some stylized facts emerge. Over the $'60s$ inflation was relatively low and stable, then during the late $'60s$ it started rising and run amok in the late $'70s$. At the same time the economy experienced a deep and long recession following the oil crisis of 1974. During the first half of the $'80s$ the economy went through a difficult disinflation: inflation went back to the levels that were prevailing before the $'70s$ at the cost of two severe recessions. From the mid-$'80s$, until the recent financial crisis, the economy has been characterized by remarkable economic stability. Economists like to refer to this last period, after 1984, with the term ‘Great Moderation’, while the name ‘Great Inflation’ is often used to label the turmoil of the $'70s$. The contrast between the two periods is evident.
The causes of these remarkable changes in the reduced form properties of the macroeconomy is a debated argument: some authors suggest that these changes are the result of exogenous shocks, the so call ‘Bad Luck’ interpretation, whereas some others claim that from the mid-$'80s$ the policy makers have reached a better understanding of the economy such that they were able to reduce the volatility of inflation and interest rate. This latter interpretation goes under the name of ‘Bad Policy’ and it is supported by the observation that the sharp decline in inflation started shortly after Paul Volcker was appointed chairman of the Federal Reserve in August 1979. It is tempting to link the two events and conclude that a substantial change in the conduct of monetary policy must have occurred in those years.

	Output Growth		**Inflation Rate**	
Period	Mean (%)	Standard Dev (%)	Mean (%)	Standard Dev (%)
1947 - 2011	3.1	3.9	3.3	2.7
1947 - 1959	3.7	5.5	2.2	3.0
1960 - 1969	4.0	3.5	2.3	1.4
1970 - 1983	2.9	4.8	6.4	2.6
1984 - 1992	2.9	2.1	3.4	1.1
1993 - 2011	2.5	2.7	2.0	1.4

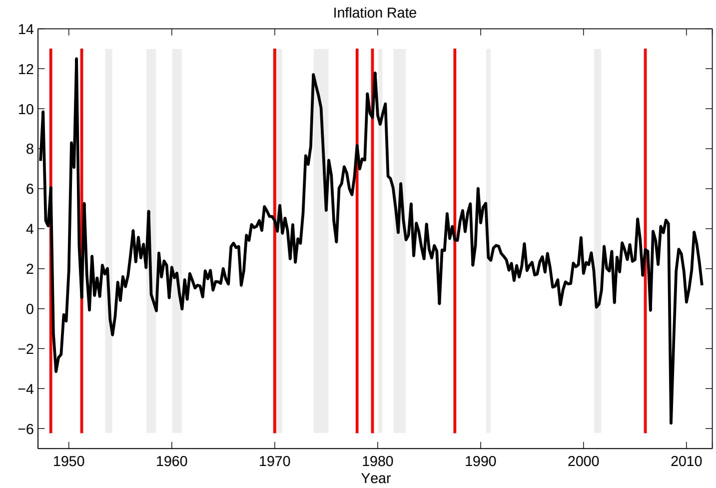
Table 1: Summary statistics for output growth and inflation rate.

Given these stylized facts, there is a growing literature that allows for parameters instability and many declinations of change-point modeling have been applied to the study of instability in U.S. real activity and inflation, generating heterogeneous and contradictory results. Let us considering only few examples, in order to foresee the variety of methodologies and results.
Kim and Nelson (1999) and Kim, Nelson and Piger (2004), assuming a single

[1] Thomas B. McCabe (April 15, 1948 - April 2, 1951); William McChesney Martin, Jr. (April 2, 1951 - February 1, 1970); Arthur F. Burns (February 1, 1970 - January 31, 1978); G. William Miller (March 8, 1978 - August 6, 1979); Paul A. Volcker (August 6, 1979 - August 11, 1987); Alan Greenspan (August 11, 1987 - January 31, 2006); Ben S. Bernanke (February 1, 2006 -)



(a)



(b)

Figure 1: Output growth 1(a) and inflation rate 1(b), 1947Q2 - 2011Q3.

change-point, investigate breaks in various measures of GDP. For most of the measures they consider, the likelihood of a break is overwhelming and Bayesian and frequentist analyses produce very similar results. Whereas Blanchard and Simon (2001) present evidences that the decline in variance might have been more gradual, as a part of an ongoing trend, discarding the idea of a single break point. Clark (2003) supports the presence of time-varying volatility in inflation and Sims and Zha (2006), using Markov switching VAR, identify changes in the volatility of the structural disturbance as the key driven behind the stabilization of the U.S. economy. Cogley and Sargent (2001, 2005) find variation in both conditional mean and conditional variance for the time series of inflation. Using a battery of break tests applied to time-varying autoregressive models Stock (2001) obtains little evidence for variation in the conditional mean of inflation, as well as in Stock and Watson (2003), where the authors conclude that for variables that measure the real economic activity, the moderation generally is associated with reductions in the conditional variance in the time-series models, not with changes in the con-

ditional means. This result is supported by Primiceri (2005), who concludes that there is a sensible variation in the conditional variance but little in the conditional mean.
In the cited studies the issue of structural instability in macroeconomics time series is handled with three main econometric methodologies: structural breaks, regime-switching and time-varying parameters (TVP). All these models have advantages and drawbacks.
The main advantage of models with a small number of structural breaks, typically from 1 to 3, is that they do not restrict the change's magnitude, that can happen after a break, in the coefficients used to model the time series but implicitly assume that after the last estimated break in the sample, there will be no more breaks. Moreover in some cases the researcher has to impose the number or the dates of breaks.
In contrast, in the TVP models, which ensure lot of flexibility, the research implicitly assumes that there is probability of a break equal to 1 in the next observation. Another disadvantage of the TVP model is that the size of the break is severely limited by the assumption that coefficients evolve according to a random walk.
Given these observations Koop and Potter (2007), KP henceforth, suggest a new model, drawn from their beliefs on some desirable features for a change-point model. They propose the following criteria:

1. The number of regimes and their maximum duration should not be restricted ex ante.
2. The regime duration distribution should not be restricted to be constant or monotonically decreasing/increasing.
3. The parameters characterizing a new regime can potentially depend on the parameters of the old regime.

Moreover, for the empirical evaluation of these models, there are other two points to consider:

4. Assuming a Bayesian approach, the prior distribution of the parameters in each regime should, if possible, be conjugate to the likelihood to minimize the computational complexity.
5. The change-point model should be easy to update in real time as new data arrive.

With respect to KP, we would like to add two more ideas:

6. The estimation of a time-varying parameters does not need to imply that in every time there is a change in regime.
7. Considering a time-varying estimation, the magnitude of the changes should not be big inside the same regime.
8. The number of regimes and their temporal boundary, without imposing any restrictions, should be easily to interpret from the economic point of view.

It can immediately be seen that standard implementations of the TVP models and models with small numbers of breaks do not have these features.
Therefore, using these criteria as building blocks, we suggest an idea, based on the Recurrent Chinese Restaurant Process (RCRP) and the particle filter procedure, which applies the notion of evolutionary cluster to the study of instability in time series.
An observation at time t is characterized by two informations: a cluster belonging and a set of parameters that identify the applied model. we assume that two observations of a time series belong to the same cluster, or regime[2], if the param-

[2]With a slight abuse of notation, we use the terms 'cluster' and 'regime' interchangeably in the paper.

eters that characterized the assumed model are 'close' to each others. Whereas if the set of parameters at time t, or a subset of it, is different with respect to the set of parameters at time $t-h$, then the observations y_t and y_{t-h} belong to different clusters. One or more observations belong to each cluster, and therefore one or more set of parameters are associated to a specific cluster, assuming that inside the same regime the parameters that define different observations could not significantly differ. In order to summarize the informations on the parameters, the vector of coefficients representative a regime is the centroid of all parameters identifying observations that belong to a specific cluster.

This approach present some advantages with respect to the previously cited methodologies. With respect to the structural breaks analysis, we don't need to impose any ex ante knowledge on the number or the dates of breaks. In opposition with the regime switching approach, as well as with respect to KP, we don't need to impose a structure on the evolution or the duration of regimes. And finally, comparing this methodology with the TVP, I do not assume an evolutionary mechanism common to all periods that drives the coefficients from one regime to another. In fact the parameters' evolution is driven by different mechanism within and between clusters. If y_{t-1} and y_t belong to the same regime the new set of parameters is obtained from the set at the previous time, evolving with the Metropolis kernel. On contrary, if the two observations belong to different clusters, the parameters at time t will be drawn, independently to the past, from a based distribution. Indeed since the transition mechanism does not need to move the parameters from a regime to another and given the assumption of parameters' homogeneity inside clusters, the volatility of the parameters' evolution mechanism is lower with respect to what you need in a standard TVP model, where, as a drawback, the size of the break is severely limited by the assumption that coefficients evolve according to a random walk.

Moreover, with this approach, the cluster could be recurrent, i.e. the time series could belong to a cluster which it had previously visited, without imposing a monotonic movement between clusters or a single jump, i.e. the series could jump from cluster j to, for example, cluster $j-2$, where $j>2$.

For the estimation procedure, as KP, we use Bayesian methods, which are attractive for change-point models since they allow for flexible relationships between parameters in various regimes and are computationally simple.

The empirical analysis we perform is based on the GDP growth and the inflation rate as measured by the PCE deflator from 1957 Q4 through 2011 Q3 for U.S., both at an annual rate. The assumed model for both variables is an AR(2) with an intercept, as in KP.

The application of the evolutionary cluster procedure to US data shows a good ability in fit the observed values, moreover this methodology produces a clusterization of the time series that could be interpreted in terms of economic history and it is able to recover key data features without making restrictive assumptions, as in 'one-break' or TVP models. In particular the model is able to replicate the decreasing in volatility for both time series, underlining the relation between stylized facts and episodes like the Fed's chairmen appointment or the end of specific recession periods. It is interesting that the model assigns the observations of the first ten years of the Great Moderation to two different clusters: a regime goes from 1983 to 1987, whereas the observations between 1987 and 1992 belong to another cluster. Nevertheless the posterior modes for the estimations of the first coefficient in the AR(2) model are equal in the two regimes. This result can be interpreted as a clue that the Fed's target for the monetary policy change during the Great Moderation and that this policy was pursued under both Volcker's and Greenspan's leadership.

Considering the open debate on the source of Great Moderation period, under

the caveat that until now we have not studied a VAR or a structural form, this approach presents conclusions that support Cogley and Sargent (2001, 2005) findings, suggesting changes in both volatility and coefficients, even if the latter are less marked.
The chapter is structured as follow: in Section II we recall some theoretical notions that are the building blocks of the main process used to construct the model, the so called Recurrent Chinese Restaurant Process. In Section III and IV the model and the inference methodology are specified, whereas in Section V we present the results for the macroeconomic series of interest. Finally the conclusions and further work in Section VI.

2 The theoretical background: from the Dirichelt Process to the Recurrent Chinese Restaurant Process

In this section we are going to introduce the Recurrent Chinese Restaurant Process from the building block: the Dirichlet Process.
The Dirichlet process (DP) is a probability distribution over probability distributions, i.e. draws from the DP are themselves discrete probability distributions. This characteristic allows the DP to be used as a prior in nonparametric Bayesian models. More formally, the DP is a stochastic process whose sample paths are probability measures with probability one, and whose finite dimensional marginal distributions are Dirichlet distributed.

Definition 1 ***Dirichlet Process (DP)***
Let $\mathbb{G}_0$ be a distribution over Θ and let $\alpha \in \mathcal{R}^+$. Further let $A_1, \dots, A_r$ be a finite measurable partition of Θ. Then the random vector $(G(A_1), \dots, G(A_r))$ is distributed according to a Dirichlet process, with base distribution $\mathbb{G}_0$ and concentration parameter α, written $G \sim DP(\alpha, \mathbb{G}_0)$, if $(G(A_1), \dots, G(A_r))) \sim Dir(\alpha\mathbb{G}_0(A_1), \dots, \alpha\mathbb{G}_0(A_r))$ for any finite measurable partition $A_1, \dots, A_r$ of Θ.

The parameters $\mathbb{G}_0$ and α can intuitively be understood as the mean and the precision of the DP respectively. Let $G \sim DP(\alpha, \mathbb{G}_0)$, therefore:

$$\mathbb{E}[G(A)] = \mathbb{G}_0(A) \qquad \text{and} \qquad Var[G(A)] = \frac{\mathbb{G}_0(A)\,[1 - \mathbb{G}_0(A)]}{\alpha + 1}$$

As $\alpha \to \infty$ we have $G(A) \to \mathbb{G}_0(A)$ for any measurable A, i.e. G converges to $\mathbb{G}_0$ pointwise.
Moreover as $G \sim DP(\alpha, \mathbb{G}_0)$, we can draw samples $\theta_1, \dots, \theta_N \sim G$ and consider the posterior distribution $G|\theta_1, \dots, \theta_N$. Let $A_1, \dots, A_r$ be a finite measurable partition of Θ, and let n_k be the number of observed values θ_i in A_k, i.e. $n_k = \#\{\theta_i | \theta_i \in A_k\}$. Then from the Dirichlet-Multinomial conjugacy we have

$$(G(A_1), \dots, G(A_r))|\theta_1, \dots, \theta_N \sim Dir(\alpha\mathbb{G}_0(A_1) + n_1, \dots, \alpha\mathbb{G}_0(A_r) + n_r)$$

which, by definition of DP, implies that the posterior distribution is a DP as well, with parameters given by

$$G|\theta_1, \dots, \theta_N \sim DP\left(\alpha + N, \frac{\alpha}{\alpha + N}\mathbb{G}_0 + \frac{N}{N + \alpha}\frac{\sum_{i=1}^N \delta_{\theta_i}}{N}\right)$$

In other words, the posterior is distributed according to a DP whose base distribution is a weighted average of the prior base distribution $\mathbb{G}_0$ and the empirical distribution of the observations $\frac{\sum_{i=1}^N \delta_{\theta_i}}{N}$, where δ_{θ_i} is a Dirac function.

Therefore the DP is a conjugate family that is closed under posterior updates.
One of the uses of the Dirichlet process is as mixing distribution in a mixture model, which in turn can be used for density estimation or clustering.

Definition 2 ***Dirichlet Process Mixture (DPM)***
A Dirichlet process mixture (DPM) is a model of the following form

$$\begin{aligned} y_i|\phi_i &\sim F(\phi_i) \\ \phi_i|G &\sim G \\ G &\sim DP(\alpha, \mathbb{G}_0) \end{aligned}$$

where y_i (for $i = 1, \ldots, N$) are observations that are modeled as exchangeable draws from a mixture of distributions $F(\phi_i)$ and the mixing distribution G, which is drawn from a DP with base distribution $\mathbb{G}_0$ and concentration parameter α.

Note that such a model is equivalent to the limit of the following finite mixture model with K components as $K \to \infty$:

$$\begin{aligned} y_i|c_i, \theta_{c_i} &\sim F(\theta_{c_i}) \\ c_i|p_1, \ldots, p_K &\sim \text{Discrete}(p_1, \ldots, p_K) \\ \theta_{c_i} &\sim \mathbb{G}_0 \\ p_1, \ldots, p_K &\sim Dir\left(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}\right) \end{aligned}$$

where $c_i \in 1, \ldots, K$ is the class label of data point y_i, θ_{c_i} are the parameters associated with the cluster c_i and p_k, for $k = 1, \ldots, K$, are the mixing proportions on which a symmetric Dirichlet prior is placed. Mixture models are a widely used tool for density estimation and clustering. However, finite mixture models, i.e. mixture models with a fixed, finite number of components K, can only be applied when the number of components is known a priori or model selection procedures have to be used to choose K. On the other hand, infinite mixture models bypass the model selection problem by, in principle, allowing an infinite number of components, where only a finite number of them will have data associated.
While DPMs are very flexible and powerful tools for modeling independent and identically-distributed data, they cannot directly be applied when there is some temporal or spatial structure associated with the data. In order to model dependent data, we need to consider a Dependent Dirichlet Process (DDP). Griffin and Steel (2006) give the following definition:

Definition 3 ***Dependent Dirichlet Process (DDP)***
A dependent Dirichlet process is a stochastic process defined on the space of probability measures over a domain, indexed by time, space or a selection of other covariates in such a way that the marginal distribution at any point in the domain follows a Dirichlet process.

Models that capture this idea have recently been developed in different papers, in particular Caron, Davy and Doucet (2007) and Ahmed and Xing (2008).
For this analysis we have implemented a Dependent Dirichlet Process to capture a time structure that can be referred to as time-varying Dirichlet process mixture models (TVDPM). Such models can, among other things, be used for 'evolutionary clustering', i.e. clustering of data that has a temporal structure.
There are two main approaches to model a TVDPM: (i) based on the stick-breaking representation of the DP, as MacEachern (2000), Griffin and Steel (2006), or (ii) based on the Chinese Restaurant Process (CRP), as Caron, Davy and Doucet (2007) or Ahmed and Xing (2008).
In particular, we have implemented the idea of Ahmed and Xing, who describe a TVDPM specifically designed for the application of clustering data with temporal

structure. Their approach operates by, on the one hand, tying together the CRP probabilities of consecutive time steps, and, on the other hand, allowing the cluster parameters to develop according to some Markovian dynamics from one time step to the next. Let define in details the recurrent Chinese Restaurant Process beginning with the main objects of the metaphor: the Chinese Restaurant. An observation at time t is pictured as a customer entering in a Chinese restaurant. The tables in the restaurant are the clusters: like a customer can pick up a table, an observation can lead to one of the existing cluster. On each table is served a specific dish, which is the centroid of the cluster, i.e. the set of parameters that characterized a specific cluster given the assumed model. If a customer choses a table, he is selecting a specific dish, which each time could be prepared with minor differences, i.e. the parameters that characterize two observations belonging to the same regime have to be similar but not identical.

In the time-dependent variant of the Chinese Restaurant Process due to Ahmed and Xing, named the Recurrent Chinese Restaurant Process (RCRP), one assumes that customers to arrive in fixed times $t = 1, \dots, T$ and there is a dependence between times. Therefore the metaphor slightly change: in the first epoch, the customers are seated according to the standard CRP, where the probability of an existing table is given by its popularity today, whereas the probability of a new table is proportional to the concentration parameter of the DP. In all remaining epochs, however, the table sizes at the previous epoch[3] are used to determine the popularity of a table today. Specifically, the conditional probability that the i^{th} data point arrived at time t belongs to the k^{th} cluster, i.e. $c_{t,i} = k$, is defined as

$$P(c_{t,i} = k | c_{t-1,1:K_{t-1}}, c_{t,1:i-1}) \propto \begin{cases} n_{k,t-1} + n_{k,t}^{(i)} & \text{if } k \text{ is an old cluster} \\ \alpha & \text{for a new cluster} \end{cases} \tag{1}$$

where $n_{k,t-1}$ is the number of customers seated at table k in time $t-1$, and $n_{k,t}^{(i)}$ is the number of customers seated at table k in epoch t after observing the first $(i-1)^{ths}$ data points. If k is an old cluster, i.e. it had already data associated yesterday, the parameters $\theta_{k,t}$ are carried over from $t-1$ to t by means of a Markovian transition kernel $P(\theta_{k,t}|\theta_{k,t-1})$, whereas if k is a new cluster, the parameters $\theta_{k,t}$ are drawn from $\mathbb{G}_0$. Therefore the conditional distribution of $\theta_{k,t}$ given the cluster and the past is define as:

$$P(\theta_{k,t}|\theta_{k,t-1}, c_t = k) = \begin{cases} P(\theta_{k,t}|\theta_{k,t-1}) & \text{if k is an old cluster} \\ \mathbb{G}_0(\theta_{k,t}) & \text{for a new cluster} \end{cases}$$

3 The model

The approach taken here is to model y_t as observations from a Recurrent Chinese Restaurant Dirichlet Process Mixture model (RCRDPM), where $t = 1, \dots, T$. In order to fully specify the model, the mixed distribution $F(\cdot)$, the base distribution $\mathbb{G}_0$ and the transition kernel $P(\theta_{k,t}|\theta_{k,t-1})$ have to be defined.

3.1 Mixed Distribution

Let us define the mixed distribution $F(\cdot)$ or, equivalently, the probability $P(y_t|c_t = k, \theta_{k,t})$ of the observed data conditional to the fact that y_t belong to the cluster k and the model is characterized by the parameters $\theta_{k,t}$ associated to the cluster k. We assume the observed data follow an $AR(p)$ model of the form:

$$y_t = \psi_{0,t} + \psi_{1,t} y_{t-1} + \psi_{2,t} y_{t-2} + \dots + \psi_{p,t} y_{t-p} + \sigma_t \varepsilon_t$$

[3]For simplicity we assume only a time dependence of one period, even if there are no restrictions on the possible time dependence.

where $\varepsilon_t \sim \mathcal{N}(0,1)$. Therefore we define:

$$P(y_t|x_t, c_t = k, \theta_{k,t}) = \mathcal{N}(y_t|x_t\Psi_{k,t}, \sigma^2_{k,t})$$

where $\theta_{k,t} = (\Psi_{k,t}, \lambda_{k,t})$, $\Psi_{k,t}$ is $(\psi_{0,t}, \psi_{1,t}, \psi_{2,t}, \ldots, \psi_{p,t})$, $\lambda_{k,t}$ is $1/\sigma^2_{k,t}$ and x_t stacks the past observations, i.e. $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$.

3.2 Base distribution $\mathbb{G}_0$

In the RCRDPM model, as well as in DP mixture models, the base distribution $\mathbb{G}_0$ acts as a prior over the parameters $\theta_{k,t}$, therefore it should reflect our prior belief of how we expect the means and precisions of the individual clusters to vary. One choice for $\mathbb{G}_0$ computationally attractive and not unreasonable from a modeling perspective is the conjugate prior to the likelihood $P(y|x,\theta)$, i.e. the Normal-Gamma Distribution given by

$$\mathbb{G}_0(\theta_{k,t}) \equiv \mathbb{G}_0(\Psi_{k,t}, \lambda_{k,t}) = \mathcal{N}(\Psi_{k,t}|\Psi_0, n_0\lambda_{k,t})\Gamma(\lambda_{k,t}|a,b)$$

The main advantage of this choice is that the predictive distribution and the posterior are analytically tractable, simplifying the inference procedures.

3.3 Transition Kernel

The transition kernel $P(\theta_{k,t}|\theta_{k,t-1})$ models the evolution of component parameters through time, inside the same cluster. Given Definition 3 of Dependent Dirichlet Process, marginally the model has to be distributed as a DP, therefore the major restriction on the choice of the transition kernel is that it has to fulfill

$$\int \mathbb{G}_0(\theta_{k,t-1})P(\theta_{k,t}|\theta_{k,t-1})d\theta_{k,t-1} = \mathbb{G}_0(\theta_{k,t}) \tag{2}$$

In other words, $\mathbb{G}_0$ has to be an invariant distribution for the Markovian chain $P(\theta_{k,t}|\theta_{k,t-1})$.
A possible choice for $P(\theta_{k,t}|\theta_{k,t-1})$ is the one based on the update rule of the Metropolis algorithm, which has the advantage that it can be combined with any base distribution and likelihood. The main drawbacks of the Metropolis kernel is that the dependence structure it introduces, i.e. how the parameters evolve from $t-1$ to t, is not easily characterizable, and that one cannot sample directly from:

$$P(\theta_{k,t}|\theta_{k,t-1}, y_t) = \frac{P(\theta_{k,t}|\theta_{k,t-1})P(y_t|\theta_{k,t})}{\int P(\theta_{k,t}|\theta_{k,t-1})P(y_t|\theta_{k,t})d\theta_{k,t}} \tag{3}$$

A possible alternative to (3) is the auxiliary variable kernel[4], which removes both the drawbacks of the Metropolis kernel but requires a possibly large number of auxiliary variables and has to be re-derived for each choice of $\mathbb{G}_0$.
The Metropolis Update Rule Kernel is given by:

$$\begin{aligned} P(\theta_{k,t}|\theta_{k,t-1}) &= S(\theta_{k,t-1}, \theta_{k,t})A(\theta_{k,t-1}, \theta_{k,t}) \\ &+ \left(1 - \int S(\theta_{k,t-1}, \tilde{\theta})A(\theta_{k,t-1}, \tilde{\theta})d\tilde{\theta}\right)\delta_{\theta_{k,t-1}}(\theta_{k,t}) \end{aligned} \tag{4}$$

[4]Given an invariant distribution $\mathbb{G}_0$ it is possible to introduce a set of auxiliary variables $z_{k,t,1:M} = (z_{k,t,1}, \ldots, z_{k,t,M})$ that fulfill $P(\theta_{k,t}|\theta_{k,t-1}) = \int P(\theta_{k,t}|z_{k,t,1:M})P(z_{k,t,1:M}|\theta_{k,t-1})dz_{k,t,1:M}$ such that $P(\theta, \boldsymbol{z}) = P(\boldsymbol{z}|\theta)\mathbb{G}_0$, where θ denotes the parameters of any given cluster at any given time step and $\boldsymbol{z}$ denotes the corresponding set of auxiliary variables. See Pitt and Walker (2005).

where $S(\theta_{k,t-1}, \theta_{k,t})$ is a symmetric proposal distribution, $A(\theta_{k,t-1}, \theta_{k,t}) = \min\left\{1, \frac{\mathbb{G}_0(\theta_{k,t})}{\mathbb{G}_0(\theta_{k,t-1})}\right\}$ is the probability of move and $\delta_{\theta_{k,t-1}}(\theta_{k,t})$ is a Dirac function, such that:

$$\delta_{\theta_{k,t-1}}(\theta_{k,t}) = \begin{cases} 0 & \text{if } \theta_{k,t-1} \neq \theta_{k,t} \\ \infty & \text{if } \theta_{k,t-1} = \theta_{k,t} \end{cases}$$

and

$$\int \delta_{\theta_{k,t-1}}(\theta_{k,t}) d\theta_{k,t} = 1$$

It can easily be shown that the Metropolis kernel fulfills the restriction in equation(3) allowing us to use it with an arbitrary base distributions $\mathbb{G}_0$ as long as they can be evaluated. Sampling from this transition kernel is straightforward: once a sample $\theta_{k,t}$ is obtained from the proposal distribution, it will be accepted with probability $A(\theta_{k,t-1}, \theta_{k,t})$, otherwise with the complementary probability the draw will be rejected and $\theta_{k,t} = \theta_{k,t-1}$. One possible and popular choice for the proposal distribution is an isotropic Gaussian centered at the old value

$$S(\theta_{k,t-1}, \tilde{\theta}) = \mathcal{N}(\tilde{\theta}|\theta_{k,t-1}, sI). \tag{5}$$

The parameter s controls the spread of the proposed jumps, and therefore the dependence among consecutive time steps induced by the kernel: small values for s will lead to mostly small changes, which will have a high probability of being accepted, leading to high correlations between consecutive time steps, while larger values will lead, on one hand, to larger jumps which decrease the correlation between $\theta_{k,t-1}$ and $\theta_{k,t}$ but, on the other hand, also to more rejections, increasing the correlation, since a rejection implies $\theta_{k,t-1} = \theta_{k,t}$.
By definition, observations that belong to the same cluster are characterized by similar parameters, indeed the change implied by the transition kernel should be small. This is one of the main improvement with respect to the standard TVP: the evolutionary structure does not need to address big variations in coefficients between time t and $t-1$, inducing a lot of variability, that is a quite criticized point of the TVP approach, since the transition kernel does not have to deal with change of regimes.

4 Inference

So far we have presented the model specification, in this section we will define the estimation strategy.
There are two main possible inference procedures: a sequential Monte Carlo (SMC) algorithm or a Metropolis-Hastings (M-H) sampler. we implement the SMC algorithm, in the form of particle filter, similar to the approach proposed in Caron, Davy and Doucet (2007).
A Sequential Monte Carlo algorithm belongs to a family of sampling methods that allow recursive sampling from a sequence of distributions $\pi_t(z_{1:t})$, for $t = 1, \ldots, T$ where $\pi_t(\cdot)$ is called target distribution. The main idea behind the SMC methods is that if the state $z_{1:t}$ up to time t consists of an old component $z_{1:t-1}$ and a new component z_t , i.e. $z_{1:t} = (z_{1:t-1}, z_t)$, then samples from $\pi_{t-1}(z_{1:t-1})$ can be used to construct samples from $\pi_t(z_{1:t})$. A widely used sequential Monte Carlo technique is known as sequential importance sampling (SIS), which recursively uses importance sampling to obtain weighted samples from the target distribution.
Both cited inference schemes have their strengths and weaknesses: the particle filter has the advantage that it can be used sequentially, i.e. predictions can be made as new data comes in, and it is computationally efficient. However, there are also some drawbacks. In particular sequential Monte Carlo methods are known to behave poorly in high dimensional settings, possibly leading to an inaccurate

representation of the posterior distribution: in fact, the posterior estimate might collapse to a single point. If one is only interested in the Maximum A Posteriori (MAP) solution this is not overly problematic as one can then simply view the particle filter as a method for finding the MAP solution.

Algorithm 1 outlines a sequential Monte Carlo inference procedure for the sequence of target distributions:

$$
\begin{aligned}
\pi_t(c_{1:\,t}, \Theta_{1:\,t}) &= \pi_{t-1}(c_{1:\,t-1}, \Theta_{1:\,t-1}) && (6)\\
& \times P(c_t|\Theta_t, c_{1:\,t-1}, y_t) && (7)\\
& \times \prod_{k=1}^{K_t} \begin{cases} P(\theta_{k,t}|\theta_{k,t-1}) & \text{if } k \leq K_{t-1} \\ \mathbb{G}_0(\theta_{k,t}) & \text{if } k = K_t \end{cases} && (8)
\end{aligned}
$$

where $\theta_t = (\theta_{1,t}, \ldots, \theta_{K_t,t})$ is the vector of parameters for all possible clusters at time t, i.e. $c_t = k$ for $k = 1, \ldots, K_t$, where $K_t \equiv K_{t-1} + 1$, i.e. the number of possible clusters today is the number of observed clusters yesterday plus one.

After the weights initialization, at each time step t, the SMC algorithm first moves each particle forward in time by sampling a label for the data point y_t and new associated parameters from suitable importance distributions (forecast step), then computes the weight of each particle (update step) and finally selects a new particle set by resampling. In the following we will specify the main choices for each step.

Algorithm 1 RCRDPM via particle filter

1: **for** $i = 1, \dots, N$ **do**
2: $\quad w_0^{(i)} = \frac{1}{N}$
3: **end for**
4: **for** $t = 1, \dots, T$ **do**
5: $\quad$ **for** $i = 1, \dots, N$ **do**
6: $\quad\quad$ Sample the cluster $c_t^{(i)} \sim P(c_t|\Theta_{t-1}^{(i)}, c_{1:t-1}^{(i)}, y_t)$
7: $\quad\quad$ **if** $c_t^{(i)} = K_t$, i.e. you have sample a new cluster **then**
8: $\quad\quad\quad$ Sample $\theta_{c_t^{(i)},t}^{(i)} \sim q_1(\theta|y_t)$
9: $\quad\quad\quad$ Compute the important ratio for the weight: r
10: $\quad\quad$ **else** $c_t^{(i)} \leq K_{t-1}$, i.e. you have sample an old cluster
11: $\quad\quad\quad$ Sample $\theta_{c_t^{(i)},t}^{(i)} \sim q_2(\theta|y_t, \theta_{c_t^{(i)},t-1}^{(i)})$
12: $\quad\quad\quad$ via the Metropolis Transition Kernel
13: $\quad\quad\quad$ Compute the important ratio for the weight: r
14: $\quad\quad$ **end if**
15: $\quad\quad$ Compute the new weights $\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \times r$
16: $\quad\quad$ Check if the estimated coefficients lead to a stable series.
17: $\quad$ **end for**
18: $\quad$ Normalize the weights: $\hat{w}_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{i=1}^N \tilde{w}_t^{(i)}}$
19: $\quad$ **if** $\left[\sum_{i=1}^N (\tilde{w}_t^{(i)})^2\right]^{-1} \leq \frac{N}{2}$ **then**
20: $\quad\quad$ Resampling: Duplicate the particles with large weights
21: $\quad\quad$ and remove the particles with small weights, resulting in a
22: $\quad\quad$ new set of particles $i_1, \dots, i_N$ with equal weights $w_t = \frac{1}{N}$.
23: $\quad$ **else**
24: $\quad\quad$ Take the old particles, with weights $w_t^{(i)} = \hat{w}_t^{(i)}$
25: $\quad$ **end if**
26: **end for**

Line 6: Sample the Cluster

The observation y_t can belong to one of the cluster obtained until time $t-1$, i.e. the cluster's label could be in the set $\{1, 2, \dots, K_{t-1}\}$, or it can belong to a new cluster and the associated label is K_t. Therefore, at time t the number of possible clusters is equal to K_t and the distribution $P(c_t|\theta_{t-1}, c_{1:t-1}, y_t, x_t)$, for $c_t = 1, \dots, K_t$, is given by:

$$P(c_t|\theta_{t-1}, c_{1:t-1}, y_t, x_t) = \begin{cases} P(y_t|\theta_{k,t-1}, x_t) & \text{if } k \leq K_{t-1} \\ \int P(y_t|\theta, x_t)\mathbb{G}_0(\theta)d\theta & \text{if } k = K_t \end{cases} \tag{9}$$

For any old cluster, i.e. $c_t \in \{1, 2, \dots, K_{t-1}\}$, the cluster's distribution is simply the likelihood of y_t, given $\theta_{k,t-1}$, i.e the set of parameters that characterized the cluster k at time $t-1$, and the past x_t .
On contrary for the new cluster, since we have chosen the base distribution to be conjugate to the likelihood, we can analytically evaluate the probability of the data point when integrating over the base distribution $\int P(y_t|\theta, x_t)\mathbb{G}_0(\theta)d\theta$, which is a Student-t distribution:

$$P(y_t|\theta_{k,t-1}, x_t) = St(y_t|x_t'\Psi_0, f(x_t)\frac{a}{b}, 2a)$$

where $f(x_t) = 1 - x_t(x_t'x_t + n_0)^{-1}x_t'$.
Therefore to sample from the distribution in (9) the steps are: evaluate $P(c_t|\theta_{t-1}, c_{1:t-1}, y_t, x_t)$ for all possible choices of $c_t = \{1, 2, \ldots, K_t\}$, apply the weights scheme of the RCRP, normalize, and then sample from the resulting discrete distribution.

Line 7-8: Importance Distribution for new cluster $q_1(\theta_{k,t}|y_t, x_t, c_t = k)$

The true distribution of the parameters $\theta_{k,t}$ for a new cluster, given a newly associated data point y_t, is given by

$$P(\theta_{k,t}|y_t, x_t, c_t = k) = \frac{P(y_t|\theta_{k,t}, x_t)\mathbb{G}_0(\theta_{k,t})}{\int P(y_t|\theta_{k,t}, x_t)\mathbb{G}_0(\theta_{k,t})d\theta_{k,t}}$$

which, due to conjugacy, can be evaluated analytically. Therefore we can set the importance distribution as the true distribution of $\theta_{k,t}$:

$$q_1(\theta_{k,t}|y_t, x_t, c_t = k) \stackrel{def}{=} P(\theta_{k,t}|y_t, x_t, c_t = k)$$

Line 11-12: Importance Distribution for old cluster $q_2(\theta_{k,t}|\theta_{k,t-1}, y_t, x_t, c_t = k)$

When an old cluster is sampled, the distribution of the parameters at time t depends on the old parameters at time $t-1$, through the transition kernel, and the data point, through the likelihood:

$$P(\theta_{k,t}|\theta_{k,t-1}, y_t, x_t, c_t = k) = \frac{P(\theta_{k,t}|\theta_{k,t-1})P(y_t|\theta_{k,t}, x_t)}{P(\theta_{k,t}|\theta_{k,t-1})P(y_t|\theta_{k,t}, x_t)d\theta_{k,t}} \tag{10}$$

Since we adopt a Metropolis kernel it is not obvious how to sample from the distribution in (10), therefore we have to use an approximation for the importance distribution $q_2(\cdot)$. A possible choice is

$$q_2(\theta_{k,t}|\theta_{k,t-1}, x_t, y_t, c_t = k) \stackrel{def}{=} P(\theta_{k,t}|\theta_{k,t-1})$$

i.e. the Metropolis kernel as in (4).

Line 9, 13 and 15: Weights

Due to the recursive nature of the algorithm and since a particle is characterized by a cluster and a set of parameters, the weights at time t are given by

$$\begin{aligned}
w_t \quad = \quad & w_{t-1} && (11)\\
& \times \frac{P(c_{t,i} = k|c_{t-1,1:K_{t-1}}, c_{t,1:i-1})}{P(c_{t,i} = k|\theta_{t-1}, c_{1:t-1}, y_t, x_t)} && (12)\\
& \times \frac{\mathbb{G}_0(\theta_{c_t,t})}{q_1(\theta_{c_t,t}|y_t)} \mathbb{1}_{k=K_t} && (13)\\
& \times \frac{P(\theta_{c_t,t}|\theta_{c_t,t-1})}{q_2(\theta_{c_t,t}|y_t)} \mathbb{1}_{k \le K_{t-1}} && (14)\\
& \times P(y_t|x_t, c_t = k, \theta_{k,t}) && (15)
\end{aligned}$$

where (11) is the weight at time $t-1$, (12) is the ratio between the conditional probability as in (1) and the probability associated with the sampled cluster from the discrete distribution, (13) is the ratio between the prior and the importance distribution for the vector of parameters $\theta_{c_t,t}$ if a new cluster was sampled, (14) is the ratio if the cluster is an old one, but since the choice for the importance distribution is exactly the transition kernel, this ratio simplifies to 1, and (15) is the likelihood.

Line 20: Resampling

The resampling step is needed since the weights of the particle filter tend to skew, i.e. after few iterations only a small number of particles presents high weights whereas most of the particles have weights close to zero, which implies that the algorithm tries to reproduce the target distribution with only a small subsample of the particles.
There are different resampling strategies: multinomial resampling, residual resampling, stratified resampling or systematic resampling. In this work we have implemented the multinomial resampling, in which the particle duplication N_i counts are drawn from a multinomial distribution $N_i \sim Mult(N; w^{(1)}, \dots, w^{(N)})$. The indices $i_1, \dots, i_N$ for the new set of particles can then easily be obtained from these counts.

5 Empirical Analysis: Inflation and GDP

5.1 Initialization and fixed coefficients

The base distribution is defined as:

$$\mathbb{G}_0(\Psi_{k,t}, \lambda_{k,t}) = \mathcal{N}(\Psi_{k,t}|\Psi_0, n_0\lambda_{k,t})\Gamma(\lambda_{k,t}|a, b)$$

where Ψ_0 is the OLS estimation for the vector of coefficients Ψ, a and b are such that the expected value, i.e. ab^{-1}, is equal to the OLS estimation for the variance of the time series, whereas n_0 is such that $n_0\lambda_{k,t}$ is a value that ensures the algorithm to span the parameters space.
The jump in the transition kernel driven by the parameter s in (5) is chosen assuming that inside the same cluster, i.e. when the transition kernel takes place, the parameters are closed to each other, in other words the difference between the parameters at time $t-1$ and t has to be small if the two observations belong to the same regime. The number[5] of particles, N, is fixed at 15000. The latter parameter to choose is α, the concentration parameter of the DPM, that reflects the believe that a new cluster could appear: the larger α is, the smaller the variance and the DP will concentrate more of its mass around the mean. Moreover this parameter controls the number of cluster in a direct manner, indeed a larger α implying a larger number of cluster a priori. We assume a value such that even in the worst scenario, i.e. all the particle at time $t-1$ and all the particles, up to the last, at time t belong to the same cluster, there exists a probability non equal to zero that a new cluster appears, given by $\frac{\alpha}{\alpha+N+N} = 0.0015$.

5.2 Inflation

In this section we are going to present the main results for the inflation rate, as PCE deflator, for the period[6] $1957Q4 : 2011Q3$, expressed as an annual rate. Following KP, we assume an AR(2) model with constant for the inflation series. Figure 2 shows the actual data, the time series generated by the standard OLS for an AR(2) model, and the series obtained with the evolutionary clusters approach. The methodology based on clusters presents a good performance in reproducing

[5] Until now the number of particles is limited by the amount of ram needed to run the algorithm and save all the important informations.

[6] With respect to the figure 1(b) the studied temporal horizon skips the first ten year, i.e. 1947-1957, because the high volatility that characterized the years after the WWII is quite difficult to explain.

the actual data. In particular, comparing to the performance of a standard OLS we obtain a better fit for the period of high level and volatility of inflation, i.e. 1970-1982.

Figure 2: Percentage inflation rate: actual data (green), estimated series via OLS (blue) and estimated series via evolutionary cluster methodology (red).

At the MAP, the procedure identifies 13 clusters that are highlight with different color over the original series in figure 3. It is interesting note that it is possible to interpret the clusterization in terms of economic history and stylized facts. For example a new cluster is generated in proximity both Volcker and Greenspan appointment dates as chairman, a new cluster corresponds to the 'standard' starting date for the Great Moderation period, i.e. 1983, and another one after the end of the 1992 recession. It is important to note that the possibility of interpret the clusters, due to a relative small number that however it is not imposed a priori, it is really an advantage, in particular with respect to the result of KP, who estimated 124 regimes losing any possible interpretations.

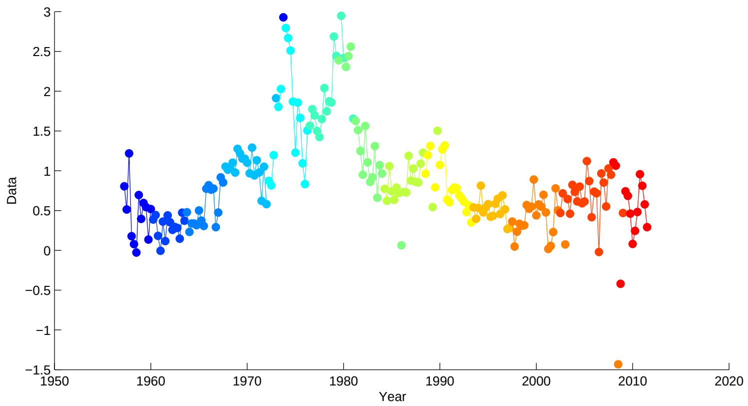


Figure 3: Percentage inflation rate divided in the estimated clusters.

Let now consider the estimated coefficients. Give the MAP for the sequence of clusters that evolves over time, each observation y_t belongs to a cluster and it is characterized by a distribution for the parameters. We summarized the informations relative to the parameters with the cluster's centroid.

Let us analyze the precision parameter, λ, in figure 4.

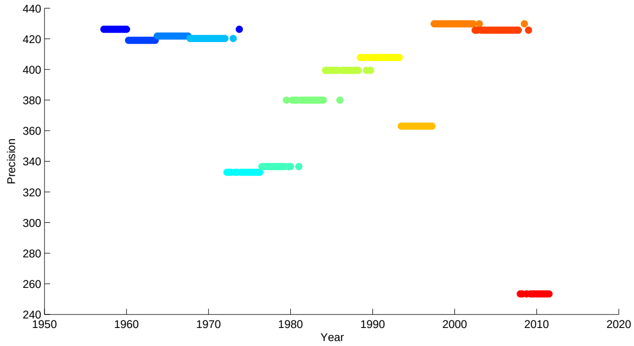


Figure 4: Centroids for the estimated clusters: the precision parameter λ.

The cluster procedure is able to reproduce the low volatility characterizing the $'60s$, the suddenly increase in $'70s$ and the decreasing path due to the Great Moderation after 1983. Moreover, the algorithm is quite sensible to the increase in volatility in 1992, after the crisis of $1990-1991$, and in the last years, due to the recent recession.
In figure 5 are plotted the values of the clusters' centroid for the coefficient of the first lag in the AR(2) model.

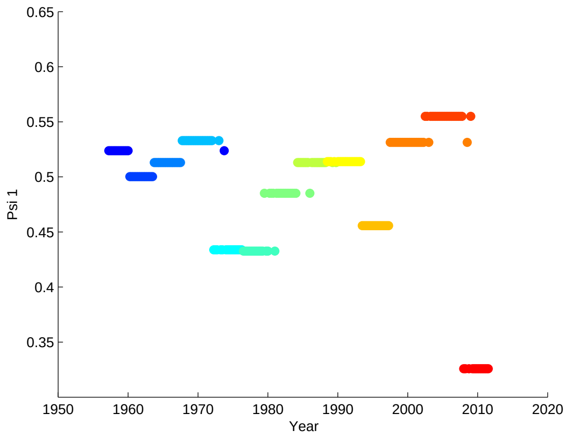


Figure 5: Centroids for the estimated clusters: the coefficient associated to the first lag in the AR(2) ψ_1.

It is interesting note that the values of this parameters during the first ten years of Great Moderations are stable even if they belong to two different clusters, whereas there is a significant difference with the values characterizing the $'70s$, where, as well as for the Great Moderation years, we observe that the estimations for ψ_1 are homogeneous across the two clusters that identify these years. A possible interpretation for this consistency within periods, i.e. the $'70s$ and from 1983 to 1992, and heterogeneity between periods, is that around 1983 there was a change in terms of policy effort devoted to maintain the inflation at a stable and low level.
With regards to the other two autoregressive coefficients ψ_0 and ψ_2, the posteriors highlight small changes between clusters.
Stock and Watson (2003) find for the implicit price deflator for the personal consumption expenditure (GDC) a break in both conditional mean and conditional

variance in the first years of the $'70s$. Even in this analysis we obtain a cluster's change for that period, but since we do not impose a single breaks, we recover other important changes, as the one in the early $'80s$, that Stock and Watson miss. Indeed this comparison is a good example of the limits of a single break approach.
Considering the debate on Bad Luck or Bad Policy, even if the analysis is not implemented neither in a VAR nor in a structural form and therefore we can not address directly the question on the origin of the Great Moderation, nevertheless this result could be an indicator that the Great Inflation period has to be interpreted in terms of both Bad Policy and Bad Luck. In fact we can observe an important change in volatility, i.e. the source of Great Inflation was Bad Luck before $'84$, as well as in coefficients, i.e. the source of Great Inflation was a Bad Policy.

5.3 GDP

In this section we are going to present the main results for the GDP growth (percentage), for the period $1957Q4 : 2011Q3$, expressed as an annual rate. In figure 6 there is the time series for GDP colored for the 5 clusters estimated at the MAP. As well as for the inflation, it is possible to interpret the clusters in terms of economic history. The first cluster covers the $'60s$, the second contains the 'Great Inflation' period, whereas the third one starts around 1983 with the 'Great Moderation', the fourth cluster appears after the end of the 1992 recession, whereas the last regime starts after the end of the recession of 2001-2002. Comparing my results with KP, as previously pointed out, the number of regimes that we obtain is more intelligible, in fact KP found a posterior mean number of regimes equal to 45. Nevertheless it is interesting to observe that in both analyses there is a consistent reduction in the number of regimes with respect to the result for inflation.
Let us consider the parameters that identify the clusters' centroid, starting with the precision parameter. It is possible to observe a continuous reduction in variance[7], as shown in figure 7.

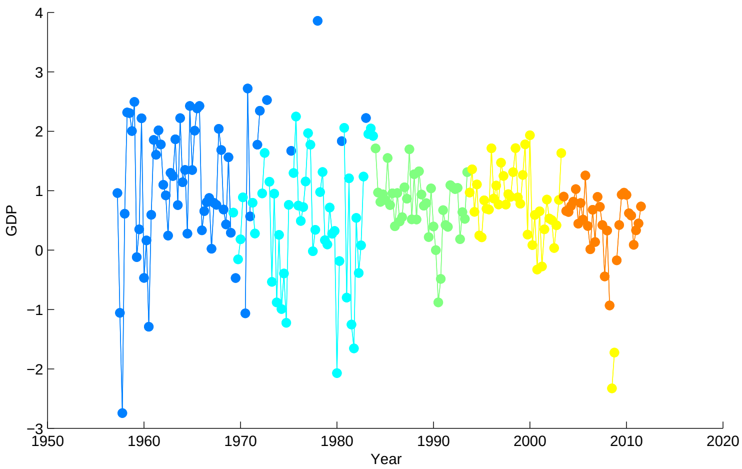


Figure 6: Percentage GDP growth divided in the estimated clusters.

[7]We transform λ in order to compare these results with Koop and Potter (2007), multiplying $1/\lambda$ by 100 and taking the exponential.

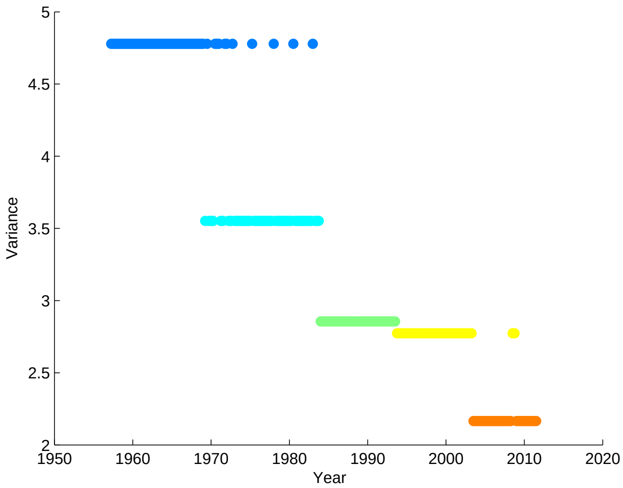
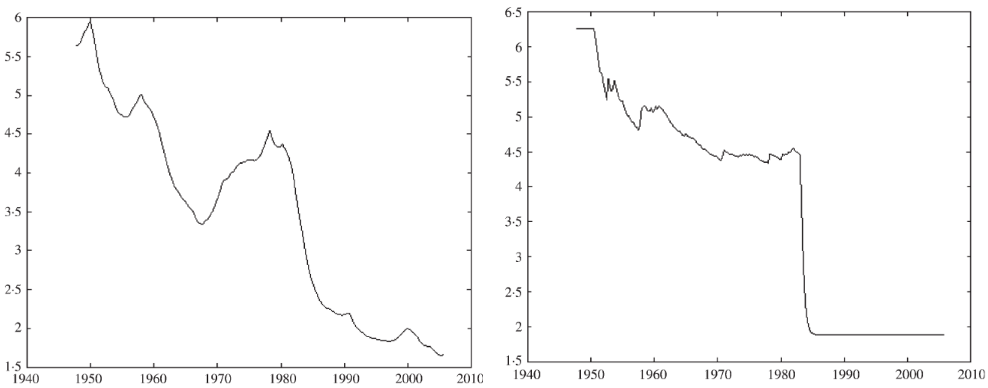


Figure 7: Transformed variance, considering the centroids for the estimated clusters.

We can compare this result with the estimations obtained for the TVP and the one-break model for real GDP growth, presented in KP: indeed in figure 8 there are the graphs for the posterior mean volatility of GDP under the two alternative models as in Koop and Potter (2007). Both models indicate that volatility is decreasing substantially over time, with a particularly dramatic drop occurring around 1984. However, with the TVP model this decline is much more smooth and non-monotonic than with the one-break model.



(a) (b)

Figure 8: Posterior mean for GDP variance: TVP in 8(a) and one break in 8(b), from Koop and Potter (2007)

.

The cluster approach remarks less then the two alternative approaches the volatility decline in 1984. It is interesting note that the changes in clusters, obtained with the evolutionary cluster methodology, correspond to the changes in the slope for the graph of the estimated variance via TVP as in figure 8(a): the first cluster catches the first decline in volatility between 1957 and 1968, the second cluster corresponds to the period of increase in volatility during the $'70s$, that nevertheless implies a lower level of variance than during the $'60s$, and the third, fourth and fifth groups are related to the reduction in volatility experimented from

1984, with two little waves after 1990 and 2000.
Let now consider the values for ψ_1, plotted in figure 9, where it is possible to observe a small non-monotonic reduction over the sample. Note that the changes in coefficients for GDP are lower than what it is experimented in inflation. This result is supported by others papers: for example KP find small changes in coefficients under both models TVP and one-break. The results for ψ_0 and ψ_2 are similar to what observed for ψ_1: there is a low volatility of the parameters between clusters and the impact of both the constant and the observation at time $t-2$ decreases during the years.
This small change for the coefficients, opposing to the remarkable change in volatility, is stressed in Stock and Watson (2003) with a similar but stronger results: the authors fail to reject the null hypothesis of no break in the coefficients of the conditional mean, whereas the null hypothesis of no break in the conditional variance is rejected at the 1% significance level, with a estimate break date in 1983.
Concluding, in order to formulate possible explanation on the source of Great Moderation, a structural economic model is needed, since the reduction of GDP variance, mostly due to changes in the precision parameter, could arise either for reductions in the variance of certain structural shocks or from changes in the structure of monetary policy.

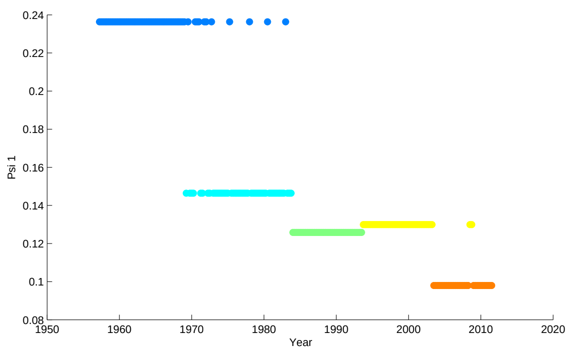


Figure 9: Centroids for the estimated clusters of ψ_1

.

6 Conclusion

In this study we have presented a new methodology in order to analyze the instability in time series, applying the idea of evolutionary cluster to the study of output and inflation for US after War World II. Based on the Recurrent Chinese Restaurant Process, we have specified a model for an autoregressive process and estimated it via particle filter using a conjugate prior.
This procedure displays some advantages, in particular does not require a strong ex ante structure in order to neither detect the breaks nor manage the parameters' evolutions. Whereas the main drawback is that this methodology is based on a descriptive approach, thus it can not be use for forecast.
The application of the evolutionary cluster procedure to GDP growth and inflation rate for US from 1957 to 2011 generate estimations which well track the data. Moreover this methodology produces a clusterization of the time series that can be interpreted in terms of economic history and it is able to recover key data features without making restrictive assumptions, as in 'one-break' or TVP models.
This analysis supports the findings of Koop and Potter (2007), such that a model

between the one-break and TVP models is most sensible and successfully capturing the properties of a reasonable data generating process. Therefore, as in KP we let the data tell us what the key properties of the data are, rather than assuming them. The empirical results compared with Stock and Watson (2003) also show the problems of working with models with a small number of breaks when, in reality, the evolution of parameters is much more gradual. Moreover with respect to KP, the proposed methodology stress the idea that parameters associated to observations that belong to the same regime have a different relationship then parameters that generate observations of different clusters, and the results on the number of regimes is more intelligible under the presented study.

Considering the open debate on the source of the Great Moderation, under the caveat that until now we have not studied a VAR or a structural form, this approach presents conclusions that support the findings in Cogley and Sargent (2001, 2005), suggesting changes in both volatility and coefficients before and after 1984, even if the changes in coefficients are less marked.

We leave to future research the implementation of this approach to a structural model in order to address directly the question on Great Inflation period.

References

Ahmed, Amr, and Eric P. Xing. 2008. "Dynamic Non-Parametric Mixture Models and the Recurrent Chinese Restaurant Process." *Proceedings of The Eighth SIAM International Conference on Data Mining*, 33–40.

Blanchard, Olivier, and John Simon. 2001. "The Long and Large Decline in U.S. Output Volatility." *Brookings Papers on Economic Activity*, 32(1): 135–174.

Caron, Francois, Manuel Davy, and Arnaud Doucet. 2007. "Generalized Polya Urn for Time-varying Dirichlet Process Mixtures." *UAI*, 33–40.

Cogley, Timothy, and Thomas J. Sargent. 2001. "Evolving Post-World War II U.S. Inflation Dynamics." *NBER Macroeconomics Annual*, 16: 331–388.

Cogley, Timothy, and Thomas J. Sargent. 2005. "Drift and Volatilities: Monetary Policies and Outcomes in the Post World War 2 U.S." *Review of Economic Dynamics*, 8(2): 262–302.

Griffin, J.E., and M.F.J. Steel. 2006. "Order-Based Dependent Dirichlet Processes." *Journal of the American Statistical Association*, 101: 179–194.

Kim, Chang-Jin, and Charles R. Nelson. 1999. "Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle." *The Review of Economics and Statistics*, 81(4): 608–616.

Kim, Chang-Jin, Charles R Nelson, and Jeremy Piger. 2004. "The Less-Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations." *Journal of Business & Economic Statistics*, 22(1): 80–93.

Koop, Gary, and Simon Potter. 2007. "A flexible approach to parametric inference in nonlinear time series models."

MacEachern, Steve N. 2000. "Dependent Dirichlet Processes." *Technical Report, Department of Statistics, The Ohio State University.*

Pitt, Michael K., and Stephen G. Walker. 2005. "Constructing Stationary Time Series Models Using Auxiliary Variables With Applications." *Journal of the American Statistical Association*, 100(470): 554–56.

Primiceri, Giorgio. 2005. "Why Inflation Rose and Fell: Policymakers' Beliefs and US Postwar Stabilization Policy." *NBER Working Papers*, , (11147).

Sims, Christopher A., and Tao Zha. 2006. "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96(1): 54–81.

Stock, James H. 2001. "Comment on Cogley and Sargent." *NBER Macroeconomic Annual.*

Stock, James H., and Mark W. Watson. 2003. "Has the business cycle changed?" *Proceedings*, 9–56.