Time-Consistency Problem and the Behavior of US Inflation from 1970 to 2008

Nima Nonejad

CREATES Research Paper 2013-25

Time-Consistency Problem and the Behavior of US Inflation from 1960 to 2013

Nima Nonejad[†][*]

Aarhus University and CREATES

Abstract

Restrictions implied by the theory of time-consistent monetary policy are imposed on data. Model estimation is conducted using Bayesian MCMC techniques. We identify a major change in the policy of the Federal Reserve with regards to inflation stabilization. In fact, since the beginning of 1982 the Federal Reserve places more weight on its goals for inflation stabilization. Furthermore, volatility of the shocks that affect the fundamentals and the correlation between these shocks has fallen since the early 1980s.

Keywords: Monetary policy, Structural breaks, Time-consistency
(*JEL:* C11, C22, C51, E42, E52)

1 Introduction

Movements in the quarter-to-quarter inflation rate in the United States from 1960 to 2013 are remarkably different. First, a period of rising inflation from the 1960s to the early 1980s, thereafter a period of falling inflation from mid 1980s to the end of the 1990s and again a period of slight rising inflation from the mid 2000s. In their seminal paper Barro and Gordon (1983) explain that the activist, discretionary monetary policy can have contributed to the high rates of inflation during the 1970s and early 1980s.

Ireland (1999) performs a detailed econometric analysis using the Barro and Gordon (1983) model and tests the hypothesis that the time-consistency problem may explain the behavior of the inflation rate in the United States from 1960 to 2000. In this paper we extend the main model of Ireland (1999). Specifically, we try to answer the following question: can we build on Ireland (1999) by allowing for time-variation in the key policy parameters of the Federal Reserve and thus obtain more information regarding policy decisions at different points in time? Furthermore, to distinguish between policy changes and external shocks, exogenous macroeconomic parameters are also subject to structural breaks.

Ideally, we should be able to detect a structural break in the beginning of the 1980s. Furthermore, we would also expect lower volatility of the shocks that affect the macroeconomic fundamentals such as the Philips curve and changes in the natural level of unemployment since the beginning of the 1980s. From a computational point of view this extension also allows us to demonstrate the flexibility of Markov chain Monte Carlo (MCMC) methods to adapt to more complicated model structures.

Estimation results are presented in section (2.2). Section (3) concludes. Appendices at the end of the paper present the theoretical framework of Ireland (1999), details about the MCMC algorithm, DIC computation and prior sensitivity analysis.

2 Methodology

The main equations that form the basis of Ireland (1999) are

$$\pi_t - \alpha A U_t = \left(1 + \alpha^2 A\right) \eta_t - \alpha A \varepsilon_t \tag{2.1}$$

$$\triangle U_t = \triangle U_t^n - \alpha \eta_t + \alpha \eta_{t-1} \tag{2.2}$$

π_t and U_t denote the inflation and unemployment rate respectively. α is the slope of the expectational Phillips curve. $A = (1-k)/b > 0$ where $b > 0$ and $0 < k < 1$ are the weights that the Fed attaches to its loss function, $L_t = (1/2)(U_t - kU_t^n)^2 + (b/2)\pi_t^2$. Thus, any deviation of inflation

(unemployment) from its target value generates increasing marginal disutility in L_t. U_t^n is the natural rate of unemployment, $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$, $\eta_t \sim N\left(0, \sigma_\eta^2\right)$ with $E\left[\varepsilon_t \eta_t\right] = \sigma_{\varepsilon\eta}$. We can write (2.1) and (2.2) in state space form as

$$
y_t = \begin{bmatrix} \pi_t - \alpha A U_t \\ \triangle U_t \end{bmatrix} = H\beta_t = \begin{bmatrix} 0 & -\alpha A & 1+\alpha^2 A & 0 \\ 1 & 0 & -\alpha & \alpha \end{bmatrix} \begin{bmatrix} \triangle U_t^n \\ \varepsilon_t \\ \eta_t \\ \eta_{t-1} \end{bmatrix}
\tag{2.3}
$$

and

$$
\beta_t = \begin{bmatrix} \triangle U_t^n \\ \varepsilon_t \\ \eta_t \\ \eta_{t-1} \end{bmatrix} = F\beta_{t-1} + K\varepsilon_t = \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \triangle U_{t-1}^n \\ \varepsilon_{t-1} \\ \eta_{t-1} \\ \eta_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}
\tag{2.4}
$$

Allowing for structural breaks in (2.3)-(2.4) is straightforward.

Assume that there are $m-1$, $m \in \{1, 2, ...\}$ change points at unknown times, $\Omega_m = \{\tau_1, \tau_2, ..., \tau_{m-1}\}$. Separated by these change points, there are m different regimes. The density of y_t, $t = 1, ..., T$ depends on $\theta_k = \left(\alpha_k, b_k, k_k, \lambda_k, \sigma_{\varepsilon k}, \sigma_{\eta k}, \sigma_{\varepsilon\eta k}\right)'$ where

$$
\theta = \begin{cases} \theta_1 & \text{if} & t < \tau_1 \\ \theta_2 & \text{if} & \tau_1 \leq t < \tau_2 \\ \vdots & \vdots & \vdots \\ \theta_{m-1} & \text{if} & \tau_{m-2} \leq t < \tau_{m-1} \\ \theta_m & \text{if} & \tau_{m-1} \leq t \end{cases}
\tag{2.5}
$$

Let $S = (s_1, ..., s_T)'$ where $s_t = k$ indicates that y_t is from regime k. The one-step-ahead transition probability matrix for s_t is given as

$$
P = \begin{bmatrix} p_{11} & p_{12} & 0 & \cdots & & 0 \\ 0 & p_{22} & p_{23} & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & 0 & p_{m-1,m-1} & p_{m-1,m} \\ 0 & 0 & \cdots & & 0 & 1 \end{bmatrix}
\tag{2.6}
$$

where $p_{lk} = p\left(s_t = k \mid s_{t-1} = l\right)$ with $k = l$ or $k = l+1$ is the probability of moving from regime l at time $t-1$ to regime k at time t. P ensures that given $s_t = k$ at time t, the next period, $t+1$,

3

$s_t + 1$ remains in the same state or jumps to the next state. Once the last regime is reached, one stays there forever, that is $p_{m,m} = 1$. Testing for the dimension of (2.6) is then a test for the number of structural breaks.

2.1 Bayesian Estimation

To conduct model estimation we jointly estimate $\theta = \{\theta_k\}_{k=1}^m$, S and P. Although the joint posterior density of the model, $p(P, \theta, S \mid Y_T)$ is not a well-known density, samples from it can be obtained using Gibbs sampling and Metropolis-Hastings (M-H), see Appendix (B.1) for more details. The parameters are divided into $m + 2$ blocks: $\theta_1, ..., \theta_m$, S and P. Choose a set of starting values: $\theta_1^{(0)}, ..., \theta_m^{(0)}$ and $P^{(0)}$. For $j = 1, ..., N$ iterate from

- $S^{(j)} \mid P^{(j-1)}, \theta^{(j-1)}, Y_T$

- $\theta_k^{(j)} \mid S^{(j)}, Y_T$ for $k = 1, ..., m$

- $P^{(j)} \mid S^{(j)}$

- $j = j + 1$

In step 1, we use the algorithm of Chib (1998) to draw $S^{(j)}$. $\theta_k^{(j)}$ is sampled all-at-once using Metropolis-Hastings conditional on $S^{(j)}$ and Y_T. Finally, $p_{kk}^{(j)} \mid S^{(j)} \sim Beta\left(a_0 + n_{kk}^{(j)}, b_0 + 1\right)$, $k = 1, ..., m - 1$ where $n_{kk}^{(j)}$ is the number of one-step transitions from state k to state k in $S^{(j)}$.

2.2 Results

The inflation rate is measured using yearly percentage changes in the quarterly seasonally adjusted GDP implicit price deflator from 1960q1 to 2013q4. For the unemployment rate we use the civilian unemployment rate also from 1960q1 to 2013q4. Both data series are downloadable from FRED.

We estimate (2.3) and (2.4) for $m = 0, 1, 2$. For each specification, the MCMC algorithm consists of $N = 5000$ iterations with a burn-in of 1000 assuming the following priors:

$$\alpha_k \sim TN_{]0,\infty[}(0, 100) \quad , \quad b_k \sim TN_{]0,\infty[}(0, 100)$$
$$k_k \sim TN_{]0,1[}(0, 100) \quad , \quad \lambda_k \sim TN_{]-1,1[}$$
$$\sigma_{\varepsilon k} \sim N(0, 100) \quad , \quad \sigma_{\eta k} \sim N(0, 100)$$
$$\sigma_{\varepsilon \eta k} \sim N(0, 100) \quad , \quad p_{kk} \sim Beta(8, 0.1)$$

where TN refers to the truncated Normal distribution on the domain $[., .]$. In this setting, the prior for each element of θ_k is very non-informative, while the prior for p_{kk} favors infrequent structural

Model	no-break					break				
Parameter	mean	std. dev	95% credibility	R_B	Geweke	mean	std. dev	95% credibility	R_B	Geweke
$\bar{\alpha}_1$	0.322	0.009	[0.307,0.337]	2.800	0.237	0.361	0.011	[0.343,0.378]	9.292	1.368
$\bar{\alpha}_2$						1.786	0.010	[1.769,1.803]	2.560	0.419
$\bar{b}_1$	0.159	0.007	[0.147,0.171]	2.554	-2.470	0.118	0.006	[0.107,0.129]	4.574	1.598
$\bar{b}_2$						1.149	0.011	[1.130,1.166]	6.939	-0.105
$\bar{k}_1$	0.719	0.009	[0.705,0.733]	2.606	0.172	0.708	0.009	[0.693,0.725]	3.029	-0.281
$\bar{k}_2$						0.730	0.006	[0.719,0.741]	1.868	0.054
$\bar{\lambda}_1$	0.342	0.010	[0.326, 0.359]	3.137	0.766	0.214	0.014	[0.193,0.236]	12.004	-1.045
$\bar{\lambda}_2$						0.551	0.015	[0.525,0.576]	15.536	0.374
$\bar{\sigma}_{\varepsilon 1}$	0.296	0.008	[0.283,0.309]	2.418	-0.582	0.382	0.009	[0.367,0.398]	3.496	-1.168
$\bar{\sigma}_{\varepsilon 2}$						0.441	0.009	[0.426,0.457]	4.493	0.372
$\bar{\sigma}_{\eta 1}$	1.976	0.009	[1.960,1.991]	2.635	-1.126	1.818	0.014	[1.800,1.836]	11.726	-1.359
$\bar{\sigma}_{\eta 2}$						0.672	0.010	[0.655,0.688]	2.332	-0.182
$\bar{\sigma}_{\varepsilon \eta 1}$	0.544	0.009	[0.528,0.561]	2.714	-0.039	0.647	0.013	[0.624,0.670]	11.369	0.941
$\bar{\sigma}_{\varepsilon \eta 2}$						0.083	0.009	[0.067,0.099]	1.831	-0.093
$\bar{p}_{11}$						0.991	0.009	[0.971,0.999]	1.116	-0.075
Loglikelihood	-495.969					-472.437				
DIC	993.104					947.980				
M-H-ratio, $m=1$	0.526					0.522				
M-H-ratio, $m=2$						0.585				

Table 2.1: Posterior mean (mean), standard deviation (std. dev), 95% credibility intervals, inefficiency factors, R_B, Geweke's convergence statistics, loglikelihood, $\log p\left(Y_T \mid \theta\right)$, DIC and M-H acceptance ratios for no-break and the break model. The parameters associated with each regime are labeled with subscript $1,\ldots,m$ and $\bar{x} = \Sigma_{j=1}^{N} x^{(j)}$.

breaks. In Appendix (B.3) we evaluate sensitivity of the results to different prior specifications. Overall, we find that evidence in favor of structural breaks is quite strong and the choice of prior hyperparameters is of limited importance.

Results suggest existence of one change point according to DIC. The DIC of (2.3)-(2.4) (no-break) is 993.104 while the DIC for its one change-point alternative (break) is 947.980. We also estimate (2.3)-(2.4) conditional on two change points and obtain a DIC of 972.499. Table 2.1 reports some summary statistics concerning the posterior distribution of the parameters and some diagnostics. Figure 2.1 displays the data, posterior density of the change-point date and the cumulative probability of the change point for (2.3)-(2.4) conditional on one change point. The posterior mode of $\left\{S^{(j)}\right\}_{j=1}^{N}$ is associated with 1982q2.

Let $\bar{b}_k$ denote the posterior estimate of b in regime k. For the break model we see that $\bar{b}_k$'s $k = 1, 2$ are distinctly different with $\bar{b}_2 = 1.149 > \bar{b}_1 = 0.118$. More importantly, the period during the chairmanships of William M. Martin, Arthur Frank Burns and George William Miller which covers from 1960 to 1979 is clearly associated with $\bar{b}_1$. Thereafter, results clearly reveal a switch towards $\bar{b}_2$ shortly after the appointment of Paul Volcker as chairman of the Fed and throughout the chairmanships of Alan Greenspan and Ben Bernanke. On the other hand, estimates of k do not differ across regimes or models. Furthermore, estimates of $k \approx 0.8$ are also close to those obtained by Primiceri (2006).

With regards to the other parameters, we find that $\bar{\alpha}_2$ is significantly flatter than $\bar{\alpha}_1$. The estimates of λ are positive, as expected. However, $\bar{\lambda}_2$ shows a slight more persistence in $\triangle U_t^n$ since 1982q3. Furthermore, estimates of σ_η show a decrease in volatility of the shocks that affect π_t and thus contributing to a more stable evolution in π_t since 1982q3. Finally, positive estimates of $\sigma_{\varepsilon\eta}$ indicate that unfavorable shocks to U_t^n tend to coincide with unfavorable shocks to π_t.

According to these results the increase in the rate of inflation in the 1960s and 1970s can also partially be explained by a string of "bad luck", in form of higher values of σ_ε and σ_η along with stronger correlation between ε_t and η_t. Specifically, we find that the estimated correlation coefficient, $\bar{\rho}_{\varepsilon\eta}$ is 0.93 for $m = 1$ while on the other hand $\bar{\rho}_{\varepsilon\eta}$ is 0.28 for $m = 2$.

Overall, results in Table 2.1 and Figure 2.1 indicate that compared to the rest of the sample the Fed is focusing more on inflation stabilization (indicated by $\bar{b}_2 > \bar{b}_1$) since the early 1980s. Furthermore, we can cautiously infer that the hypothesis that the disinflation since this period is solely the result of "good luck" in form of lower volatility of the shocks that affect the fundamentals and the correlation between these shocks can be ruled out. Rather, this disinflation can also be attributed to policy changes at the Fed. However, as noted by Ireland (1999) the version of Barro and Gordon's model that we have used, is extremely simple and does not account for many important factors.
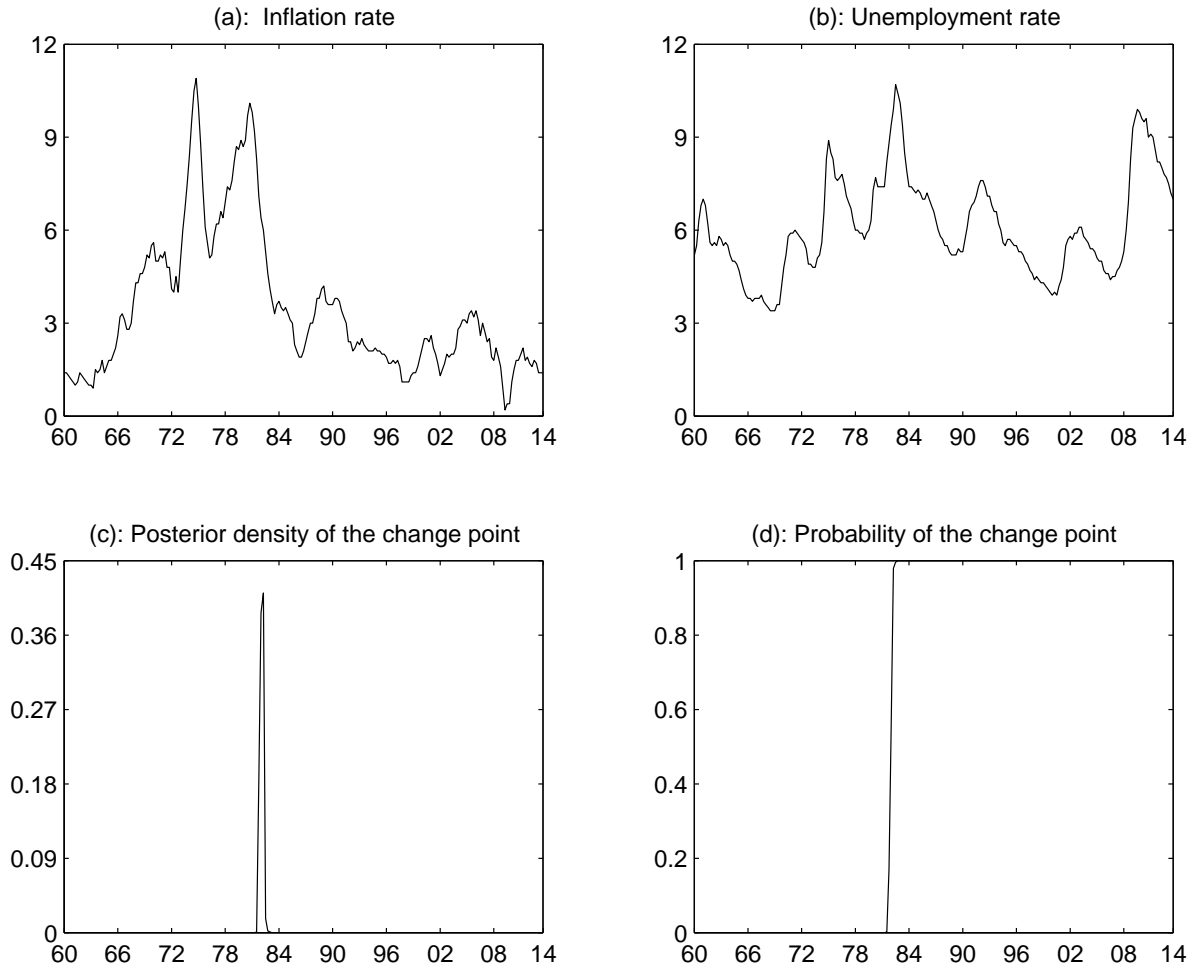
Figure 2.1: MCMC estimates (break model)

3 Conclusion

Overall, this paper provides some interesting results. Parameter estimates indicate that compared to the rest of the sample the Fed is relatively focusing more on inflation stabilization since the early 1980s. Furthermore, volatility of the shocks that affect the fundamentals has also decreased in this period.

From a computational point of view by using basic tools such as Gibbs sampling and Metropolis-Hastings we show that one is able to extensively build on an already well-known model and thus obtain better interpretation of important economic phenomena. Specifically, we show that by allowing for time-variation in policy variables as well exogenous parameters using the conditioning features of MCMC we can obtain an in-depth understanding of the dynamics of US inflation using a well-known macroeconomic framework. Hence, these methods can be applied to more complicated models such as Primiceri (2006) which we leave for future research.

References

[1] Barro, R. J., and D. B. Gordon. 1983. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy 91*: 589-610.

[2] Chib, S., and E. Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician 49(4)*: 327-335.

[3] Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics 86*: 221-24

[4] Hamilton, J. D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica 57(2)*: 357-384.

[5] Ireland, P. N. 1999. "Does The Time-Consistency Problem Explain the Behavior of Inflation in the United States?" *Journal of Monetary Economics 44*: 279-291.

[6] Kim, J. C., and C. R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications.* MIT press.

[7] Koop, G. 2003. *Bayesian Econometrics*. John Wiley & Sons Ltd.

[8] Primiceri, G. 2006. "Why Inflation Rose and Fell: Policymaker's Beliefs and US Postwar Stabilization Policy." *The Quarterly Journal of Economics 121*: 867-901.

[9] Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde. 2002. "Bayesian Measures of Model Complexity and Fit (with comments)". *Journal of the Royal Statistical Society B 64*: 583-639.

A Appendix

The model of Ireland (1999) is as follows: let U_t denote the unemployment rate which fluctuates around the unobserved natural rate, U_t^n in response to deviations of the actual inflation rate, π_t from the expected rate, π_t^e

$$U_t - U_t^n = -\alpha \left(\pi_t - \pi_t^e \right) \tag{A.1}$$

where $\alpha > 0$. The natural rate, in turn, fluctuates over time in response to real shocks according to the autoregressive process

$$U_t^n - U_{t-1}^n = \lambda \left(U_{t-1}^n - U_{t-2}^n \right) + \varepsilon_t, \quad |\lambda| < 1, \quad \varepsilon_t \sim N \left(0, \sigma_\varepsilon^2 \right) \tag{A.2}$$

At the beginning of each time period after private agents have formed their expectation, π_t^e but before realization of ε_t the Fed chooses a planned rate of inflation, π_t^p. Actual inflation for that time period is then determined as the sum of π_t^p and a control error, η_t, such that

$$\pi_t = \pi_t^p + \eta_t \tag{A.3}$$

where $\eta_t \sim N \left(0, \sigma_\eta^2 \right)$ and $E \left[\varepsilon_t \eta_t \right] = \sigma_{\varepsilon\eta}$. The Fed selects π_t^p in order to minimize its loss function, L_t. Following Barro and Gordon (1983) L_t penalizes variations of unemployment rate and inflation around target values $kU_t^n < U_t^n$ and zero

$$L_t = (1/2) \left(U_t - kU_t^n \right)^2 + (b/2) \pi_t^2$$

where $1 > k > 0$ and $b > 0$ are the weights that the Fed attaches to inflation and unemployment rate fluctuations. Using (A.1) and (A.3), the Fed's problem can be written as

$$\min_{\pi_t^p} E_{t-1} \left[(1/2) \left((1-k) U_t^n - \alpha \left(\pi_t^p - \pi_t^e + \eta_t \right) \right)^2 + (b/2) \left(\pi_t^p + \eta_t \right)^2 \right]$$

where E_{t-1} denotes the expectation at the end of period $t-1$. If $k = 1$ then the Fed is just trying to keep unemployment at the natural rate. However, we assume that $0 < k < 1$ which means that the Fed is targeting an unemployment rate that is actually lower than the natural rate of unemployment. Barro Gordon (1983) justify this assumption by appealing to the idea that the government (not the Fed), is trying to push unemployment lower than its natural rate and the Fed considers k as an external constraint. The first-order condition for this problem is

$$\alpha E_{t-1} \left[(1-k) U_t^n - \alpha \left(\pi_t^p - \pi_t^e + \eta_t \right) \right] = b E_{t-1} \left[\pi_t^p + \eta_t \right] \tag{A.4}$$

Private agents know the structure of the economy and understand the Fed's time-consistency problem. Therefore, they correctly anticipate its actions and adjust their inflation expectations by solving out the above optimization problem, such that in equilibrium $\pi_t^e = \pi_t^p$. Using this equilibrium condition, along with the fact that $E_{t-1}[\eta_t] = 0$, equation (A.4) simplifies to

$$\pi_t^p = \alpha A E_{t-1}[U_t^n] \tag{A.5}$$

where $A = (1-k)/b > 0$. Equation (A.5) reveals that the inflationary bias resulting from the Fed's inability to commit depends positively on $E_{t-1}[U_t^n]$. Equations (A.1) and (A.3) imply that

$$U_t = U_t^n - \alpha\eta_t \tag{A.6}$$

while combining (A.2), (A.3) and (A.5) yields

$$\pi_t = \alpha A U_{t-1}^n + \alpha A \lambda \triangle U_{t-1}^n + \eta_t \tag{A.7}$$

where $\triangle U_{t-1}^n = U_{t-1}^n - U_{t-2}^n$. In the same manner, combining (A.2) and (A.6) yields

$$U_t = U_{t-1}^n + \lambda \triangle U_{t-1}^n + \varepsilon_t - \alpha\eta_t \tag{A.8}$$

Equations (A.7) and (A.8) reveal that while both π_t and U_t are nonstationary the linear combination $\pi_t - \alpha A U_t$ is stationary

$$\pi_t - \alpha A U_t = (1 + \alpha^2 A)\eta_t - \alpha A \varepsilon_t \tag{A.9}$$

Furthermore, taking first difference of (A.6) yields

$$\triangle U_t = \triangle U_t^n - \alpha\eta_t + \alpha\eta_{t-1} \tag{A.10}$$

According to (A.9), π_t and U_t ought to be nonstationary but cointegrated. We follow Ireland (1999) and check for unit roots and test if U_t and π_t are cointegrated. In Table A.1, we find that both π_t and U_t are nonstationary. The null hypothesis of no cointegration is also rejected.

Panel A: unit root test	ρ	t-test	M	Z
π_t	0.986	-1.052	8	-1.793
U_t	0.968	-1.819	10	-2.526

Panel B: cointegration test	λ_1	λ_2	cointegrating vector	$LR = -T \ln\left(1 - \lambda_1\right)$
	0.043	0.001	$0.433\pi_t - 0.264U_t$	9.4406

Panel A reports the estimates of ρ, the slope coefficient from a regression variable on a constant and its own lagged value, t-test for $\rho = 1$ and Phillips and Perron's (1988) statistic, Z adjusted to allow for serial correlation in the regression errors. M denotes the truncation lag required to form the Newey-West estimator. Panel B reports the two eigenvalues, λ_1 and λ_2, the estimated cointegrating vector and the likelihood ratio statistic, LR for testing the null hypothesis of no cointegration.

Table A.1: Unit root and cointegration test

B Appendix

B.1 Markov chain Monte Carlo algorithm

In the following let $\theta = \{\theta_k\}_{k=1}^{m}$, $S = (s_1, ..., s_T)'$ and $Y_T = (y_1, ..., y_T)'$. The idea behind MCMC is to simulate a trajectory of a Markov chain $\left\{\theta^{(j)}, S^{(j)}, P^{(j)}\right\}_{j=1}^{N}$ from a given set of starting points, $\left\{\theta^{(0)}, S^{(0)}, P^{(0)}\right\}$, with limiting invariant density, the joint augmented posterior density, $p(\theta, S, P \mid Y_T)$. Once convergence is achieved, the algorithm provides a sample of serially dependent draws for θ, S and P, which can be used for posterior inference. Below more details are provided on each step of the Gibbs sampler.

Step 1: $S \mid \theta, P, Y_T$

Chib (1998) shows that a joint draw of all states can be achieved using

$$p(S \mid \theta, P, Y_T) = p(s_T \mid \theta, P, Y_T) \prod_{t=1}^{T-1} p(s_t \mid s_{t+1}, \theta, P, Y_t) \tag{B.1}$$

in which one samples sequentially from each density on the right-hand-side of (B.1) beginning with $p(s_T \mid \theta, P, Y_T)$, and then $p(s_t \mid s_{t+1}, \theta, P, Y_t)\ t = T-1, ..., 1$. At each step one conditions on the previously drawn state s_{t+1}, until a full draw of S is obtained. The individual densities in (B.1) are obtained based on the following steps:

  (a) Initialization: at $t = 1$, set $p(s_1 = 1 \mid \theta, P, Y_1) = 1$.
  (b) Compute the Hamilton (1989) filter, $p(s_t = k \mid \theta, P, Y_t)$. This involves a prediction and an

update step in which one iterates on the following from $t = 2, ..., T$,

$$p(s_t = k \mid \theta, P, Y_{t-1}) = \sum_{l=k-1}^{k} p(s_{t-1} = l \mid \theta, P, Y_{t-1}) p_{lk}, \quad k = 1, ..., m \qquad (B.2)$$

$$p(s_t = k \mid \theta, P, Y_t) = \frac{p(s_t = k \mid \theta, P, Y_{t-1}) p(y_t \mid \theta, Y_{t-1}, s_t = k)}{\sum_{l=1}^{m} p(s_t = l \mid \theta, P, Y_{t-1}) p(y_t \mid \theta, , Y_{t-1}, s_t = l)},$$
$$k = 1, ..., m$$

The last equation is obtained from Bayes' rule. Note that in (B.2) the summation is only from $k-1$ to k, due to the restricted nature of the transition matrix and $p(y_t \mid \theta, , Y_{t-1}, s_t = k)$ is obtained by applying Kalman recursions.

(c) Finally, Chib (1998) shows that the individual densities in (B.1) are

$$p(s_t \mid s_{t+1}, \theta, P, Y_t) \propto p(s_t \mid \theta, P, Y_t) p(s_{t+1} \mid s_t, P)$$

Thus, given $s_T = m$, s_t is drawn backwards over $t = T - 1, T - 2, ..., 2$ as

$$s_t \mid s_{t+1}, \theta, P, Y_t = \begin{cases} s_{t+1} & \text{with probability} \quad c_t \\ s_{t+1} - 1 & \text{with probability} \quad 1 - c_t \end{cases}$$

where

$$c_t = \frac{p(s_t = k \mid \theta, P, Y_t) p(s_{t+1} = k \mid s_t = k, P)}{\sum_{l=k-1}^{k} p(s_t = l \mid \theta, P, Y_t) p(s_{t+1} = k \mid s_t = l, P)}$$

Finally, note that $p(s_1 = 1 \mid s_2, \theta, P, Y_1) = 1$.

Step 2: $\theta \mid S, Y_T$

Sampling α_k, b_k, k_k and λ_k is more involved, since their conditional posteriors are not known in closed form. These parameters are updated using the Metropolis-Hastings (M-H) algorithm. The Metropolis-Hastings algorithm is an alternative to the Gibbs sampler. It is very useful in cases where the posterior itself is hard to sample from. In contrast to the Gibbs sampler, which draws samples directly from the conditional posterior, M-H draws a candidate, θ^* from the candidate generating density, $q\left(\theta \mid \theta^{(j-1)}\right)$ which is close to the conditional posterior followed up by an acceptance-rejection procedure with acceptance probability

$$a_{MH}\left(\theta^*, \theta^{(j-1)}\right) = \min\left\{1, \frac{p(\theta^* \mid Y_T) q\left(\theta^{(j-1)} \mid \theta^*\right)}{p\left(\theta^{(j-1)} \mid Y_T\right) q\left(\theta^* \mid \theta^{(j-1)}\right)}\right\}$$

where $p(\theta^* \mid Y_T)$ is the notation for the posterior density evaluated at $\theta = \theta^*$ and $q\left(\theta^{(j-1)} \mid \theta^*\right)$ is the candidate generating density function for $\theta^{(j-1)}$ given θ^*. If the candidate is rejected, the old draw that was sampled in the previous iteration is used as the next item of the chain.

In this paper we follow Koop (2003) and perform Independence Chain Metropolis–Hastings. Hence, $q\left(\theta \mid \theta^{(j-1)}\right) = q(\theta)$ and $q\left(\theta^{(j-1)} \mid \theta\right) = q\left(\theta^{(j-1)}\right)$, that is the candidate generating density does not depend on $\theta^{(j-1)}$. We set $q(\theta) \sim N\left(\hat{\theta}_{ML}, \Sigma_q\right)$, where $\hat{\theta}_{ML}$ is the maximum likelihood estimate of θ and $\Sigma_q \approx c \cdot var\left(\hat{\theta}_{ML}\right)$, $c \in R^+$ is chosen by the researcher in a manner to ensure a sufficient acceptance rate. We follow Koop (2003), page 98 and adjust Σ_q to get acceptance rates roughly around 40 to 50%. We do this by experimenting with different values of c until we find one which yields a reasonable acceptance rate probability. In general, the M-H steps can be summarized in the following way:

1. Draw θ^* from $q(\theta)$. That is $\theta^* \sim N\left(\hat{\theta}_{ML}, \Sigma_q\right)$.

2. Draw $u \sim U(0,1)$, where u is a draw from a Uniform distribution, U.

3. If $a_{MH}\left(\theta^*, \theta^{(j-1)}\right) > u$, accept and set $\theta^{(j)} = \theta^*$, else set $\theta^{(j)} = \theta^{(j-1)}$.

Under certain conditions, see Chib and Greenberg (1995), iterations of these steps produce samples from $p(\theta \mid Y_T)$.

For the change-point model the conditional posterior of θ_k, $k = 1, ..., m$ in each state depends only on information in regime k. Let $\hat{Y}_k = \{y_t : s_t = k\}$. Hence, at the jth step of the algorithm, θ_k is sampled using

$$a_{MH}\left(\theta_k^*, \theta_k^{(j-1)}\right) \quad = \quad \min\left\{1, \frac{p\left(\theta_k^* \mid \hat{Y}_k\right) q\left(\theta_k^{(j-1)}\right)}{p\left(\theta_k^{(j-1)} \mid \hat{Y}_k\right) q\left(\theta_k^*\right)}\right\} \tag{B.3}$$

Step 3: $P \mid S$

Simulating P from its conditional posterior is straightforward. Assume independent Beta distribution for the prior of p_{kk}, i.e. $p(p_{kk}) \sim Beta(a_0, b_0)$ where a_0 and b_0 are known hyperparameter values. The conditional posterior for each diagonal component of P is then $Beta(a_0 + n_{kk}, b_0 + 1)$, $k = 1, ..., m-1$ where n_{kk} is the number of one-step transition from state k to state k in the sequence of S.

B.2 Deviance information criterion

The deviance information criterion (DIC) of Spiegelhalter (2002) is a very useful model assessment tool. It is a Bayesian alternative to AIC and BIC. Calculation of the DIC in a MCMC application is trivial. It does not require maximization over the parameter space. It is defined as a combination of a classical estimate of fit, based on the likelihood function, $p(Y_T \mid \theta, P)$ and a penalty term, p_D. More precisely, p_D is the difference between the measure of fit and the deviance at the estimates. It describes the complexity of the model and serves as a penalization term that corrects deviance's propensity towards models with more parameters. More precisely, p_D is defined as $p_D = \overline{D(\theta, P)} - D(\bar{\theta}, \bar{P})$ where $\overline{D(\theta, P)}$ is approximated by $N^{-1} \Sigma_{j=1}^{N} - 2\log p\left(Y_T \mid \theta^{(j)}, P^{(j)}\right)$ and $D(\bar{\theta}, \bar{P}) = -2\log p\left(Y_T \mid \bar{\theta}, \bar{P}\right)$ where $\bar{\theta}$ and $\bar{P}$ are estimated from the Gibbs output using mean or mode of the posterior draws. The DIC is defined as $D(\bar{\theta}, \bar{P}) + 2p_D$. A smaller DIC indicates a better fit to the data set.

A crucial point for estimation of DIC in the context of the change-point model is the evaluation of $p\left(Y_T \mid \theta^{(j)}, P^{(j)}\right)$. In this paper the approximated solution proposed in Kim and Nelson (1999) (pp. 97-106) in which the Kalman and the Hamilton filters are merged together is applied.

B.3 Prior sensitivity analysis

In this section sensitivity of the results to prior specification is evaluated by investigating alternative priors on the transition probabilities, p_{kk}, $k = 1, ..., m - 1$. p_{kk} is the key parameter of the model because it controls the duration of each regime in S. In Table B.1, we experiment with different hyperparameter values for $p_{kk} \sim Beta(a_0, b_0)$ and report the DIC for each of these hyperparameter values by estimating 2.3-2.4 from 0 to 2 change points. For instance, the first alternative prior that is considered is $p_{kk} \sim Beta(0.1, 0.1)$ which is relatively flat. With this prior we also find evidence of one change point in the data and the change-point date corresponds also to 1982q2. In fact, regardless the values of a_0 and b_0 we still find that the one change-point specification performs best in terms on DIC. Furthermore, the change-point date for each of these specifications corresponds to 1982q1.

# CP	$Beta(0.1, 0.1)$	$Beta(8, 0.1)$	$Beta(10, 2)$	$Beta(20, 0.1)$	$Beta(100, 0.1)$
0	993.104	993.104	993.104	993.104	993.104
1	947.675	947.980	947.672	947.678	948.101
2	972.797	972.499	972.561	972.761	973.375

This table compares DIC for different values of a_0 and b_0 where $p_{kk} \sim Beta(a_0, b_0)$. The priors of the other parameters are set according to section 2.2. #CP: the number of change points that are conditioned on.

Table B.1: Prior sensitivity analysis (break model)

Research Papers
2013

CREATES
Center for Research in Econometric
Analysis of Time Series

2013-07:	Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns
2013-08:	Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation
2013-09:	Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox
2013-10:	Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models
2013-11:	Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis
2013-12:	Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications
2013-13:	Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions
2013-14:	Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments
2013-15:	Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation
2013-16:	Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression
2013-17:	Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model
2013-18:	Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models
2013-19:	Asger Lunde and Kasper V. Olesen: Modeling and Forecasting the Volatility of Energy Forward Returns - Evidence from the Nordic Power Exchange
2013-20:	Anders Bredahl Kock: Oracle inequalities for high-dimensional panel data models
2013-21:	Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen: Lassoing the Determinants of Retirement
2013-22:	Johannes Tang Kristensen: Diffusion Indexes with Sparse Loadings
2013-23:	Asger Lunde and Anne Floor Brix: Estimating Stochastic Volatility Models using Prediction-based Estimating Functions
2013-24:	Nima Nonejad: A Mixture Innovation Heterogeneous Autoregressive Model for Structural Breaks and Long Memory
2013-25:	Nima Nonejad: Time-Consistency Problem and the Behavior of US Inflation from 1970 to 2008