Modelling Inflation Volatility

Eric Eisenstat*
University of Bucharest and RIMIR

Rodney W. Strachan
University of Queensland

February 18, 2014

*Corresponding author: Faculty of Business Administration, University of Bucharest, Blvd. Regina Elisabeta nr. 4-12, Bucharest (Romania), eric.eisenstat@faa.ro. Rodney Strachan is a Fellow of the Rimini Centre for Economic Analysis and the Centre for Applied Macroeconomic Analysis.

ABSTRACT

This paper discusses estimation of US inflation volatility using time varying parameter models, in particular whether it should be modelled as a stationary or random walk stochastic process. Specifying inflation volatility as an unbounded process, as implied by the random walk, conflicts with priors beliefs, yet a stationary process cannot capture the low frequency behaviour commonly observed in estimates of volatility. We therefore propose an alternative model with a change-point process in the volatility that allows for switches between stationary models to capture changes in the level and dynamics over the past forty years. To accommodate the stationarity restriction, we develop a new representation that is equivalent to our model but is computationally more efficient. All models produce effectively identical estimates of volatility, but the change-point model provides more information on the level and persistence of volatility and the probabilities of changes. For example, we find a few well defined switches in the volatility process and, interestingly, these switches line up well with economic slowdowns or changes of the Federal Reserve Chair.

Keywords: Inflation volatility, monetary policy, time varying parameter model, Bayesian estimation, Change-point model.

JEL Classification: C11, C22, E31

Summary

This paper considers the relative virtues of modelling inflation volatility as a stationary or non-stationary (random walk) process, both of which are conceptually problematic: random walks have increasing probability bounds while stationary processes cannot capture the large, low frequency movements we observe. We propose a change-point process that captures low-frequency movements while being always bounded in probability. To implement it, we develop a new methodology for sampling bounded state vectors. Our results suggest that for empirical purposes, either stationary or non-stationary processes are adequate, but our new process provides interesting inference on the changes in level and persistence of volatility.

1 Introduction

The literature on modelling inflation is voluminous as inflation has an important place in many macroeconomic issues. For example, it is central to studies of the transmission of monetary policy shocks (Cogley and Sargent (2001 and 2005), Primiceri (2005), Sargent, Williams and Zha (2006), and Koop, Leon-Gonzalez and Strachan (2009)), there has been a resurgence in interest in the Phillips curve (King and Watson (1994), Staiger, Stock and Watson (1997), Koop, Leon-Gonzalez and Strachan (2010)), and there is a large literature devoted to forecasting inflation (e.g., Ang, Bekaert, and Wei (2007), Stock and Watson (2007 and 2009), D'Agostino, Gambetti, and Giannone (2009), Croushore (2010), Clark and Doh (2011), Chan (2013) and Wright (2013)).

Time varying parameter models of macroeconomic variables such as inflation have proven useful on a range of questions of interest to policymakers and the state space representation for these model has been a popular choice of specification. While there has been much attention to modelling the conditional mean of inflation, recently there has also been increasing interest in the variance with some evidence that the variance changes more (Primiceri (2005)) and more often (Koop, *et al.* (2009)) than do the mean coefficients. Therefore a feature that has proven important in such models is to allow for heteroscedasticity and a common specification in macroeconomics of this is stochastic volatility using a random walk for the state equation for log volatility (see for example, Cogley and Sargent (2005), Primiceri (2005) and Koop, *et al.* (2009)). This specification is attractive because of its parsimony, ease of computation and the smoothness it induces in the estimated volatility over time.

While the random walk specification is useful for practical reasons, it can be criticised as inappropriate since it implies that the range of likely values for volatility increases over time and is in the limit unbounded[1], which is clearly not what we observe. An alternative specification for stochastic volatility, which is commonly used in finance, is a stationary autoregressive model for the log volatility. Such a model implies inflation is bounded in probability at all horizons and has an easily derived stationary distribution. This property is appropriate for many financial processes where the variance shows only brief deviations far from its mean and then rapid mean reversion.

The behaviour of US inflation volatility, however, is not well described by a stationary, quickly mean reverting process. Although it is an unobserved latent process, common patterns have emerged in estimates presented in the literature on the behaviour of this process over time. Representative estimates of the volatility of inflation are presented in Figure 1. The pattern is an increase in the level of volatility that persisted during the 1970s and early 1980s, followed by a decline towards a lower lever over the late 1980s and early 1990s, and finally another increase in the 2000s. Other estimates in the literature differ in the detail, but what is generally evident in estimates of the volatility of inflation are large, low frequency movements. While the random walk model of volatility can be criticized for being incoherent, this model could be viewed as an approximation to the true process. In contrast, when we consider this behaviour, a time invariant stationary model of volatility does not appear appropriate for modelling inflation volatility.

[FIGURE 1 HERE]

This paper makes several contributions. i) We discuss the relative advantages and disadvantages of the random walk and stationary specifications of inflation volatility.

[1] By this we mean that the process is bounded in probability for any finite horizon, but these probability bounds widen over time such that in the limit, the process is not bounded in probability.

ii) This discussion leads us to present a change-point model of log inflation volatility that switches between stationary models with different levels and dynamics for inflation volatility. This model meets the theoretical concern that volatility should be bounded, but also permits the model to capture the occasional, large movements in the volatility level that have been observed over the past forty years. iii) While the specification and sampler are based upon the model of Koop and Potter (2007) for changes in the measurement equation, we develop a new specification that speeds computation when stationarity constraints are imposed and we expect that this algorithm will prove useful in a wider range of settings. iv) We compare outputs from our model with those from the random walk and stationary specifications and find that estimates of volatility differ little among the specifications and the estimated parameter values from the stationary model are close to the nonstationary region. The results suggest either a random walk or stationary model is a practically sensible specification to use for estimating inflation volatility, when the volatility itself is not of central concern.

Another contribution is v) a characterisation of regimes of inflation volatility since 1960. An advantage of our model over the random walk and stationary specifications is that it provides much more information on the level and persistence of inflation volatility. Our model also informs us on points at which inflation regimes change. Conditional upon the mean of volatility the change probabilities sharpen and provide an interesting insight. The dates at which the model switches strongly suggest that the changes occur soon after economic slowdowns or a new Federal Chair appointment.

In the next section, Section 2, we discuss and compare the attractive and less attractive properties of the random walk and stationary specifications for inflation volatility. In Section 3 we introduce our change-point model for the latent volatility process. This model is an adaptation of the model proposed by Koop and Potter

(2007) for the measurement equation. However, we apply it to a latent process, log volatility, with the addition that we impose stationarity restrictions. The stationarity restriction complicates the estimation and so we present an new specification that simplifies and speeds estimation. In Section 4 we present and discuss the empirical results from the three models. Section 5 concludes the paper.

2 The Random Walk and Stationary Models

For both discussion and estimation, we will use a common measurement equation specification for inflation, y_t. In general, we write this as

$$y_t = \mu_t + \exp\left(h_t/2\right)\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1),$$

where μ_t captures the conditional mean equation. For the moment, we leave the latter unspecified and focus on the specification of the state equation for the latent log volatility process h_t. We restrict the discussion of the stationary and nonstationary models of h_t to one specification of each, although the general points we make carry over to a wider range of specifications that a researcher might consider. The random walk specification for log volatility, which is commonly used in macroeconometrics, is

$$\begin{aligned} h_t &= h_{t-1} + \nu_t \quad \nu_t \sim \mathcal{N}\left(0, \sigma_h^2\right), \qquad (1) \\ h_0 &\sim \mathcal{N}\left(\underline{h}, V_{h_0}\right). \end{aligned}$$

Many applications, particularly in finance, use a stationary specification. For our purposes we will use the following process for log volatility:

$$h_t = \eta + \rho\left(h_{t-1} - \eta\right) + \nu_t, \quad \nu_t \sim \mathcal{N}\left(0, \sigma_h^2\right), \tag{2}$$
$$h_0 \sim \mathcal{N}\left(\eta, V_{h_0}\right).$$

Clearly the process (1) nests within the process (2) when $\rho = 1$, at which point η is no longer identified. To ensure stationarity of h_t in (2) we impose the restriction $|\rho| < 1$.

One argument made for the random walk specification is that it 'captures the idea that "the coefficients today have a distribution that is centered over last period's coefficients"' (Koop, Léon-González and Strachan (2011)). This view takes seriously the idea that the state equation is the prior expression of our beliefs about how log volatility evolves over time. If we do consider the state equation as an expression of prior beliefs, however, one might conclude (1) is not coherent. With finite expectation $E\left(h_s\right) = \underline{h}$ for any $s < t$, the model (1) implies h_t is a Martingale and so has finite expectation[2] for all $t < \infty$. Sending $s \to -\infty$, h_t is no longer bounded in probability, and is no longer a Martingale since the first absolute unconditional moment, $E\left(|h_t|\right)$, no longer exists (see Billingsley (1986)), which also implies the process does not have a finite mean. That the random walk implies the volatility process is unbounded in probability in the limit is a feature that some econometricians express reservations about (see, for example, the discussion in Primiceri (2005)). The specification for h_t in (2), by contrast, always has a finite mean η and variance and is bounded in probability.

In finite samples, the random walk specification in (1) is always bounded in prob-

[2]Conditional upon h_s.

ability. However, it still implies some properties for volatility that the researcher may not like. For example, conditional upon h_0, the variance of the log volatility grows linearly in time as $V(h_t|h_0) = t\sigma_h^2$. This property can be interpreted as reflecting that uncertainty about the volatility is increasing through time. For the stationary process, the variance is $V(h_t|h_0) = \sigma_h^2 \frac{1-\rho^{2t}}{1-\rho^2}$ which is also increasing over time but converges to the constant $\frac{\sigma_h^2}{1-\rho^2}$ in the limit. For values of ρ and T that one typically finds in macroeconometrics, $\rho^{2T} \approx 0$ and so the variance can be regarded as effectively constant[3] towards the end of the sample.

Another important property of the prior that merits consideration is the correlation structure. If two states, h_s and h_t, are *a priori* independent, there is nothing in the prior to ensure that information in y_s about h_s will be transmitted to h_t. If two states, h_s and h_t, are *a priori* dependent, then information in y_s about h_s will be transmitted to h_t and the strength of this transmission will depend upon the strength of the dependence. In the extreme, if h_s is perfectly correlated with h_t, all information in y_s about h_s will be transmitted to h_t as there is effectively only one state in this case. A prior specification with strong dependence, then, can roughly be thought of as achieving parsimony without losing any parameters.

In the state space model, correlation between the states induces dependence[4] and the correlation structure is induced by the state equation. Conditional upon h_0, the stationary process in (2) implies a correlation between h_t and h_{t-q} of

$$r_{|\rho|<1,t,t-q} = \rho^q \sqrt{\frac{1-\rho^{2(t-q)}}{1-\rho^{2t}}}$$

[3]We may consider large values of ρ and small values of T and still get this result. For example, with $\rho = 0.95$ and $T = 40$, we have $\rho^{2T} = 0.0165$.

[4]Initial conditions and the distribution of the state error are other sources.

and for (1) this is

$$r_{\rho=1,t,t-q} = \sqrt{\frac{t-q}{t}}.$$

As the sample increases in size, the correlation between the last and any earlier states in (1) becomes stronger. In the limit $\lim_{T\to\infty} r_{\rho=1,T,T-q} = 1$. This property suggests the time varying parameter model with a random walk state equation converges to a time invariant model. While this may not reflect the researcher's beliefs, it has advantages. In particular, the increasing correlation increases the transmission of information from earlier states to later ones. That the correlation with previous states increases could be viewed as a positive feature in that it compensates for the increasing variance of, or uncertainty about, the later states.

For the purposes of estimation, the random walk specification can be thought of as a parsimonious approximation to a stationary specification with a high persistence. Further, the random walk is attractive in that it implies greater smoothness than the stationary model with low persistence. Finally, the random walk specification implies stronger correlation among all log variances than does the stationary specification in (2). That is, the correlations between h_t and h_{t-q} for the models (2) and (1) are ordered as $r_{|\rho|<1,t,t-q} < r_{\rho=1,t,t-q}$. As discussed earlier, estimation of latent states in the state space model is aided by a stronger correlation structure.

3 A New Model of Inflation Volatility

A goal of this paper is to investigate the support for an alternative model that captures the feature of persistent shifts in the level of inflation volatility, but also implies that volatility has a stationary distribution at any point in time. To specify a model that is stationary at all times but permits changes in the level of volatility, we employ a change-point model based upon that developed in Koop and Potter (2007). In this section, we present our change-point model of inflation volatility. Existing techniques could be used to estimate this model, but they turn out to be computationally slow. We therefore present an alternative representation that results in a more efficient sampler.

A significant difference between the model we present and the models used in Koop and Potter (2007) is that we apply the change-point process to the parameters governing the evolution of the volatility, which is a latent process. That is, the state equation for h_t is given as

$$h_t = \eta_{s_t} + \rho_{s_t} \left(h_{t-1} - \eta_{s_{t-1}}\right) + \nu_t, \quad \nu_t \sim \mathcal{N}\left(0, \sigma^2_{h,s_t}\right), \tag{3}$$

where $s_t \in \{1, \cdots, M\}$ indicates the regime at period t and M is the maximum number of regimes. Stationarity in each regime is imposed by assuming $|\rho_{s_t}| < 1$ for all s_t. Following Koop and Potter (2007), let

$$\begin{aligned} \eta_m &= \eta_{m-1} + \xi_{\eta_m}, & \xi_{\eta_m} &\sim \mathcal{N}\left(0, \sigma^2_\eta\right), \\ \rho_m &= \rho_{m-1} + \xi_{\rho_m}, & \xi_{\rho_m} &\sim \mathcal{N}\left(0, \sigma^2_\rho\right), \\ \ln \sigma^2_{h,m} &= \ln \sigma^2_{h,m-1} + \xi_{\sigma_m}, & \xi_{\sigma_m} &\sim \mathcal{N}\left(0, \sigma^2_\sigma\right), \end{aligned}$$

and define the vectors $\eta = (\eta_1, \ldots, \eta_M)'$, $\rho = (\rho_1, \ldots, \rho_M)'$, $\sigma^2_h = \left(\sigma^2_{h,1}, \ldots, \sigma^2_{h,M}\right)'$, $h = (h_1, \ldots, h_T)'$, and $s = (s_1, \ldots, s_T)'$. These vectors are important as we do not use a Kalman filter based algorithm. Instead we use the more efficient precision based samplers (see Rue (2001), Chan and Jeliazkov (2009) and McCausland, Millera, and Pelletier (2011)).

An important feature of the Koop and Potter (2007) approach is the explicit specification of a prior on the duration of each regime. Define the time of the change-point from one regime to the next as $\tau_m = \{t : s_{t+1} = m+1, s_t = m\}$ such that the duration is defined as $d_m = \tau_m - \tau_{m-1}$. A hierarchical prior is specified for the duration. At the first level, the duration is *a priori* a Poisson process with mean λ_m. The parameter λ_m has a Gamma distribution $\mathcal{G}\left(\underline{\alpha}_\lambda, \beta_\lambda\right)$ in which $\underline{\alpha}_\lambda$ is fixed and the *rate* parameter β_λ is given a Gamma distribution $\mathcal{G}(\underline{\xi}_1, \underline{\xi}_2)$. This setup has a number of advantages and addresses several issues in modelling change-point processes as discussed in detail in Koop and Potter (2007). The notation we use for these parameters is identical to that of Koop and Potter (2007) and we refer the reader to that paper for further details. For our purposes we note that this structure implies the prior mean duration is

$$\underline{d}_m = E\left(d_m\right) = 1 + \underline{\alpha}_\lambda \left(\frac{\underline{\xi}_2}{\underline{\xi}_1 - 1}\right).$$

For η and σ_h^2, the above model implies no particular complication and we can complete the specification of the priors on these two parameter vectors with

$$\eta_0 \sim \mathcal{N}(\kappa_{\eta_0}, V_{\eta_0}), \qquad \ln \sigma_{h,0}^2 \sim \mathcal{N}(\kappa_{\sigma_0}, V_{\sigma_0}),$$
$$\sigma_\eta^2 \sim \mathcal{IG}(\gamma_\eta, \delta_\eta), \qquad \sigma_\sigma^2 \sim \mathcal{IG}(\gamma_\sigma, \delta_\sigma),$$

where the κ's, V's, γ's and δ's are given constants. For a given volatility regime m to be stationary, however, we need to impose $|\rho_m| < 1$, which requires some care in specifying the prior on ρ. One way to proceed is to follow Chan, *et al.* (2013) and put a univariate truncated normal prior directly on ξ_{ρ_m}, along with

$$\rho_0 \sim \mathcal{N}(\kappa_{\rho_0}, V_{\rho_0}),$$
$$\sigma_\rho^2 \sim \mathcal{IG}(\gamma_\rho, \delta_\rho),$$

then apply the Chan and Strachan (2012) algorithm for sampling both ρ and the hyper-parameters (ρ_0, σ_ρ^2) from non-standard, but tractable distributions. While conceptually straightforward, this approach could be computationally intensive, and given that we need to nest it within a regime-searching algorithm that is already computationally demanding, we consider a simpler alternative. Moreover, there is an aspect to the algorithm in Chan, *et al.* (2013) that could make it particularly unsuitable in this application. We return to this point shortly.

Consider an alternative prior on $\left(\sigma_\rho^2, \rho_0, \rho\right)$, namely

$$\begin{aligned} p\left(\sigma_\rho^2, \rho_0, \rho\right) &\propto p\left(\sigma_\rho^2\right) p\left(\rho_0\right) p\left(\rho|\rho_0, \sigma_\rho^2\right) 1(|\rho| < \iota_M), \qquad (4) \\ p\left(\sigma_\rho^2\right) &= \mathcal{IG}_{\sigma_\rho^2}\left(\gamma_\rho, \delta_\rho\right), \\ p\left(\rho_0\right) &= \mathcal{N}_{\rho_0}\left(\kappa_{\rho 0}, V_{\rho 0}\right), \\ p\left(\rho|\rho_0, \sigma_\rho^2\right) &= \mathcal{N}_\rho\left(\rho_0 \iota_M, \sigma_\rho^2 \left(H'H\right)^{-1}\right), \end{aligned}$$

where ι_M denotes a $M \times 1$ vector of ones and $|\rho| < \iota_M$ is intended to mean that each element of ρ is less than one in absolute value. The the $M \times M$ matrix H is

$$H = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}.$$

The above specification simplifies computation to the extent that conditional on ρ, the hyper-parameters $\left(\rho_0, \sigma_\rho^2\right)$ can be sampled in a standard way. However, the conditional distribution

$$\rho|\sigma_\rho^2, \rho_0, \eta, h, \sigma_h^2, s \sim \mathcal{N}_{|\rho|<\iota_M}\left(\overline{\rho}, \overline{V}_\rho\right), \tag{5}$$
$$\overline{\rho} = \overline{V}_\rho\left(\frac{\rho_0}{\sigma_\rho^2}\widetilde{\iota}_M + X'\widetilde{\Sigma}_h^{-1}(h-\widetilde{\eta})\right),$$
$$\overline{V}_\rho = \left(\frac{1}{\sigma_\rho^2}H'H + X'\widetilde{\Sigma}_h^{-1}X\right)^{-1},$$

where $\widetilde{\eta} = (\eta_{s_1}, \ldots, \eta_{s_T})'$ and $\widetilde{\Sigma}_h = \text{diag}(\sigma_{h,s_1}^2, \ldots, \sigma_{h,s_T}^2)$, is *multivariate truncated normal.*[5] Sampling from this distribution directly is difficult. A brute-force approach would be to draw ρ from an unrestricted multivariate normal distribution until we get a ρ that satisfies $|\rho| < \iota_M$. This could work reasonably well as long as the constraints are only binding with a fairly low probability in the posterior and this is the approach used in Chan, *et al.* (2013) to obtain a proposal for the ARMH step in their algorithm. If we were to follow Chan, *et al.* (2013) and specify univariate truncated normal priors on $(\rho_m|\rho_{m-1})$, we would inevitably need to use this approach also. The algorithm in Chan and Strachan (2012) would dictate that proposals of ρ be drawn from exactly the distribution in (5). However, there are issues unique to our specification that suggest this will not be very efficient.

In Chan, *et al.* (2013), sampling from an unrestricted multivariate normal and rejecting draws that do not satisfy the constraints works precisely because the constraints are not binding with a high probability. In our model, however, there is an important difference that could make this approach difficult to implement. Specifically, for any regime-switching algorithm with an unknown number of change points,

[5] $\widetilde{\iota}_M$ is the $M \times 1$ vector $\widetilde{\iota}_M = (1, 0, \ldots, 0)'$, and X is a $T \times M$ matrix with the element at row t, column m given by

$$x_{t,m} = \begin{cases} h_{t-1} - \eta_{s_{t-1}} & \text{if} \quad s_t = m \\ 0 & \text{otherwise} \end{cases}.$$

Clearly, $X'X$ is diagonal, and therefore, the precision for ρ is sparse and banded. Note that in the above definition, we assume $s_0 = 1$ for completeness.

the sampler will always have m^* states covering the observed sample and m^* will be less than or equal to M. Unless M is set to be very small, then it will often be the case that m^* will be *strictly* less than M which means that for all out of sample regimes m, that is where $m^* < m < M$, there is no data to pin down estimates of ρ_m. Since the posterior for ρ_m in these cases relies only on the prior, which essentially assumes that ρ_m evolves as a random walk across regimes, a sufficiently low m^* relative to M will imply a high variance for all out of sample ρ_m with m closer to M and farther from m^*, even if σ_ρ^2 is "small". For such ρ_m, the probability of falling outside the interval $(-1, 1)$ will be high, and in consequence, the probability that *at least one* element in ρ will not satisfy the constraint will likely be too high to make the brut-force rejection sampling method practical. This contrasts with the model in Chan, *et al.* (2013) which we may regard as a regime-switching model with a *fixed* number of regimes $m^* = T$ (thus, all ρ_m are related to a data point).

Returning to the prior in (4), one alternative approach would be to sample ρ in M blocks as in Geweke (1991). Such an algorithm would always generate draws of ρ in the correct space, but given that M would typically be high, and the elements of ρ may be highly correlated then, for reasons discussed in the previous section, sampling efficiency could be quite poor. In light of this, we propose a third approach that exploits the particular covariance structure in ρ. Specifically, we augment the parameter space with a $M \times 1$ latent vector of static factors $f = (f_1, \ldots, f_M)$, such that

$$p(\sigma_\rho^2, \rho_0, f, \rho) \propto p\left(\sigma_\rho^2\right) p\left(\rho_0\right) p\left(f|\sigma_\rho^2\right) p\left(\rho|\rho_0, f, \sigma_\rho^2\right) 1(|\rho| < \iota_M), \tag{6}$$

$$p\left(\sigma_\rho^2\right) = \mathcal{IG}_{\sigma_\rho^2}\left(\gamma_\rho, \delta_\rho\right),$$

$$p\left(\rho_0\right) = \mathcal{N}_{\rho_0}\left(\kappa_{\rho_0}, V_{\rho_0}\right),$$

$$p\left(f|\sigma_\rho^2\right) = \mathcal{N}_f\left(0, \sigma_\rho^2 I_M\right),$$

$$p\left(\rho|\rho_0, f, \sigma_\rho^2\right) = \mathcal{N}_\rho\left(\rho_0 \iota_M + H^{-1} A f, 0.25\sigma_\rho^2 I_M\right),$$

where A is a lower-triangular matrix such that $AA' = I_M - 0.25HH'$. Note that $I_M - 0.25HH'$ is guaranteed to be *positive-definite* and therefore A can be easily computed by the Cholesky decomposition. This is, in fact, closely related to the Stern (1992) decomposition and it is straightforward to verify that integrating (6) over f yields the original prior in (4). The latter implies two things: (i) the priors are equivalent, and (ii) all parameters besides ρ - including the hyper-parameters $\left(\rho_0, \sigma_\rho^2\right)$ - can be sampled marginally of f exactly as before.

The only role for the draws of f is to permit efficient sampling of ρ. Given a draw of f, the conditional (on f) distribution for ρ is

$$p\left(\rho|f, \sigma_\rho^2, \rho_0, \eta, \sigma_{h,m}^2, h, s\right) = \prod_{m=1}^{M} p\left(\rho_m|f, \sigma_\rho^2, \rho_0, \eta, \sigma_{h,m}^2, h, s\right), \tag{7}$$

$$\rho_m|f, \sigma_\rho^2, \rho_0, \eta, \sigma_{h,m}^2, h, s \sim \mathcal{N}_{|\rho_m|<1}\left(\overline{\rho}_m, \overline{V}_{\rho_m}\right),$$

$$\overline{\rho}_m = \overline{V}_{\rho_m}\left(\frac{4}{\sigma_\rho^2} k_m + \frac{1}{\sigma_{h,m}^2} x_m'(h - \widetilde{\eta})\right), \tag{8}$$

$$\overline{V}_{\rho_m} = \frac{\sigma_\rho^2 \sigma_{h,m}^2}{4\sigma_{h,m}^2 + \sigma_\rho^2 x_m' x_m}. \tag{9}$$

In (8)-(9), k_m refers to the m-th element of the vector $k = \rho_0 \widetilde{\iota}_M + H'Af$, while x_m is the m-th column of X. These quantities are straightforward to compute, and hence, sampling ρ from independent univariate truncated normal distributions is straightforward (e.g. Robert (1995)).

Likewise, there is no difficulty in simulating f conditional on ρ. Noting that

$$\left(I_M + 4A'\left(HH'\right)^{-1}A\right)^{-1} = I_M - A'A,$$

the appropriate conditional distribution may be written as

$$p\left(f|\rho, \sigma_\rho^2, \rho_0, s\right) \sim \mathcal{N}\left(4D(\rho - \rho_0\iota_M), \sigma_\rho^2\left(I_M - A'A\right)\right), \tag{10}$$
$$D = (I_M - A'A)(H^{-1}A)'.$$

Because $A'A$ is sparse and banded, simulation from (10) is fast even for large M. Moreover, the quantities D and $I_M - A'A$ involve only known constants and, hence, need to be computed only once before commencing the MCMC. Details regarding the full Gibbs sampler are provided in the appendix.

4 Empirical Results

In this section we report the estimates of inflation volatility from the random walk, stationary and change-point models. For estimation purposes, we specify μ_t as simply a constant-coefficient AR(4) equation restricted to stationarity. However, to verify the robustness of the results presented below, we also ran three alternative specifications:

1. time-varying intercept with constant intercept and AR(4) coefficients restricted to stationarity and the state equation for the intercept specified as a stationary AR(1) process;

2. bounded trend inflation (Chan, *et al.* (2013)) with the time-varying trend constrained to $[0, 5]$ and the time-varying AR coefficient to $[0, 1)$;

3. a fully time-varying, unrestricted AR(2).

These three conditional mean specifications involve a larger computational burden, but yield results very similar to those reported in this section.

The data are the quarterly inflation rate, computed as $400 \ln (CPI_t / CPI_{t-1})$ where CPI_t is U.S. CPI data where the period covers 1947Q2-2013Q2 which after losing four lags for the mean equation, gives $T = 261$.[6] Using the change-point model, we also show estimates of the evolving persistence and level of volatility. Finally, we consider the probabilities of switching regimes at each point in time.

The estimates of the parameters governing the behaviour of the latent inflation volatility process, such as ρ_m and η_m, are sensitive to the prior settings. In particular, these parameters are sensitive to the prior expected duration. We therefore consider four combinations of hyper-parameters controlling the regime-search algorithm implying four different expected regime durations. Recall M is the maximum number of regimes allowed while $\underline{\alpha}_\lambda, \underline{\xi}_1, \underline{\xi}_2$ determine the prior expected regime duration as

$$\underline{d}_m = E(d_m) = 1 + \underline{\alpha}_\lambda \left(\frac{\underline{\xi}_2}{\underline{\xi}_1 - 1} \right).$$

For all cases, $M = 30$ and $\gamma_\eta = \gamma_\rho = \gamma_\sigma = 5$. We set $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 \in \{15, 30, 60, 120\}$ implying prior expected durations of approximately 17, 32, 62 and 122 quarters, that is $\underline{d}_m \in \{17, 32, 62, 122\}$. Corresponding to each of these four configurations, we set $\delta_\eta = \delta_\rho \in \{0.1, 0.25, 0.5, 1\}$, thus allowing for greater cross-regime transitions in the parameters for models with longer *a priori* durations. The prior on σ_h^2 is fixed for all

[6]The data were downloaded from the Federal Reserve Economic Data, St. Louis Fed.

models with $\delta_\sigma = 0.01$.[7]

As the prior expected regime duration increases the model will switch regimes less often and increasingly approximate the time invariant stationary model which has $\underline{d}_m = d_m \geq T$ and $M = 1$. At the other extreme, the random walk model has $\underline{d}_m = d_m = 1$ and $M = T$. Although the random walk model does not technically nest within (2) or (3) since the supports of ρ and ρ_m do not include the point $\rho_m = 1$, if the data prefers the random walk model this parameter will approach this boundary and provide a good approximation to (1).

In Figure 2 we report the posterior estimates, $E(h_t|y)$, of the log volatility from the four change-point models, the stationary model and the random walk model. It is immediately apparent that the estimates differ very little. Comparing the extremes of the stationary model and the random walk model in the bottom two panels, we see that the estimates are close and there is considerable overlap of the error bands. The estimated parameters of the stationary model are close to the random walk model with the estimated posterior mean of ρ at $E(\rho|y) = 0.94$. The estimates from the random walk model are slightly smoother than from the stationary model. This is to be expected but it is not a dominant or distinguishing feature of the estimates.

[FIGURE 2 HERE]

As discussed in the introduction, most estimates of inflation volatility in the literature show large low frequency movements and this behaviour is not consistent with an exponentially mean-reverting process such as (2). Note also, a unit root is excluded from the support of ρ. That the estimated value of this parameter in (2) is pushed towards the unit root is how this model accommodates the low frequency

[7] We also set throughout $\kappa_{\eta_0} = \kappa_{\rho_0} = 0$, $V_{\eta_0} = V_{\rho_0} = 10$ and $\kappa_{\sigma_0} = -2.45, V_{\sigma_0} = 0.29$. The latter implies a log-normal prior on $\sigma^2_{h,0}$ that is approximately equivalent to $\mathcal{IG}(5, 0.4)$—i.e., with $E\left(\sigma^2_{h,0}\right) = 0.1$ and $V\left(\sigma^2_{h,0}\right) = 3.3 \times 10^{-3}$.

behaviour in volatility. The posterior estimates from the change-point model come from four models with very different prior settings as captured by the range of prior expected durations, $\underline{d}_m$. The estimates $E(h|y)$ are clearly not sensitive to the prior setting and again there is considerable overlap of the error bands.

To provide some summary measure of model preference, we report in Table 1 the deviance information criteria (DIC) (Spiegelhalter, *et al.* (2002)). The DIC has previously been used in a related set up to compare stochastic volatility models (e.g. Berg, *et al.* (2004)), and it is generally used to assess the ability of the model to predict future data that would be generated by the same mechanism as the existing data. It is particularly useful with large complex models with correlated latent parameters, such as state space models, where simply counting the number of parameters as a measure of model complexity is not appropriate due to the prior correlation structure.

[TABLE 1 HERE]

A smaller value of the DIC for one model than for another model suggests the first model is preferable. The first five models in Table 1 have stationary (within each regime) log volatility. The results in this table show little difference among the range of stationary models, although there is a slight preference for the short duration change-point models (with $\underline{\alpha}_\lambda = 15$ and $\underline{\alpha}_\lambda = 30$) over the other three stationary models. The random walk model is the least preferred model and the long duration change-point model (with $\underline{\alpha}_\lambda = 120$) and single regime stationary model (2) are equally ranked. These results suggest there is a preference in the data for stationarity but changing regimes are also very important.

To further assess the predictive abilities of the six models in practice, we compare predictive likelihoods for two different periods: 2008Q3–2010Q2 and 2011Q3–2013Q2. The first eight-quarter period is interesting because inflation dropped suddenly from

6.1% in 2008Q3 to -9.9% in 2008Q4, making it a very difficult period to forecast. In contrast, the second eight-quarter period is characterized by relatively stable inflation volatility. The results in Table 2 summarize the joint predictive likelihoods for each eight-quarter period.

[TABLE 2 HERE]

Evidently, for 2008Q3–2010Q2, the random walk specification would have predicted substantially better than the other models, outperforming the stationary model by a Bayes factor of around 4 and the shortest duration change-point model ($\underline{\alpha}_\lambda = 15$) by Bayes factor of around 2. This is not surprising taking into account that the period was marked by a great deal of turbulence in terms of inflation, and in consequence, models with short regime durations that allow for more flexibility—with the random walk specification being the most flexible in this sense—performed better.

On the other hand, the random walk model yields the weakest predictive performance for 2011Q3–2013Q2, whereas the single stationary regime model performs the best. The differences in predictive likelihoods for this period, however, are not as drastic indicating that in times of stability, all models forecast comparably well although the imposed stationarity is indeed beneficial. Moreover, in this case, the best performing change-point model is the one with the longest regime duration ($\underline{\alpha}_\lambda = 120$) and closest to the stationary specification. It appears, therefore, that whether the stationary specification or the random walk is better for forecasting depends on how volatile inflation is during the forecasted period.

Note that the change-point models always tend to fall somewhere in between these two extreme specifications in terms of forecasting performance, thus suggesting that allowing for volatility to change regimes provides an interesting balance in the following sense. Relative to the single stationary regime specification, the change-point model allows for a larger probability that volatility could increase suddenly and is, therefore, more accommodating of dramatic changes in the inflation process. On the other hand, it avoids the potential overfitting that the random walk is prone to, leading to better forecasts in periods when inflation is stable. Following this line of reasoning, one interpretation of the DICs ranking in Table 1 is that the balance provided by the change-point models makes them slightly preferable for forecasting in the long run.

[FIGURE 3 HERE]

We next consider the estimates of the level, η_m, and persistence, ρ_m, of inflation from the change-point model. Figure 3 shows the estimates of the mean level of inflation, η_m, from the four change-point models. Except for when the expected regime duration is very long, there is clearly evidence of movements in the mean of inflation volatility and this pattern is similar over the prior settings. There is a fall in volatility in the 1960s, a rise from 1970 to the early 1980s, another decline from this point until the mid to late 1990s followed by another increase. We see that the level of volatility around 2008 is similar to that of the early 1980s suggesting that the great moderation in this variable at least had clearly passed by this time.

Comparing the plots in Figure 2 with those in Figure 3, we see that there are extended periods when inflation volatility is above or below its mean. In particular, in periods of low volatility (e.g., during the 1990s) the volatility is lower than its mean and in periods of high volatility (e.g., late 1970s and early 1980s) the volatility is higher than its mean. Even assuming a change-point model of stationary inflation volatility, we see long periods of persistent deviations from the mean.

Figure 4 reports the estimated mean and $(16\%, 84\%)$ credible interval for the volatility persistence parameter, ρ_m. The general pattern in the mean $E(\rho_m|y)$ is one of a rise in the level of persistence since the 1960s with a slight drop after 2008-2009. The association of persistence with level is not strong, but it does appear that the persistence falls when the mean level of volatility increases. However, the error bands for ρ_m are reasonably wide except when the prior mean duration is high (i.e., when $\underline{d}_m = 122$). When the prior duration is long, the information in the data is used to estimate fewer parameters. This information improves the estimate of ρ_m rather than η_m which has quite wide error bands. This behaviour again appears to agree with the idea that a single (or few) stationary regime(s) cannot adequately capture the large low frequency movements in inflation volatility such as those we observe in these models. As a result ρ_m approaches 1 and η_m becomes less well identified near this point.

[FIGURE 4 HERE]

The estimated probabilities of a change in regime at each point in time are reported in Figure 5. A general observation is that the probability of a regime change at any point in time is always quite low, usually below 10%. As is to be expected, the probabilities fall as the prior expected duration increases and when the prior expected duration is very long, i.e. $\underline{d}_m = 122$, there is little evidence of regime change. For the remaining three cases, when $\underline{d}_m \in \{17, 32, 62\}$, there are clear spikes in probability of regime change around 1973, and 1992 and a few other points common across the plots. However, it is very difficult to discern much from the plots about likely points of regime change and this is likely due to the fact that the regime-switching structure is very high up in the prior hierarchy in our specification.

[FIGURE 5 HERE]

The marginal posterior estimated probabilities of regime change do not vary greatly over the sample. Given how vague is the information on regime-switching, we estimated the probabilities at a point of high probability mass for the volatilities. That is, we re-sampled the models conditional upon a single volatility path. We set the vector of log volatilities h to equal $E(h \,|\, y, \underline{\alpha}_\lambda = 30)$[8] and re-ran the simulations of the models $\underline{\alpha}_\lambda = 15$ and $\underline{\alpha}_\lambda = 60$ conditional upon these volatilities. The results for these two models are reported in Figure 6. Treating the vector of log volatilities h as observed significantly sharpens the estimates of the probabilities and we can see clear points of very likely regime change. We also obtain more precise estimates of η_m and ρ_m.

[FIGURE 6 HERE]

The largest probabilities in Figure 6 are those that are common over the two models and occur at 1951Q3, 1973Q1, 1991Q3 and 2008Q4 with a smaller spike in 1961Q2. These dates loosely align with significant economic events, usually following economic recoveries. Figure 7 plots the probabilities against regions declared to be recessions according the NBER dates. The spikes in 1961, 1991 and 2008 appear at the end of recessions while the spike in 1951 does not appear close to a peak or trough and the spike in 1973 precedes the next downturn by 9 months. It is difficult to conclude much of a systematic relationship between recessions and changes in inflation volatility regimes from these results except that the changes in inflation volatility appear associated with peaks or troughs in growth. These changes affect more the level of inflation volatility than its persistence. The volatility level can increase or decrease after a change but the persistence has shown a general increase until 2008.

We conclude the empirical section with an alternative, and somewhat interesting, perspective on the regime-switches which appears when we compare them with the

[8] We maintain the restriction $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2$ in all cases considered so, for example, $\underline{\alpha}_\lambda = 30$ implies $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30$. For notational brevity and presentation we simply report $\underline{\alpha}_\lambda = 30$.

terms of each of the Federal Chairperson also shown in Figure 7. The change in the inflation volatility regime appears to occur a few years after a new Chairperson takes office and the relationship is more consistent than that with the recessions. We can only make coincidental observations from our model but we feel that the cause of these break points does deserve further investigation.

[FIGURE 7 HERE]

5 Conclusion

This paper has considered the relative advantages and disadvantages of the random walk and stationary specifications of inflation volatility and introduced a new change-point model of log inflation volatility that captures some desirable features of a model for this process. The new model ensures that inflation volatility is bounded in probability while permitting infrequent but large changes in the volatility level and persistence; both of which are frequently discussed features of volatility over the past forty years. A comparison of estimated volatility from a range of models suggests that the specification matters little for this purpose. While information criteria show some preference for the single regime stationary model over the random walk model, the stationary model will produce estimates that approximate the random walk process. On the debate over which specification is appropriate, if the objective is estimation of volatility then inflation either would seem appropriate.

The change-point model of log volatility provides new insights on the evolution of this process. The pattern of volatility shows the often observed decline from the 1980s to the 1990s, but also indicates a rise over the 2000s. Inflation persistence generally increases over most of the sample. Using an empirical Bayes approach, we estimate regime-switching probabilities that align with periods near economic slowdowns, and interestingly, tend to follow changes of the Federal Reserve Chair.

References

Ang A., G. Bekaert and M. Wei (2007) "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?", *Journal of Monetary Economics* 54: 1163-1212.

Berg A., R. Meyer and J. Yu (2004) "Deviance Information Criterion for Comparing Stochastic Volatility Models", *Journal of Business & Economic Statistics* 22(1): 107-120.

Billingsley, P. (1979) Probability and Measure. Wiley, New York.

Chan J. C. C. (2013) "Moving Average Stochastic Volatility Models with Application to Inflation Forecast", *Journal of Econometrics* 176(2): 162–172.

Chan J. C. C. and R. Strachan (2012) "Efficient estimation in non-linear non-Gaussian state-space models", Working paper, Research School of Economics, Australian National University.

Chan J. C. C., G. Koop, and S. M. Potter (2013) "A new model of trend inflation", *Journal of Business & Economic Statistics* 31(1): 94–106.

Chan J. C. C. and I. Jeliazkov (2009) "Efficient simulation and integrated likelihood estimation in state space models", *International Journal of Mathematical modelling and Numerical Optimisation* 1: 101–120.

Clark T. E. and T. Doh (2011) "A Bayesian Evaluation of Alternative Models of Trend Inflation", Federal Reserve Bank of Cleveland, working paper no. 11-34.

Cogley T. and T. Sargent (2001) "Evolving post-World War II inflation dynamics", *NBER Macroeconomic Annual* 16: 331-373.

Cogley T. and T. Sargent (2005) "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.", *Review of Economic Dynamics* 8, 262-302.

Croushore D. (2010) "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data", *The B.E. Journal of Macroeconomics*, De Gruyter, May 10(1): 1-32.

D'Agostino A., Gambetti, L. and Giannone, D. (2009) "Macroeconomic forecasting and structural change", ECARES working paper 2009-020.

Geweke J. (1991) "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints", In E. M. Keramidas, editor, *Computer Science and Statistics, Proceedings of the Twenty-Third Symposium on the Interface*, pages 571–578. Fairfax: Interface Foundation of North America, Inc..

King R. and Watson, M. (1994) "The post-war U.S. Phillips curve: a revisionist econometric history", *Carnegie-Rochester Conference Series on Public Policy*, 41, 157-219.

S. Kim, N. Shephard and S. Chib (1998) "Stochastic volatility: likelihood inference and comparison with ARCH models", *The Review of Economic Studies*, 65(3): 361–393.

Koop G., R. Léon-González and R. W. Strachan (2009) "On the Evolution of Monetary Policy", *Journal of Economic Dynamics and Control* 33, 997-1017.

Koop G., R. Léon- González and R. W. Strachan (2010) "Dynamic probabilities of restrictions in state space models: An application to the Phillips curve", *Journal of Business and Economic Statistics* Vol. 28, No. 3: 370-379.

Koop G., R. Léon-González and R. W. Strachan (2011) "Bayesian Inference in the Time Varying Cointegration Model", *The Journal of Econometrics* 165, 210-220.

Koop G. and S. M. Potter (2006) "Estimation and forecasting in models with multiple breaks", Working Paper.

Koop G. and S. M. Potter (2007) "Estimation and forecasting in models with multiple breaks", *Review of Economic Studies* 74: 763–789.

McCausland W. J., S. Millera, and D. Pelletier (2011) "Simulation smoothing for state-space models: A computational efficiency analysis", *Computational Statistics and Data Analysis* 55: 199–212.

Primiceri G. (2005) “Time varying structural vector autoregressions and monetary policy”, *Review of Economic Studies* 72: 821-852.

Robert C. P. (2005) “Simulation of truncated normal variables”, *Statistics and Computing* 5(2): 121–125.

Rue H. (2001) “Fast sampling of Gaussian Markov random fields with applications”, *Journal of the Royal Statistical Society Series B*, 63: 325-338.

Sargent T., Williams, N. and Zha, T. (2006) “Shocks and government beliefs: The rise and fall of American inflation”, *American Economic Review* 96: 1193-1224.

Staiger D., Stock, J. and Watson, M. (1997) “The NAIRU, unemployment and monetary policy”, *Journal of Economic Perspectives* 11: 33-49.

Spiegelhalter D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002), “Bayesian Measures of Model Complexity and Fit” (with discussion), *Journal of the Royal Statistical Society, B* 64: 583-639.

Stern S. (1992) “A method for smoothing simulated moments of discrete probabilities in multinomial probit models”, *Econometrica* 60(4): 943–952.

Stock J. H. and M. W. Watson (2007) “Has Inflation Become Harder to Forecast?”, *Journal of Money, Credit and Banking* 39: 3-34.

Stock J. H. and M. W. Watson (2009), “Phillips Curve Inflation Forecasts,” in *Understanding Inflation and the Implications for Monetary Policy*, eds., J. Fuhrer, Y. Kodrzycki, J. Little, and G. Olivei, Cambridge: MIT Press, ch. 3: 99–202.

Wright J. H. (2013) “Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior”, *Journal of Applied Econometrics* 28: 762-776.

Appendix: Gibbs sampling of the regime-switching volatility model

Starting with the general state-space representation

$$\begin{aligned} y_t &= \mu_t + \exp{(h_t/2)}\,\epsilon_t, & \epsilon_t &\sim \mathcal{N}(0,1), \\ h_t &= \eta_{s_t} - \rho_{s_t}(h_{t-1} - \eta_{s_{t-1}}) + \nu_t, & \nu_t &\sim \mathcal{N}(0,\sigma^2_{h,s_t}), \end{aligned}$$

we consider the steps needed to sample

$$(h, h_0, s, \lambda, \beta_\lambda, \eta, \eta_0, \sigma^2_\eta, \rho, \rho_0, \sigma^2_\rho, \sigma^2_h, \sigma^2_{h,0}, \sigma^2_\sigma \,|\, \mu, y).$$

For completeness, define $\lambda = (\lambda_1, \ldots, \lambda_M)'$ and $\mu = (\mu_1, \ldots, \mu_T)'$. We assume the latter is obtained by an appropriate sampling of the mean equation conditional on the volatilities h. The algorithm for sampling the volatilities and the related hyper-parameters proceeds as follows:

1. Sample $(h, h_0 \,|\, s, \eta, \rho, \sigma^2_h, \mu, y)$ by first re-writing the joint prior as

$$(h_0, h) \sim \mathcal{N}\left(H_h^{-1}\zeta_h, \left(H_h'^{-1}H_h\right)^{-1}\right), \tag{11}$$

where $\zeta_h = (\eta_1, (1-\rho_1)\eta_1, \eta_2-\rho_2\eta_1, \ldots, \eta_{s_T}-\rho_{s_T}\eta_{s_{T-1}})'$, $\Omega = \mathrm{diag}(V_{h_0}, \sigma^2_{h,1}, \ldots, \sigma^2_{h,s_t})$, and

$$H_h = \begin{pmatrix} 1 & & & \\ -\rho_1 & 1 & & \\ & \ddots & \ddots & \\ & & -\rho_{s_T} & 1 \end{pmatrix}.$$

Further letting $y_t^* = \ln((y_t - \mu_t)^2 + c)$ and $y^* = (y_1^*, \ldots, y_T^*)'$, where c is a small offset constant (e.g. in our empirical work, we set $c = 10^{-4}$), the volatility state equation may be expressed as

$$y^* = h + \widetilde{\epsilon}, \tag{12}$$

where the distribution of $\widetilde{\epsilon}_t = \ln \epsilon_t^2$ is closely approximated by a seven-point mixture of normal distributions (e.g. see Kim *et al.* (1998) for details). Let θ_j be the mean and ψ_j^2 be the variance of the jth component, and let g_t denote a draw of the component in period t. Defining further $\theta = (\theta_{g_1}, \ldots, \theta_{g_T})'$, $\Psi = \text{diag}(\psi_{g_1}^2, \ldots, \psi_{g_T}^2)$, one proceeds by first drawing a vector $g = (g_1, \ldots, g_T)'$ of components conditional on h and other parameters (e.g. as described in Kim *et al.* (1998)), followed by sampling (h_0, h) conditional on the components from

$$\begin{aligned}
(h_0, h) &\sim \mathcal{N}(\overline{h}, \overline{V}_h), \\
\overline{h} &= \overline{V}_h \left(H_h'^{-1} \zeta_h + \begin{pmatrix} 0 \\ \Psi^{-1} \end{pmatrix} (y^* - \theta) \right), \\
\overline{V}_h &= \left(H_h'^{-1} H_h + \begin{pmatrix} 0 & \\ & \Psi^{-1} \end{pmatrix} \right)^{-1}.
\end{aligned} \tag{13}$$

2. Sample the regimes by first obtaining a draw

$$\beta_\lambda \sim \mathcal{G}\left(\underline{\xi}_1 + M \underline{\alpha}_\lambda, \underline{\xi}_2 + \sum_{m=1}^{M} \lambda_m \right). \tag{14}$$

Next, sample

$$(s, \lambda \,|\, \beta_\lambda, h, h_0, \eta, \rho, \sigma_h^2)$$

by first drawing s marginal of λ, followed by $(\lambda \mid s)$. The algorithm for sampling the regime indicators s is an extension of Chib's (1996) algorithm and is described in more detail in the Koop and Potter (2006) working paper. Here, we only present the practical aspects of its implementation in the context of our regime-switching volatility model. In particular, recall from Koop and Potter (2007) that the hierarchical prior on regime durations implies a Markov process for the evolution of regimes, where a Markov state is the pair $(s_t, \widetilde{d}_{s_t})$—i.e., a regime number and a partial duration of that regime (up to time t). Indexing a Markov state by i, such that

$$i = (m-1)(T+1) - \sum_{n=1}^{m-1} n + \widetilde{d}_m,$$

we will denote the reverse mapping as $(m(i), \widetilde{d}_m(i))$. The transition probabilities matrix P, therefore, is $L \times L$, with $L = M(T+1) - \sum_{m=1}^{M} m$, and contains the elements

$$p_{i,j} = \begin{cases} \pi_i & \text{if} \quad m(j) = m(i)+1, \widetilde{d}_m(j) = 1, j \neq i+1 \\ 1-\pi_i & \text{if} \quad j = i+1, \widetilde{d}_m(j) \neq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{15}$$

In this case, π_i represents the probability of switching to regime $m(i)+1$ at period $t+1$ given that $m(i)$ prevailed in period t. We compute it as

$$\begin{aligned} \pi_i &= 1 - \Pr\left(d_{m(i)} = \widetilde{d}_m(i) + 1 \mid d_{m(i)} \geq \widetilde{d}_m(i)\right) = (1+\pi_{i,0})^{-1}, \\ \pi_{i,0} &= \frac{1 - F_{NB}\left(\widetilde{d}_m(i) - 1; \underline{\alpha}_\lambda, \beta_\lambda/(1+\beta_\lambda)\right)}{p_{NB}\left(\widetilde{d}_m(i) - 1; \underline{\alpha}_\lambda, \beta_\lambda/(1+\beta_\lambda)\right)}, \end{aligned} \tag{16}$$

where $p_{NB}(\ \cdot\ )$ denotes the Negative Binomial pmf and $F_{NB}(\ \cdot\ )$ denotes the negative binomial cdf. Note that the above formulation uses the fact that λ_m can be integrated out analytically to yield Negative Binomial priors on $d_1, \ldots, d_m$, which remain independent conditional on β_λ. Moreover, for all i corresponding to $m(i) = M$, we set $\pi_i = 1$.

Next, construct a series of $L \times 1$ vectors $F_1, \ldots, F_T$ recursively. That is, starting with $F_1 = (1, 0, \ldots, 0)$, compute for $t = 2, \ldots, T$

$$\widetilde{F}_t = P' F_{t-1} \odot z_t,$$
$$F_t = \widetilde{F}_t / \sum_{j=1}^{L} \widetilde{F}_{t,j},$$

where $\odot$ denotes element-wise multiplication and z_t is a $L \times 1$ vector of state equation evaluations such that

$$z_{t,i} = \begin{cases} \phi\left(h_t; \eta_{m(i)} + \rho_{m(i)}(h_{t-1} - \eta_{m(i)-1}), \sigma^2_{h,m(i)}\right) & \text{if} \quad \widetilde{d}_m(i) = 1 \\ \phi\left(h_t; \eta_{m(i)} + \rho_{m(i)}(h_{t-1} - \eta_{m(i)}), \sigma^2_{h,m(i)}\right) & \text{if} \quad \widetilde{d}_m(i) > 1 \end{cases} \tag{17}$$

where $\phi(x_t, m, v)$ is a normal pdf with mean m and variance v. Now, F_T contains the pmf for the conditional distribution of $(s_T, \widetilde{d}_{s_T})$. A draw from this distribution, therefore, locates both the final in-sample regime number m^* as well as the last in-sample change-point τ_{m^*-1}. Thus, for all $\tau_{m^*-1} \leq t \leq T$, set $s_t = m^*$. The object $\varpi = P_{m^*} \odot F_{\widetilde{t}}$, where P_j denotes the jth column of P, is a vector of weights for the distribution of the pair $(s_{\widetilde{t}}, \widetilde{d}_{s_{\widetilde{t}}})$. Subsequently, in period $\widetilde{t} = \tau_{m^*-1} - 1$, a draw of $(s_{\widetilde{t}}, \widetilde{d}_{s_{\widetilde{t}}})$ is obtained from the distribution ϖ. Proceeding this way until τ_1 is reached produces a sample of $s_1, \ldots, s_T$.

A final remark on the above procedure is that P will typically be a very large, but sparse matrix (the same applies, albeit to a lesser extent, to F_t as well). For example, in our empirical application with $T = 261$ and $M = 30$ we have $L = 7395$, and in consequence, P contains over 54 million elements. Even storing a full matrix of this magnitude is difficult on a typical personal computer; performing a simple operation such multiplication is at best impractical. One must therefore take care in utilizing appropriate sparse matrix routines in the course of the regime-switching algorithm. In our case, for example, only a maximum of $2(L-M) = 14,730$ of the elements in P can be non-zero, and this number is much lower in practice since many transition probabilities (e.g. π_i) are also set to zero. Working with sparse matrix routines provides the necessary efficiency for this algorithm to be operational.

The sampling of the regimes is completed by drawing λ in two steps. First sample

$$\lambda_{\underline{m}} \sim \mathcal{G}(\underline{\alpha}_\lambda + d_{\underline{m}} - 1, \beta + 1), \qquad \underline{m} = 1, \ldots, m^* - 1, \tag{18}$$

$$\lambda_{\overline{m}} \sim \mathcal{G}(\underline{\alpha}_\lambda, \beta), \qquad \overline{m} = m^* + 1, \ldots, M, \tag{19}$$

where $d_{\underline{m}}$ are obtained from the draws of $s_1, \ldots, s_T$. To draw λ_{m^*} we employ an M-H step as suggested in Koop and Potter (2006). Given a previous draw of λ_{m^*}, the candidate $\lambda^c_{m^*}$ is drawn independently from the prior

$$\lambda^c_{m^*} \sim \mathcal{G}(\underline{\alpha}_\lambda, \beta), \tag{20}$$

and accepted with probability

$$\Pr\left(\text{accept } \lambda_{m^*}^c\right) = 1 - \frac{1 - F_{Po}\left(\widetilde{d}_{m^*} - 2; \lambda_{m^*}^c\right)}{1 - F_{Po}\left(\widetilde{d}_{m^*} - 2; \lambda_{m^*}\right)}, \tag{21}$$

where $F_{Po}(\ \cdot\ )$ denotes the cdf of the Poisson distribution and $\widetilde{d}_{m^*}$ is the partial duration of the final in-sample regime.

3. Sample η, η_0 and σ_η^2 in two blocks: $(\eta_0, \eta \,|\, \sigma_\eta^2)$ followed by $(\sigma_\eta^2 \,|\, \eta_0, \eta)$. To obtain a draw of the former, begin with the joint prior:

$$(\eta_0, \eta) \sim \mathcal{N}\left(\kappa_{\eta_0} \iota_{M+1}, \sigma_\eta^2 \left(H_\eta' H_\eta\right)^{-1}\right), \tag{22}$$

where H_η is the $M+1 \times M+1$ versions of the matrix H defined in the text (i.e. with 1 on the main diagonal and -1 on the first lower diagonal). Furthermore, the volatility state equation can be expressed as

$$h^* = W\eta + \nu, \tag{23}$$

where $h^* = H_h h$, W is a $T \times M$ matrix with elements

$$w_{t,m} = \begin{cases} 1 - \rho_m & \text{if} \quad s_t = s_{t-1} = m \\ -\rho_m & \text{if} \quad s_t = m+1, s_{t-1} = m \\ 1 & \text{if} \quad s_t = m, s_{t-1} = m-1 \\ 0 & \text{otherwise} \end{cases}, \tag{24}$$

and $s_0 = 1$ by definition. Also, define $\Sigma_\eta = \text{diag}(V_{\eta_0}, \sigma_h^{2\prime})$ and the augmented matrix $\widetilde{W} = (0, W)$. Accordingly, a draw of (η_0, η) is obtained from

$$
\begin{aligned}
(\eta_0, \eta) &\sim \mathcal{N}\left(\overline{\eta}, \overline{V}_\eta\right), \\
\overline{\eta} &= \overline{V}_\eta \left(\frac{\kappa_{\eta_0}}{V_{\eta_0}} \widetilde{\iota}_{M+1} + \widetilde{W}' \widetilde{\Sigma}_h^{-1} h^* \right) \\
\overline{V}_\eta &= \left(H_\eta' \Sigma_\eta^{-1} H_\eta + \widetilde{W}' \widetilde{\Sigma}_h^{-1} \widetilde{W} \right)^{-1}.
\end{aligned}
\tag{25}
$$

Finally, define $\overline{\xi}_\eta = (\eta_1 - \eta_0 - \kappa_{\eta_0}, \eta_2 - \eta_1, \ldots, \eta_M - \eta_{M-1})'$ and sample

$$
\sigma_\eta^2 \sim \mathcal{IG}\left(\gamma_e ta + M/2, \delta_\eta + \overline{\xi}_\eta' \overline{\xi}_\eta / 2\right). \tag{26}
$$

4. Sample ρ, ρ_0 and σ_ρ^2 in three blocks: $(\rho, \mid \rho, \sigma_\rho^2)$, $(\rho_0, \mid \rho_0, \sigma_\rho^2)$ and $(\sigma_\rho^2 \mid \rho_0, \rho)$. The conditional sampling of ρ is discussed in detail in Section 3 of the text. In our empirical application we use an adaptive algorithm where we first attempt to obtain a draw of ρ directly from (5) with brute force accept-reject, up to a maximum of c^* attempts. If the latter step fails, then a draw of ρ is obtained with the augmented step in (7)-(10). The potential advantage of the adaptive approach is that whenever the accept-reject step succeeds, the Markov chain partially *regenerates* in the sense that all future draws of (f, ρ) are no longer correlated with any previous draws of f. The drawback, of course, is that whenever the accept-reject step fails, the c^* attempts are a computational waste.

   In our experience, setting $c^* = 1000$ appears to improve mixing in models with shorter expected durations (e.g. $\underline{\alpha}_\lambda = 15, \underline{\alpha}_\lambda = 30$), but does not provide a noticeable advantage in models with longer expected durations (e.g. $\underline{\alpha}_\lambda = 60, \underline{\alpha}_\lambda = 120$). This is because—inline with the discussion in Section 3—in models with shorter durations, m^* is higher with a larger posterior probability, and therefore, over the course of the MCMC sampler, m^* achieves (with some frequency) values sufficiently close to M, which reduces the probability that $|\rho| < \iota_M$ is binding. In models with longer expected durations, on the other hand, m^* is low relative to M with a high probability and the accept-reject algorithm fails for most MCMC iterations, making the approach wasteful.

Given ρ, proceed by sampling

$$\rho_0 \sim \mathcal{N}\left(\frac{\overline{V}_{\rho_0}}{\sigma^2_\rho}\rho_1, \overline{V}_{\rho_0}\right), \tag{27}$$

$$\sigma^2_\rho \sim \mathcal{IG}\left(\gamma_\rho + M/2, \delta_\rho + \overline{\xi}'_\rho\overline{\xi}_\rho/2\right), \tag{28}$$

$$\overline{V}_{\rho_0} = \left(\frac{1}{V_{\rho_0}} + \frac{1}{\sigma^2_\rho}\right)^{-1},$$

$$\overline{\xi}_\rho = (\rho_1 - \rho_0 - \kappa_{\rho_0}, \rho_2 - \rho_1, \ldots, \rho_M - \rho_{M-1})'.$$

5. Sample $\ln\sigma^2_h$, $\ln\sigma^2_{h,0}$ and σ^2_σ in two blocks: $(\ln\sigma^2_{h,0}, \ln\sigma^2_h \mid \sigma^2_\sigma)$ followed by $(\sigma^2_\sigma, \mid \ln\sigma^2_{h,0}, \ln\sigma^2_h)$. To sample the first block, denote as in the previous steps

$$(\ln\sigma^2_{h,0}, \ln\sigma^2_h) \sim \mathcal{N}\left(\kappa_{\sigma_0}\iota_{M+1}, \sigma^2_\sigma\left(H'_\sigma H_\sigma\right)^{-1}\right), \tag{29}$$

$$h^\dagger = S\ln\sigma^2_h + \widetilde{\nu}, \tag{30}$$

where $H_\sigma = H_\eta$, $h^\dagger_t = \ln((h_t - \rho_{s_t}h_{t-1} - \zeta_{h,t+1})^2 + c)$ and S is a $T \times M$ matrix with column

$$S_m = (0, \ldots, 0, s_{\tau_{m-1}}, \ldots, s_{\tau_m - 1}, 0, \ldots, 0)'.$$

The procedure for sampling $(\ln\sigma^2_{h,0}, \ln\sigma^2_h)$ follows analogously from that described in Step 1 above for sampling the log volatilities. The remaining block is sampled in a straightforward way as

$$\sigma_\sigma^2 \sim \mathcal{IG}\left(\gamma_\sigma + M/2, \delta_\sigma + \overline{\xi}_\sigma' \overline{\xi}_\sigma/2\right), \tag{31}$$

$$\overline{\xi}_\sigma = (\ln \sigma_{h,1}^2 - \ln \sigma_{h,0}^2 - \kappa_{\sigma_0}, \ln \sigma_{h,2}^2 - \ln \sigma_{h,1}^2, \ldots, \ln \sigma_{h,M}^2 - \ln \sigma_{h,M-1}^2)'.$$

Tables and Figures

Model	DIC
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 15$	983
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30$	983
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 60$	984
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 120$	986
Single regime stationary model	986
Random Walk model	991

Table 1: DIC for six models

Model	2008Q3-2010Q2	2011Q3-2013Q2
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 15$	6.60×10^{-14}	3.64×10^{-07}
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30$	5.69×10^{-14}	3.48×10^{-07}
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 60$	5.63×10^{-14}	4.67×10^{-07}
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 120$	5.48×10^{-14}	4.93×10^{-07}
Single regime stationary model	3.55×10^{-14}	6.40×10^{-07}
Random Walk model	1.40×10^{-13}	3.33×10^{-07}

Table 2: A comparison of joint predictive likelihoods for 2008Q3–2010Q2 and 2011Q3–2013Q2

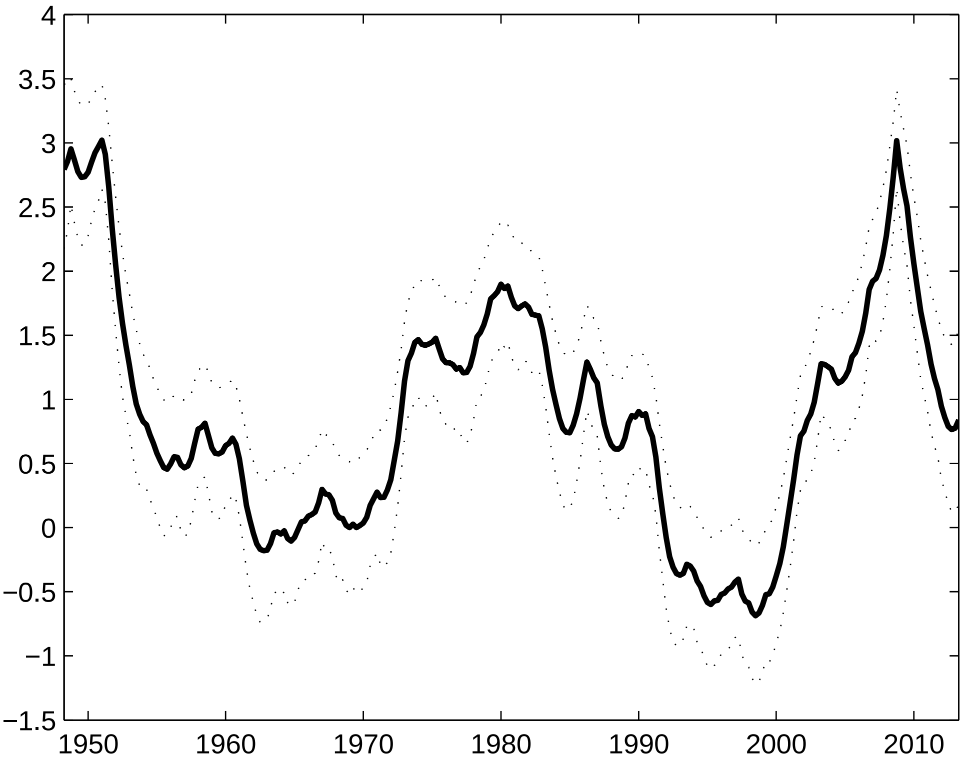
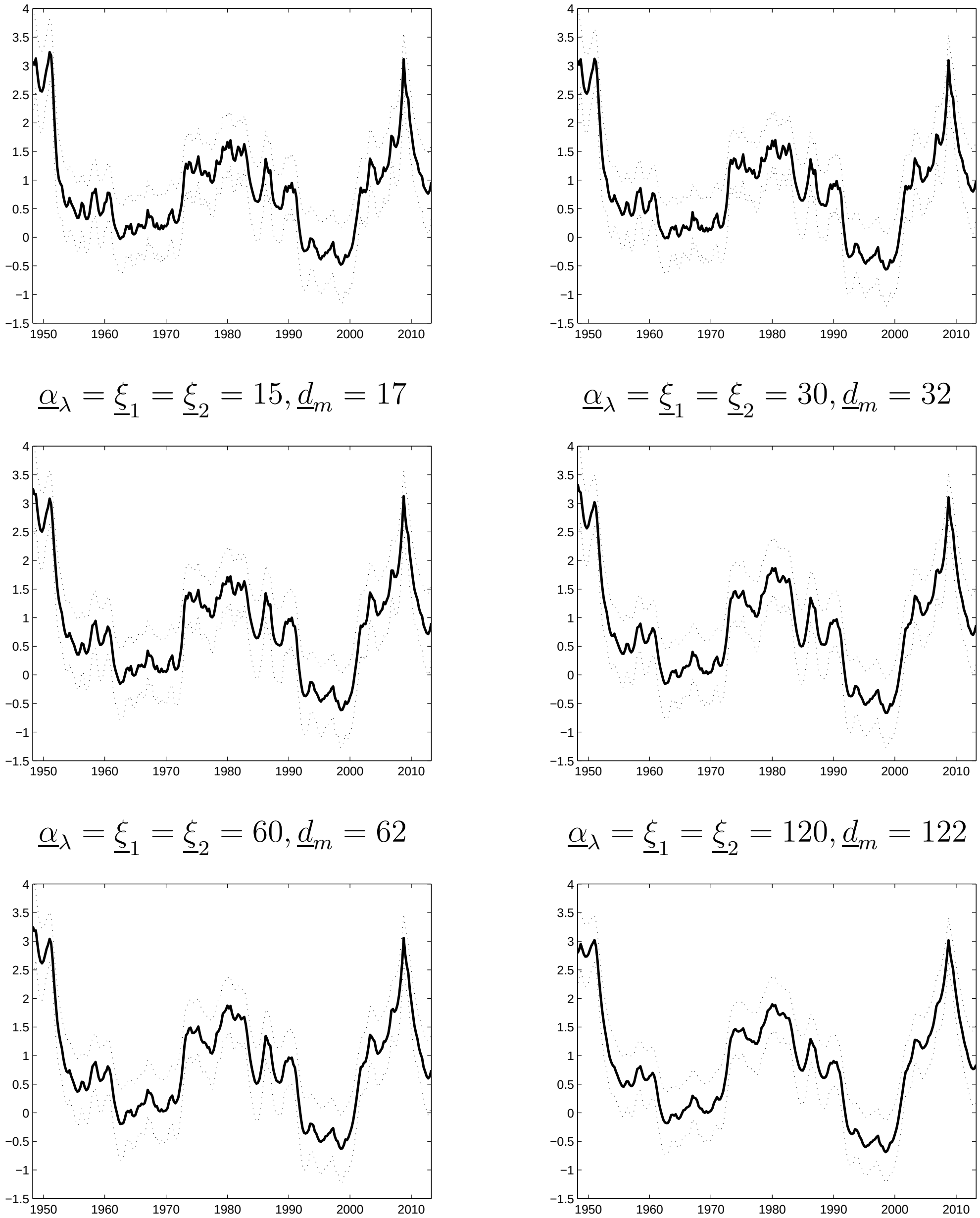


Figure 1: Plot of inflation volatility from 1948 to 2013.

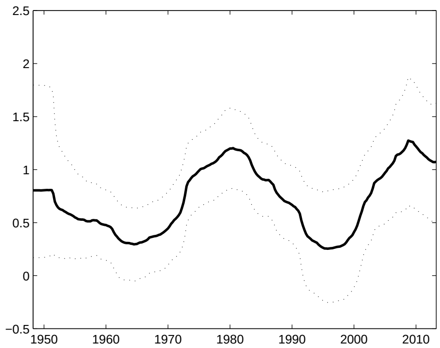


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 15, \underline{d}_m = 17$ $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30, \underline{d}_m = 32$

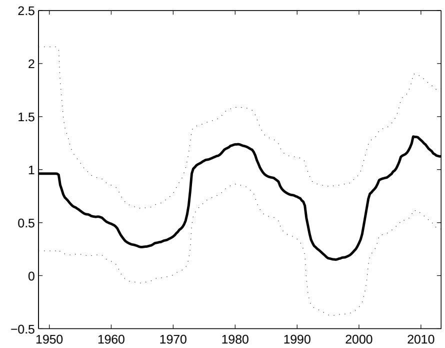
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 60, \underline{d}_m = 62$ $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 120, \underline{d}_m = 122$

Single stationary regime ($\underline{d}_m \geq T$) Random Walk ($\underline{d}_m = T$)

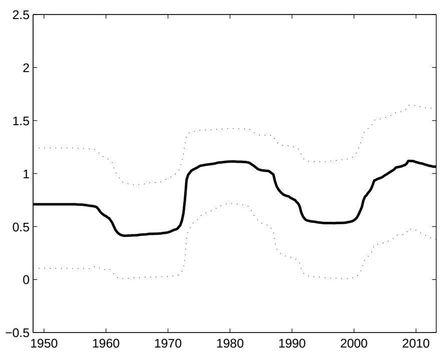
Figure 2: Posterior median and the $(16\%, 84\%)$ probability interval for the log-volatility h_t. Note: for the *single stationary regime* case (lower left), we get $\hat{\rho} = 0.94$, which is close to a "random walk".



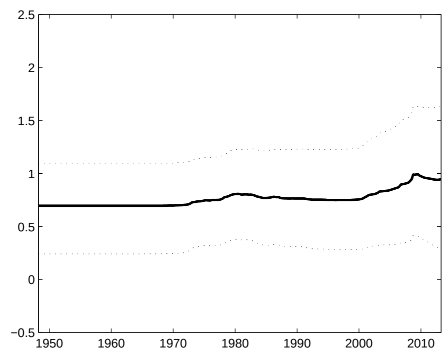
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 15, \underline{d}_m = 17$



$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30, \underline{d}_m = 32$

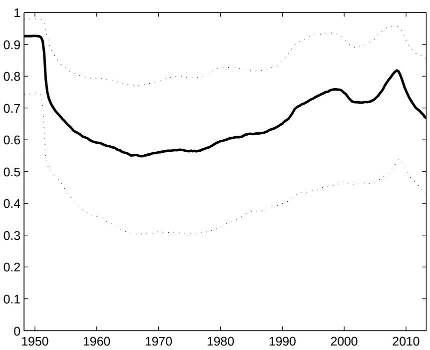


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 60, \underline{d}_m = 62$

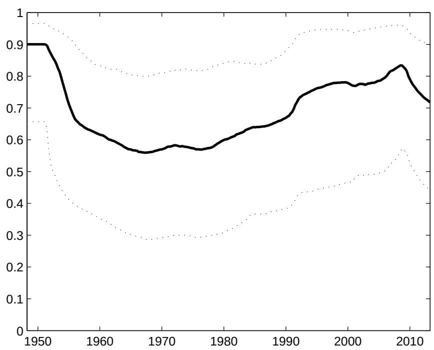


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 120, \underline{d}_m = 122$

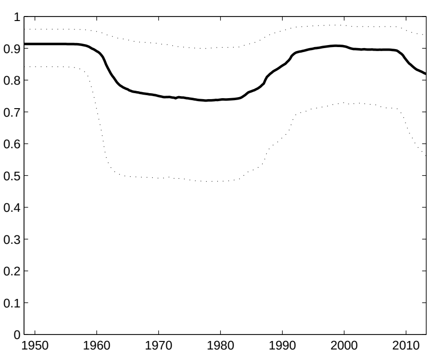
Figure 3: Posterior median and the $(16\%, 84\%)$ probability interval for η_m.



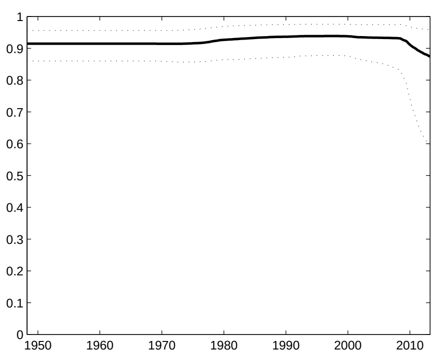
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 15, \underline{d}_m = 17$



$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30, \underline{d}_m = 32$

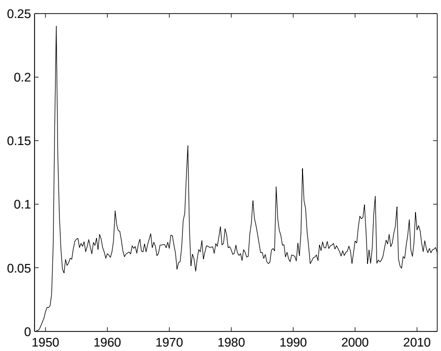


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 60, \underline{d}_m = 62$

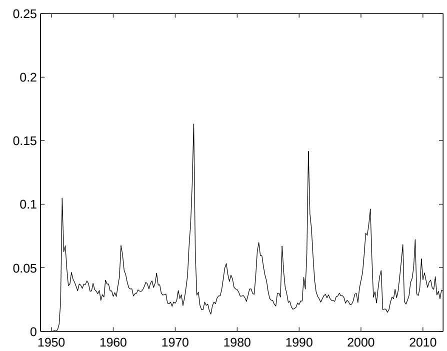


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 120, \underline{d}_m = 122$

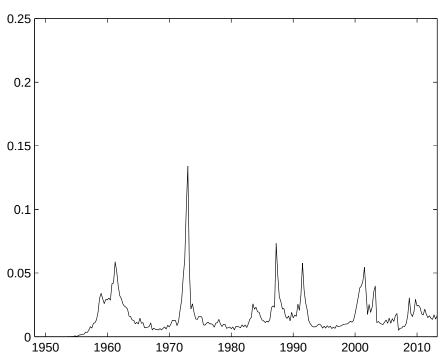
Figure 4: Posterior median and the $(16\%, 84\%)$ probability interval for ρ_m.



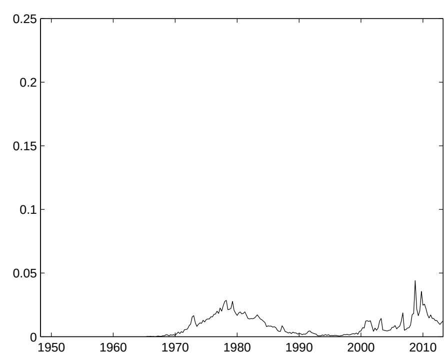
$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 15, \underline{d}_m = 17$



$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30, \underline{d}_m = 32$

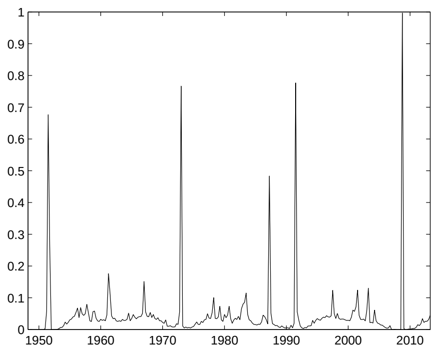


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 60, \underline{d}_m = 62$

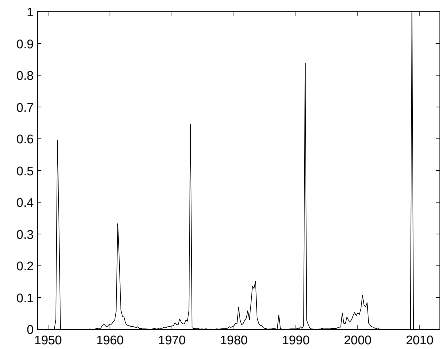


$\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 120, \underline{d}_m = 122$

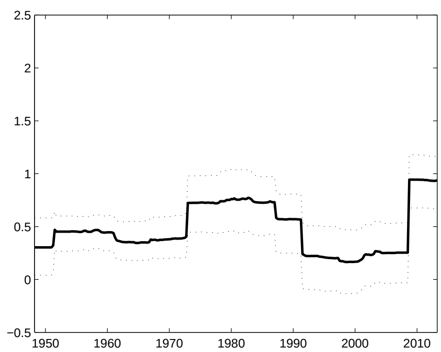
Figure 5: Change-point probabilities: probability of a regime change in each period t.



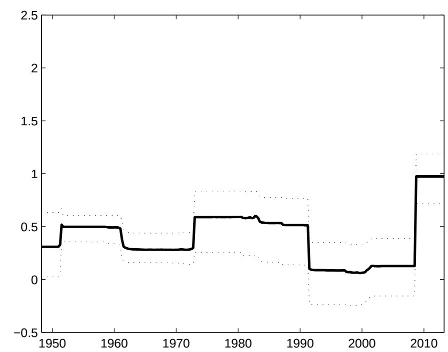
Regime-change probability ($\underline{\alpha}_\lambda = 15$)



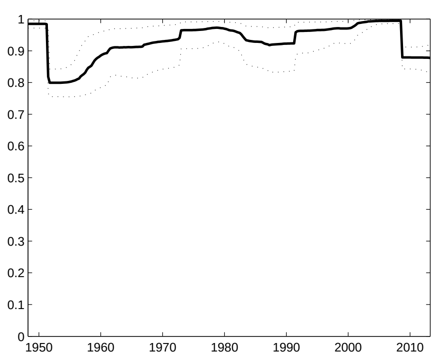
Regime-change probability ($\underline{\alpha}_\lambda = 60$)



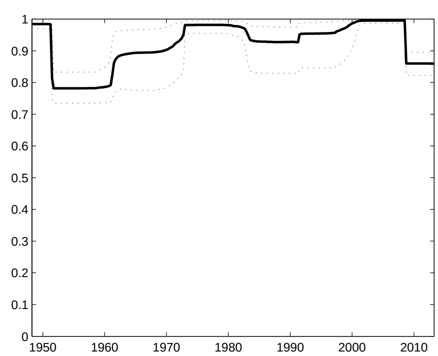
Evolution of η_m ($\underline{\alpha}_\lambda = 15$)



Evolution of η_m ($\underline{\alpha}_\lambda = 60$)



Evolution of ρ_m ($\underline{\alpha}_\lambda = 15$)



Evolution of ρ_m ($\underline{\alpha}_\lambda = 60$)

Figure 6: Posterior estimates conditional on $h = E(h \,|\, y, \underline{\alpha}_\lambda = 30)$.

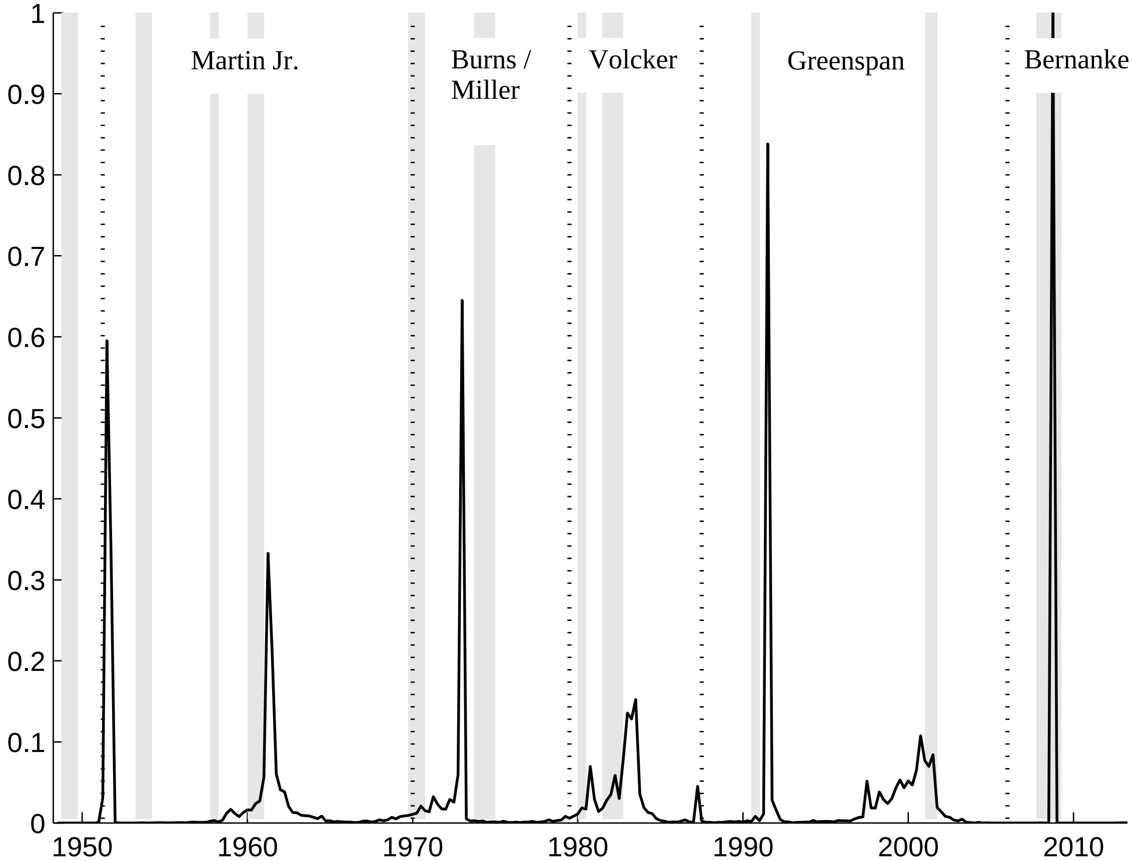


Figure 7: Conditional posterior estimates of the change-point probabilities from 1948 to 2013. Also shown are the reign of Federal Reserve Bank Chairperson (vertical dotted lines) and the NBER recessions (shaded grey bands).