Online Appendix to Modelling Inflation Volatility

Eric Eisenstat*
University of Bucharest and RIMIR

Rodney W. Strachan
University of Queensland

October, 2014

A Gibbs Sampling of the Change-Point Process Volatility Model

In this section we outline the Gibbs sampling algorithm used to obtain draws of log volatilities and related hyper-parameters for the change-point process, conditional on the mean equation. Matlab code is available at `http://www.rimir.ro/eric/research.html`. Starting with the general state-space representation

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

we consider the steps needed to sample

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

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

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

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

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

*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.

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

$$y^* = h + \widetilde{\varepsilon}, \tag{2}$$

where the distribution of $\widetilde{\varepsilon}_t = \ln \varepsilon_t^2$ 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 $\theta = (\theta_{g_1}, \ldots, \theta_{g_T})'$, $\Psi = \mathrm{diag}(\psi_{g_1}^2, \ldots, \psi_{g_T}^2)$, 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

$$\begin{aligned}(h_0, h) &\sim \mathcal{N}(\overline{h}, \overline{V}_h), \\ \overline{h} &= \overline{V}_h \left(H_h' \Omega^{-1} \zeta_h + \begin{pmatrix} 0 \\ \Psi^{-1} \end{pmatrix} (y^* - \theta) \right), \\ \overline{V}_h &= \left(H_h' \Omega^{-1} H_h + \begin{pmatrix} 0 & \\ & \Psi^{-1} \end{pmatrix} \right)^{-1}. \end{aligned} \tag{3}$$

2. Sample the regimes by first obtaining a draw

$$\beta_\lambda \sim \mathcal{G}\left(\underline{\xi}_1 + M \underline{\alpha}_\lambda, \underline{\xi}_2 + \sum_{m=1}^{M} \lambda_m \right). \tag{4}$$

Next, sample

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

by first drawing s marginal of λ, followed by $(\lambda \,|\, s)$. The algorithm for sampling the regime indicators s is an extension of the Chib [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

$$p_{i,j} = \begin{cases} \pi_i & \text{if} \quad m(j) = m(i) + 1, \widetilde{d}_m(j) = 1, j \neq i+1 \\ 1 - \pi_i & \text{if} \quad j = i+1, \widetilde{d}_m(j) \neq 1 \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

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

$$\pi_i = 1 - \Pr\left(d_{m(i)} = \widetilde{d}_m(i) + 1 \,|\, d_{m(i)} \geqslant \widetilde{d}_m(i)\right) = (1 + \pi_{i,0})^{-1},$$
$$\pi_{i,0} = \frac{1 - F_{NB}\left(\widetilde{d}_m(i) - 1; \underline{\alpha}_\lambda, \beta_\lambda/(1+\beta_\lambda)\right)}{p_{NB}\left(\widetilde{d}_m(i) - 1; \underline{\alpha}_\lambda, \beta_\lambda/(1+\beta_\lambda)\right)}, \tag{6}$$

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

$$z_{t,i} = \begin{cases} \phi\left(h_t; \eta_{m(i)} + \rho_{m(i)}(h_{t-1} - \eta_{m(i)-1}), \sigma^2_{h,m(i)}\right) & \text{if} \quad \widetilde{d}_m(i) = 1 \\ \phi\left(h_t; \eta_{m(i)} + \rho_{m(i)}(h_{t-1} - \eta_{m(i)}), \sigma^2_{h,m(i)}\right) & \text{if} \quad \widetilde{d}_m(i) > 1 \end{cases} \tag{7}$$

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 $\tau_{m^*-1} \leqslant t \leqslant T$, 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.[1]

[1] In our implementation using Matlab, we work with matrices constructed by the *sparse* command and related routines.

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

$$\lambda_{\underline{m}} \sim \mathcal{G}(\underline{\alpha}_\lambda + d_{\underline{m}} - 1, \beta + 1), \qquad \underline{m} = 1, \ldots, m^* - 1, \tag{8}$$

$$\lambda_{\overline{m}} \sim \mathcal{G}(\underline{\alpha}_\lambda, \beta), \qquad \overline{m} = m^* + 1, \ldots, M, \tag{9}$$

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

$$\lambda^c_{m^*} \sim \mathcal{G}(\underline{\alpha}_\lambda, \beta), \tag{10}$$

and accepted with probability

$$\Pr\left(\text{accept } \lambda^c_{m^*}\right) = \min\left\{1, 1 - \frac{1 - F_{Po}\left(\widetilde{d}_{m^*} - 2; \lambda^c_{m^*}\right)}{1 - F_{Po}\left(\widetilde{d}_{m^*} - 2; \lambda_{m^*}\right)}\right\}, \tag{11}$$

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^2_η in two blocks: $(\eta_0, \eta \,|\, \sigma^2_\eta)$ followed by $(\sigma^2_\eta \,|\, \eta_0, \eta)$. To obtain a draw of the former, begin with the joint prior:

$$(\eta_0, \eta) \sim \mathcal{N}\left(\kappa_{\eta_0}\iota_{M+1}, \left(H'_\eta \Sigma^{-1}_\eta H_\eta\right)^{-1}\right), \tag{12}$$

where $\Sigma_\eta = \text{diag}(V_{\eta_0}, \sigma^2_\eta \boldsymbol{\iota}'_M)$ and 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

$$h^* = W\eta + \nu, \tag{13}$$

where $h^* = H_h h$, W is a $T+1 \times M$ matrix with elements

$$w_{t,m} = \begin{cases} 1 - \rho_m & \text{if} \quad s_t = s_{t-1} = m \\ -\rho_m & \text{if} \quad s_t = m+1, s_{t-1} = m \\ 1 & \text{if} \quad s_t = m, s_{t-1} = m - 1 \\ 0 & \text{otherwise} \end{cases}, \tag{14}$$

and $s_0 = 1$ by definition. Also, define the augmented matrix $\widetilde{W} = (0, W)$. Accordingly, a draw of (η_0, η) is obtained from

$$\begin{aligned} (\eta_0, \eta) &\sim \mathcal{N}\left(\overline{\eta}, \overline{V}_\eta\right), \\ \overline{\eta} &= \overline{V}_\eta\left(\frac{\kappa_{\eta_0}}{V_{\eta_0}}\widetilde{\iota}_{M+1} + \widetilde{W}'\Omega^{-1}h^*\right) \\ \overline{V}_\eta &= \left(H'_\eta\Sigma^{-1}_\eta H_\eta + \widetilde{W}'\Omega^{-1}\widetilde{W}\right)^{-1}. \end{aligned} \tag{15}$$

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

$$\sigma^2_\eta \sim \mathcal{IG}\left(\gamma_\eta + M/2, \delta_\eta + \overline{\xi}'_\eta\overline{\xi}_\eta/2\right). \tag{16}$$

4. Sample ρ, ρ_0 and σ^2_ρ in three blocks: $(\rho, \,|\, \rho, \sigma^2_\rho)$, $(\rho_0, \,|\, \rho_0, \sigma^2_\rho)$ and $(\sigma^2_\rho \,|\, \rho_0, \rho)$. The conditional sampling of ρ is discussed in detail in Section 3 of the main text. In our empirical application we use an adaptive algorithm where we first attempt to obtain a draw of ρ directly from equation 3 of the main text 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 equations 5-8 of the main text. 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 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

$$\rho_0 \sim \mathcal{N}\left(\frac{\overline{V}_{\rho_0}}{\sigma^2_\rho}\rho_1, \overline{V}_{\rho_0}\right), \tag{17}$$

$$\sigma^2_\rho \sim \mathcal{IG}\left(\gamma_\rho + M/2, \delta_\rho + \overline{\xi}'_\rho \overline{\xi}_\rho/2\right), \tag{18}$$

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

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

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

$$(\ln \sigma^2_{h,0}, \ln \sigma^2_h) \sim \mathcal{N}\left(\kappa_{\sigma_0} \iota_{M+1}, \left(H'_\sigma \Sigma^{-1}_\sigma H_\sigma\right)^{-1}\right), \tag{19}$$

$$h^\dagger = S \ln \sigma^2_h + \widetilde{\nu}, \tag{20}$$

   where $H_\sigma = H_\eta$, $\Sigma_\sigma = \text{diag}(V_{\sigma_0}, \sigma^2_\sigma \boldsymbol{\iota}'_M)$, $h^\dagger_t = \ln((h_t - \rho_{s_t} h_{t-1} - \zeta_{h,t+1})^2 + c)$ 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

$$\sigma^2_\sigma \sim \mathcal{IG}\left(\gamma_\sigma + M/2, \delta_\sigma + \overline{\xi}'_\sigma \overline{\xi}_\sigma/2\right), \tag{21}$$

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

A.1 Estimating the UC-SV Model

In the main text, we discuss adapting the above sampler to estimate the UC-SV specification (see the main text for more details). The extension is trivial:

- Repeat Steps 1, 2-5 twice: once for $(h_{\mu,t}, \eta_{\mu,m}, \rho_{\mu,m}, \sigma^2_{h_\mu,m})$ another for $(h_{y,t}, \eta_{y,m}, \rho_{y,m}, \sigma^2_{h_y,m})$.
- In Step 2, replace (7) with

$$z_{t,i} = \begin{cases} \phi\left(h_{\mu,t}; \eta_{\mu,m(i)} + \rho_{\mu,m(i)}(h_{\mu,t-1} - \eta_{\mu,m(i)-1}), \sigma^2_{h_\mu,m(i)}\right) \times & \text{if} \quad \widetilde{d}_m(i) = 1 \\ \quad \phi\left(h_{y,t}; \eta_{y,m(i)} + \rho_{y,m(i)}(h_{y,t-1} - \eta_{y,m(i)-1}), \sigma^2_{h_y,m(i)}\right) & \\ \phi\left(h_{\mu,t}; \eta_{\mu,m(i)} + \rho_{\mu,m(i)}(h_{\mu,t-1} - \eta_{\mu,m(i)}), \sigma^2_{h_\mu,m(i)}\right) \times & \text{if} \quad \widetilde{d}_m(i) > 1 \\ \quad \phi\left(h_{y,t}; \eta_{y,m(i)} + \rho_{y,m(i)}(h_{y,t-1} - \eta_{y,m(i)}), \sigma^2_{h_y,m(i)}\right) & \end{cases} \tag{22}$$

B Estimation Results Using Core CPI Data

In this section we present results obtained by using core CPI data—i.e., excluding food and energy prices. The model is estimated with $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30$ and the bounded trend mean equation. The estimated log-volatility, level, persistance, and change point probabilities are presented in Figure 1. See the main text for a discussion of these results.

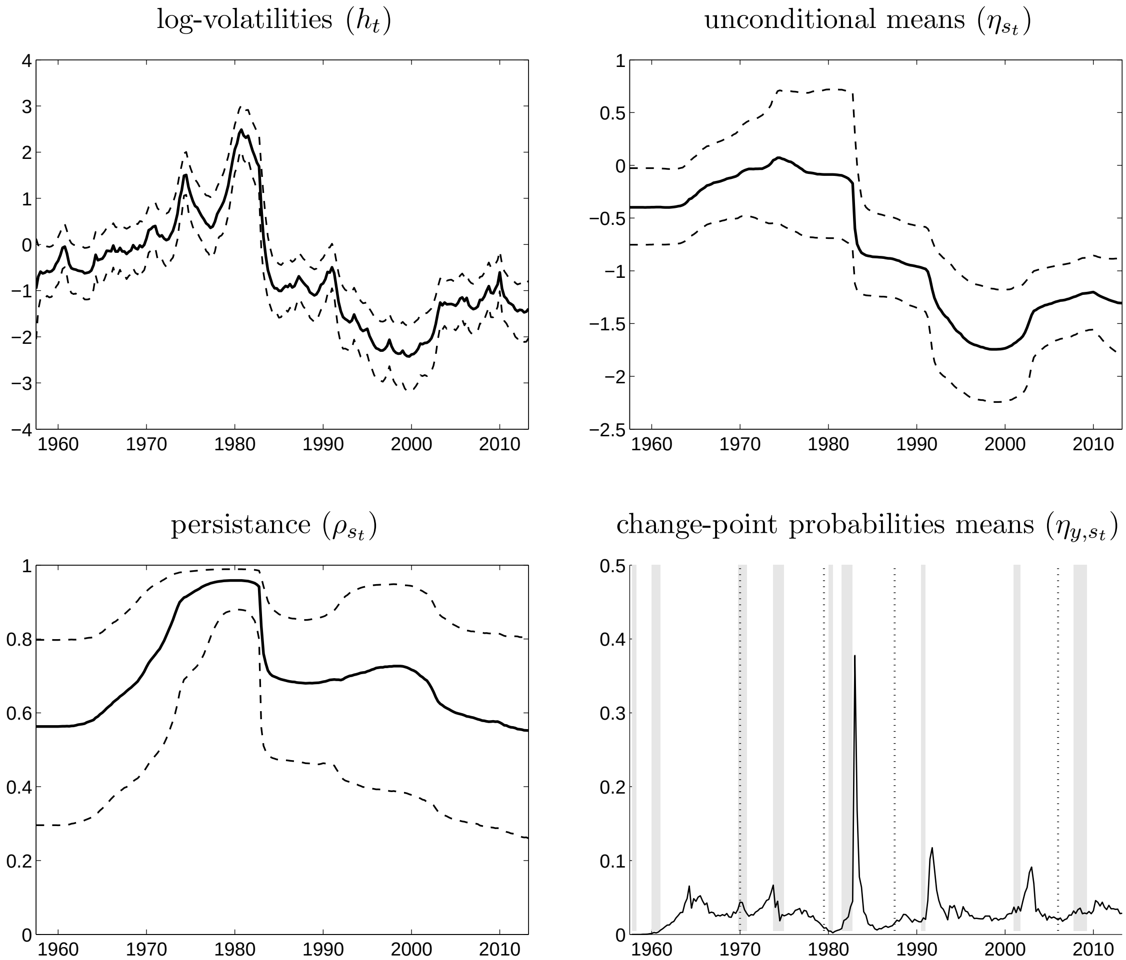

Figure 1: Core CPI data based posterior median and the $(16\%, 84\%)$ probability interval for log-volatility h_t and parameters (η_{s_t}, ρ_{s_t}) of the Multiple Stationary Regime specification. Also shown are the regime change point probabilities with Federal Reserve Bank Chairperson changeovers (vertical dotted lines) and the NBER recessions (shaded grey bands).

C Prior Sensitivity Analysis

We examine two aspects of the prior on regime durations. First, we assess the effect of changing prior hyper-parameters $\underline{\alpha}_\lambda, \underline{\xi}_1, \underline{\xi}_2$ that lead to the same expected (marginal) prior duration $\underline{d}_m = \mathrm{E}(d_m)$. In Figure 2 we present the marginal probability mass function $p(d_m)$ under various settings of the hyper-parameters. This is easily computed by numerically integrating out β_λ, λ_m in the hierarchical prior outline in the main text. The settings for $\underline{\alpha}_\lambda, \underline{\xi}_1$, and $\underline{\xi}_2$ are shown in Table 1. Each panel in Table 1 relates to densities displayed in the corresponding panel in Figure 2.

The densities plotted in the top left panel all have the same mean duration of $\underline{d}_m = 32$. Similarly, the densities plotted in the top right panel all have $\underline{d}_m = 62$ and those in the bottom left have $\underline{d}_m = 92$. The three plots in each of these panels are produced using the different settings for $\underline{\alpha}_\lambda, \underline{\xi}_1$, and $\underline{\xi}_2$ shown in Table 1. While there is some variation in the shapes of the densities, the mass of

$\underline{d}_m$	$\underline{\alpha}_\lambda$	$\underline{\xi}_1$	$\underline{\xi}_2$	$\underline{d}_m$	$\underline{\alpha}_\lambda$	$\underline{\xi}_1$	$\underline{\xi}_2$
32	30	30	30	62	60	60	60
32	30.5	60	60	62	30	15.8	30
32	30.6	90	90	62	90	133.8	90
$\underline{d}_m$	$\underline{\alpha}_\lambda$	$\underline{\xi}_1$	$\underline{\xi}_2$	$\underline{d}_m$	$\underline{\alpha}_\lambda$	$\underline{\xi}_1$	$\underline{\xi}_2$
92	90	90	90	32	30	30	30
92	30	30	88	62	60	60	60
92	60	60	89.5	92	90	90	90

Table 1: Settings of hyperparameters in Figure 2. The rows and columns correspond to those in the figure.

each is around the same general region. The bottom right panel shows three densities with mean durations $\underline{d}_m = 32$, 62 and 92 and in all three we have imposed $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2$. Under these settings we see significant changes in the densities. The biggest difference is clearly the location, but there is also some slight change in the shape. This analysis seems to reinforce the view of Koop and Potter [2007] that expected regime duration is the most important feature of their hierarchical prior, and in our empirical applications we follow them in setting $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2$ for any given value of $\mathrm{E}(d_m)$. We could analyse further but we set $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2$ in all cases and focussed upon $E(d_m)$ in setting $\underline{\alpha}_\lambda, \underline{\xi}_1$, and $\underline{\xi}_2$ as we found it easier to think about the location than the hyper-parameters.

Moreover, we re-estimate our multiple stationary regimes model under the two alternative settings $\underline{d}_m = 62$ and $\underline{d}_m = 122$. Corresponding to each of these configurations, we set $\delta_\eta = \delta_\rho \in \{0.5, 1\}$, thus allowing for greater cross-regime transitions in the parameters for models with longer *a priori* durations (also, recall that for our main specification with $\underline{d}_m = 32$, we set $\delta_\eta = 0.25$). All other prior settings are the same as those described in the main text.

Figure 3 presents two results for the AR(4) conditional mean specification and Figure 4 does the same for the bounded trend mean. For ease of comparison, we also re-produce estimates under the $\underline{d}_m = 32$ prior for each conditional mean. In both cases the MSR specification converges to a single stationary regime process as the prior expected duration is increased, but inline with our main results, the estimated log volatilities do not change significantly across these alternative priors.

Note that 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.

Also as is to be expected, the change-point 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 shorter prior duration, $\underline{d}_m = 62$, there are clear spikes in probability of regime change around 1973, and 1992 and a few other points that coincide with the more pronounced spikes under the $\underline{d}_m = 32$ specification.

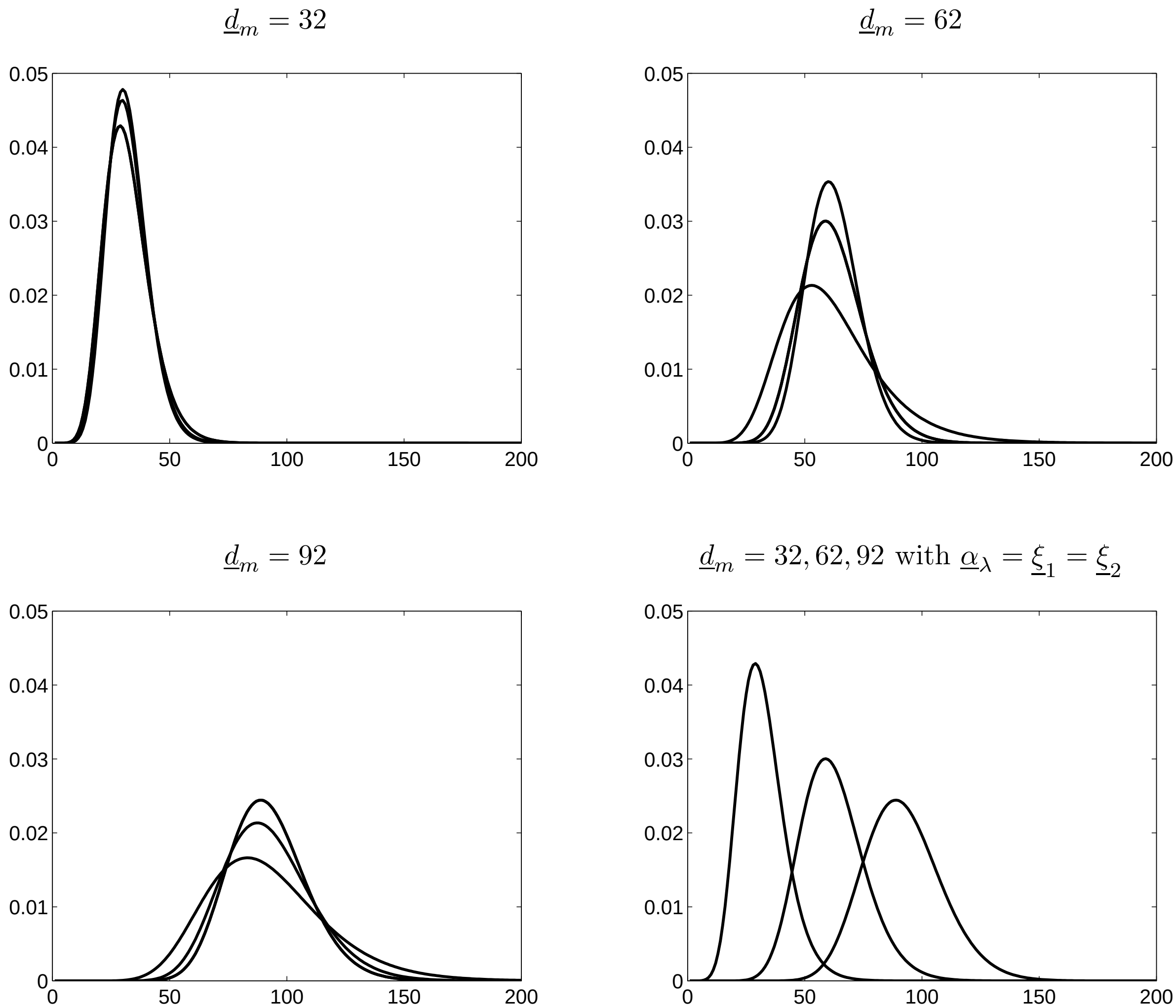

Figure 2: Marginal prior mass function for regime durations $p(d_m)$ under various settings of $\underline{\alpha}_\lambda, \underline{\xi}_1, \underline{\xi}_2$.

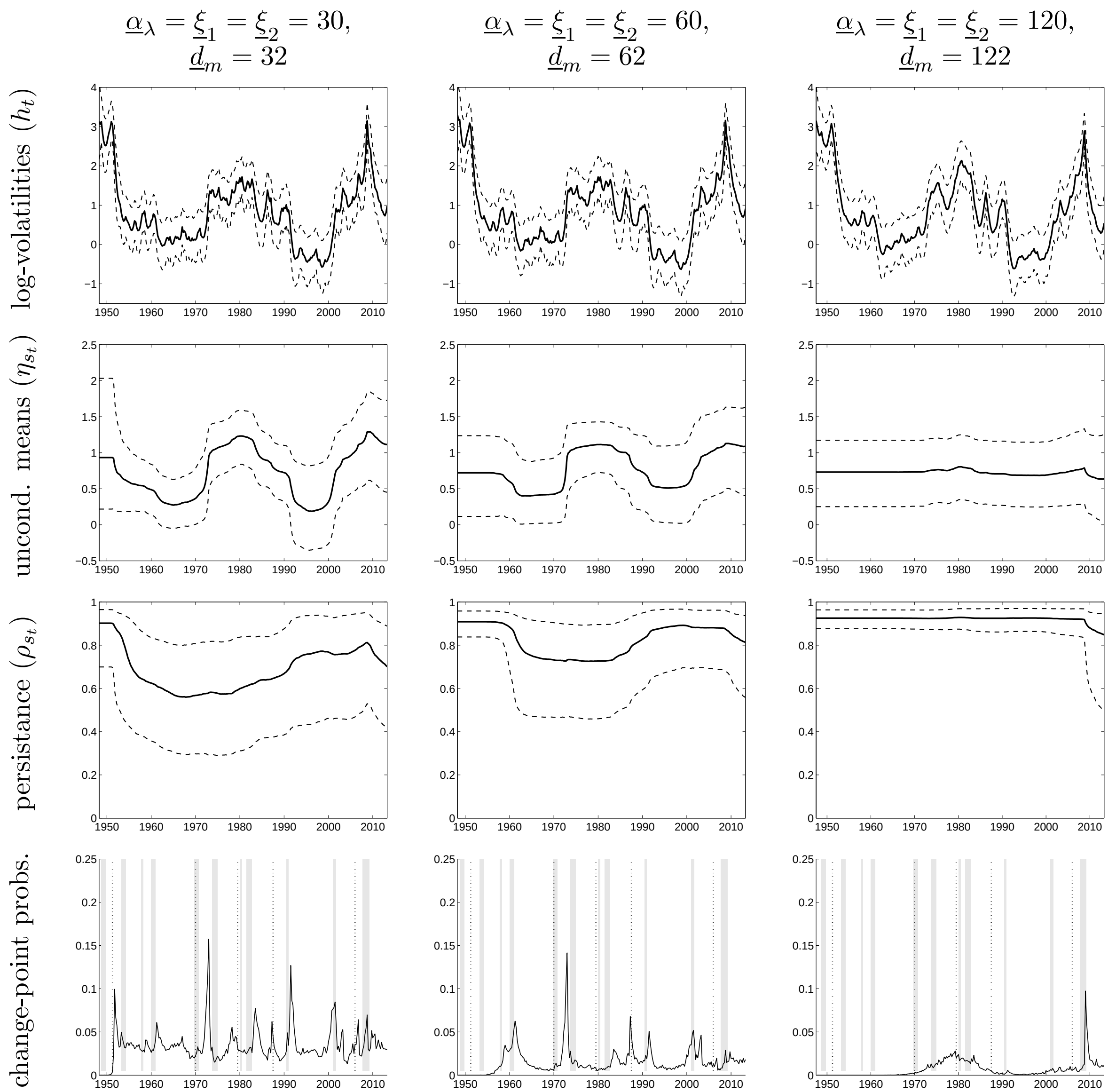

Figure 3: Prior sensitivity analysis for changing prior expected durations under the AR(4) mean model. The dashed lines depict the (16%, 84%) HPD intervals.

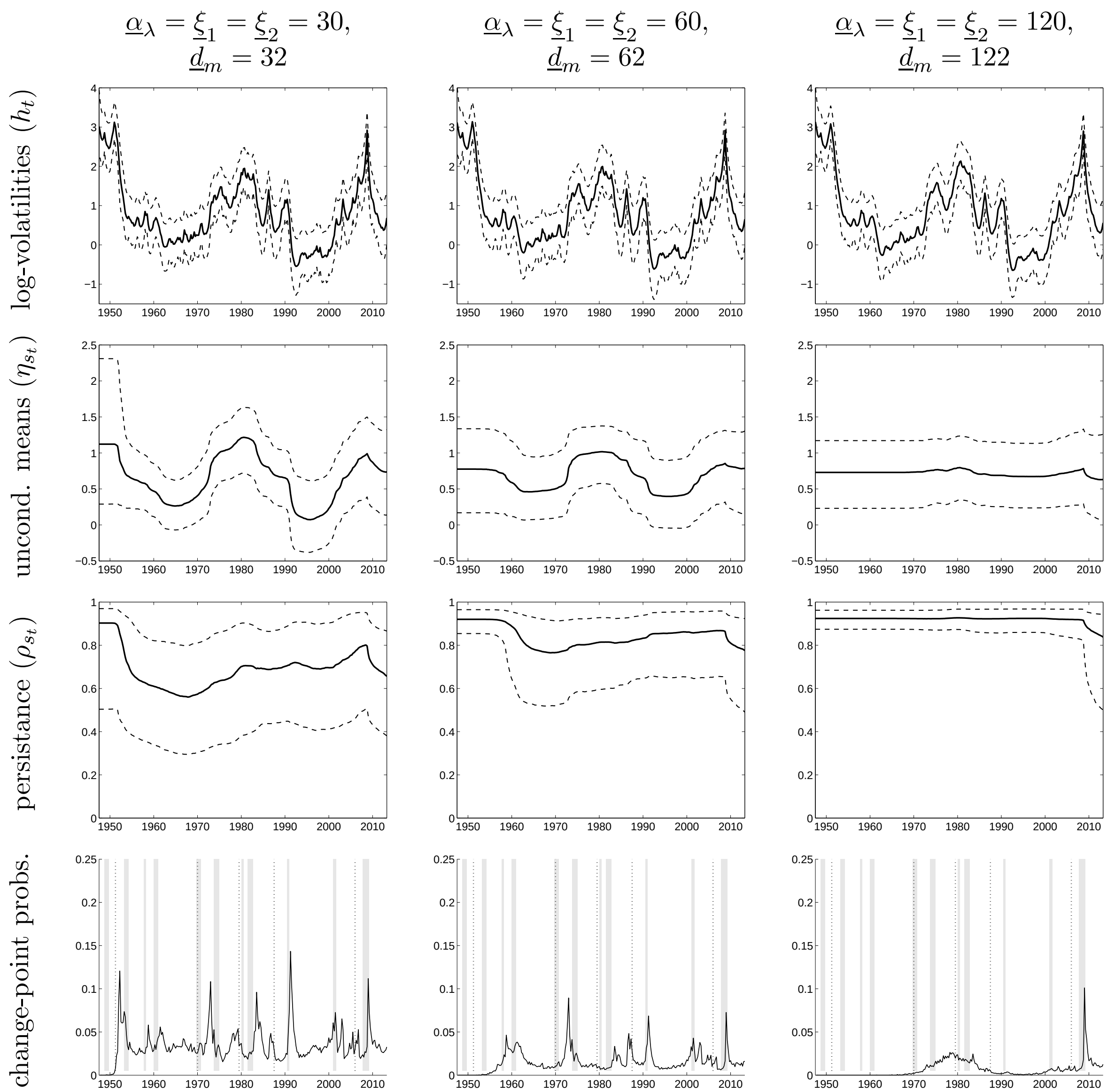

Figure 4: Prior sensitivity analysis for changing prior expected durations under the Bounded Trend conditional mean model. The dashed lines depict the (16%, 84%) HPD intervals.

D Comparision of Results Under Alternative Mean Specifications

In this section we present results for the two additional specifications of the mean equation: *stationary unobserved components* and *full time-varying parameters*. The model with stationary unobserved components is given by

$$y_t = \phi_{0,t} + \phi_1 y_{t-1} + \cdots + \phi_4 y_{t-4} + \exp(h_t/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \tag{23}$$

$$\phi_{0,t} = \pi_0 + \pi_1 \phi_{0,t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\sigma_\phi^2), \tag{24}$$

where $-1 < \pi_1 < 1$ and $\phi_1, \ldots, \phi_4$ are restricted to stationarity. The model with full time-varying parameters in the mean is given by

$$y_t = \phi_{0,t} + \phi_{1,t} y_{t-1} + \phi_{2,t} y_{t-2} + \exp(h_t/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \tag{25}$$

$$\phi_{i,t} = \phi_{i,t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\sigma_\phi^2), \tag{26}$$

and no stationarity restrictions are imposed. Figure 5 presents the results using $\underline{\alpha}_\lambda = \underline{\xi}_1 = \underline{\xi}_2 = 30$. For ease of comparison, we also reproduce estimates using the bounded trend as the conditional mean.

References

Siddhartha Chib. Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75:79–97, 1996.

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

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

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

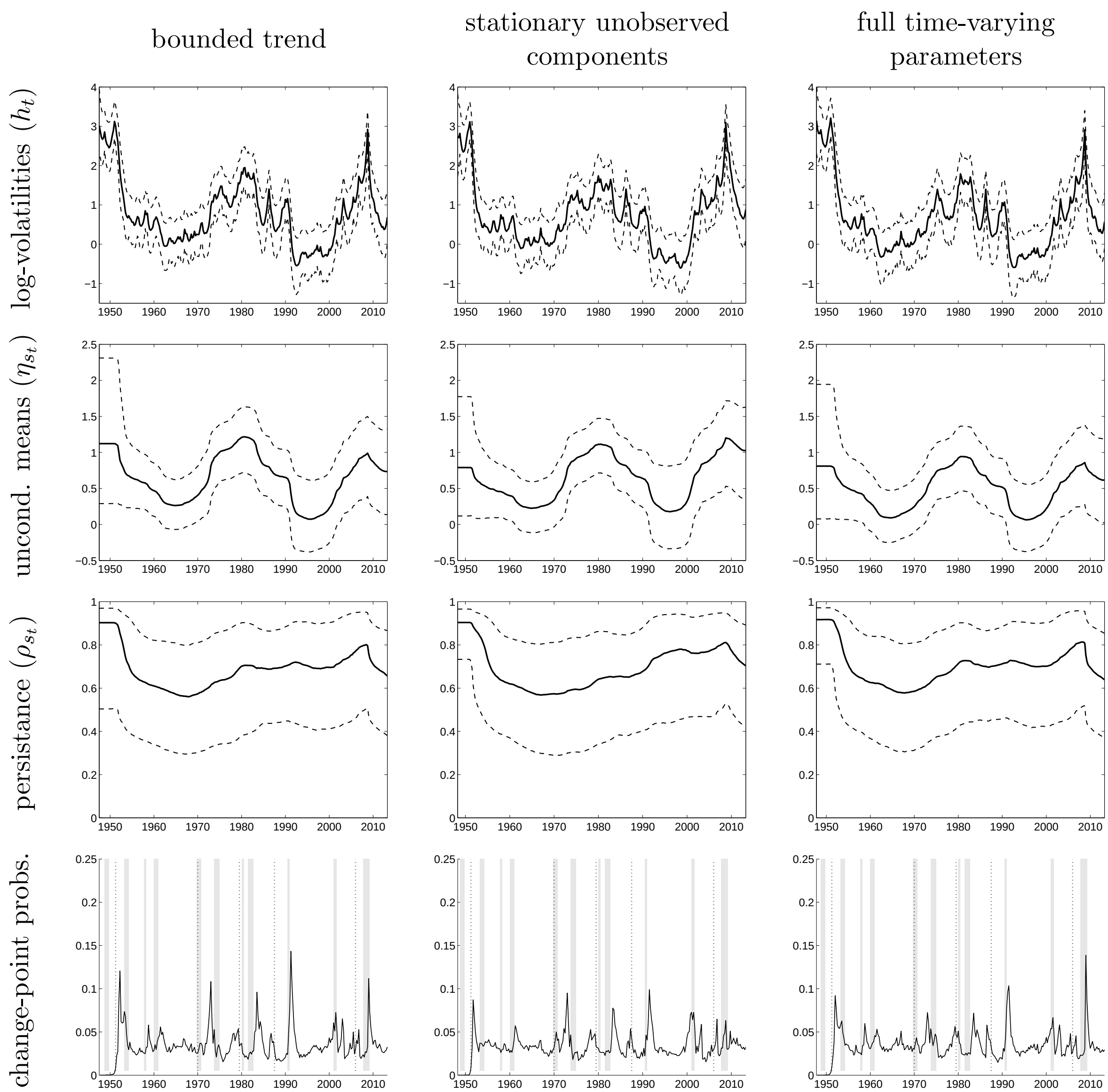

Figure 5: Comparison of alternative mean equation specification. The dashed lines depict the (16%, 84%) HPD intervals.