Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models

Paolo Giordani, Swedish Central Bank   Robert Kohn, UNSW

13/12/2006

The problem

- Optimal inference when model parameters can shift.
- Emphasis on 'tracking' the state of the system.
- Unknown: number, location and size of the shifts/breaks.
- Unknown: which parameters have changed.
- Clements and Hendry (1999): 'first-order' deterioration of forecasts.

The most common Bayesian approach

- Regime switching: Breaks as (non-reversible) regimes ('*multiple change-point*').
- Given the number of breaks:

$$
\begin{aligned}
y_t &= \beta_m^{'} z_t + \sigma_m u_t \\
\beta_m &= \beta_1,\ \sigma_m = \sigma_1,\ \text{for } \tau_1 > t \geq 1 \\
&\ldots \\
\beta_m &= \beta_M, \sigma_m = \sigma_M,\ \text{for } n \geq t \geq \tau_{M-1}.
\end{aligned}
$$

  A Markov chain determines the probability of moving to the next regime.
- Inference by Gibbs sampling standard, but often inefficient.
- Impractical for models with time-varying unobserved states (AO, seasonals, structural TS, factor models, DSGE).

Our approach: Mixture Innovation Models (MIA)

- Conditionally Gaussian State-Space form, system matrices functions of discrete latent variables K_t

$$y_t = g_t + h_t' x_t + \gamma_t u_t$$
$$x_t = f_t + F_t x_{t-1} + \Gamma_t v_t$$
$$u_t \sim nid(0,1), v_t \sim nid(0, I).$$

- Innovations in both observation and transition equations can be mixtures of normals.

- Simplest example:

$$\begin{aligned} y_t &= \mu_t + \sigma_u u_t \\ \mu_t &= \mu_{t-1} + K_t \sigma_v v_t, \end{aligned}$$

where K_t is Bernoulli,

$$\begin{aligned} K_t &= 1 \text{ prob. } \pi \\ K_t &= 0 \text{ prob } 1 - \pi. \end{aligned}$$

- Shifts in log-variance can be modeled similarly:

$$\ln \sigma_{u,t}^2 = \ln \sigma_{u,t}^2 + \sigma_v K_{2,t} v_t,$$

$$\begin{aligned} K_{2,t} &= 1 \text{ prob. } \pi \\ K_{2,t} &= 0 \text{ prob } 1 - \pi. \end{aligned}$$

Some advantages of MIA

1. Easier for parameters to change independently.
2. Easier for independent shifts in variance and in conditional mean.
3. Easier to model outliers, breaks and (some) non-linearity jointly.
4. Easier to allow for several types of breaks.
5. Can generalize Markov-switching. E.g.: Not all high inflation (volatility) regimes have the same mean inflation (volatility).

Inference

- Inference is complex: $p(y|K,\theta)$ available, but $p(y|\theta)$ only in special cases (for straightforward maximization), and $E_K(\ln p(y,K|\theta))$ only in special cases (for maximization via EM). Simulation methods needed.
- Gibbs breaks down completely for models with breaks.
- Gerlach, Carter and Kohn (2000) and this paper develop efficient MCMC.
- Efficient: (i) computations are $O(n)$ (ii) rapid convergence (iii) adaptive: we sample most where most needed.

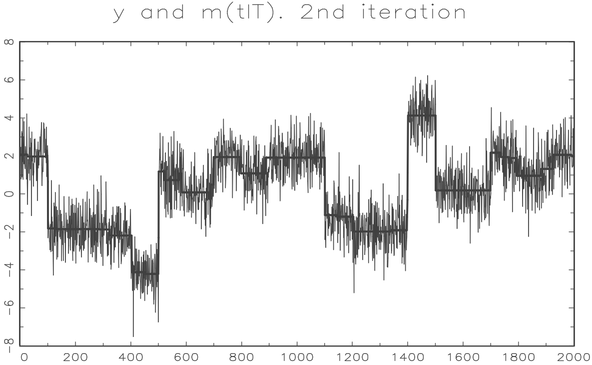

Figure: Breaks in mean

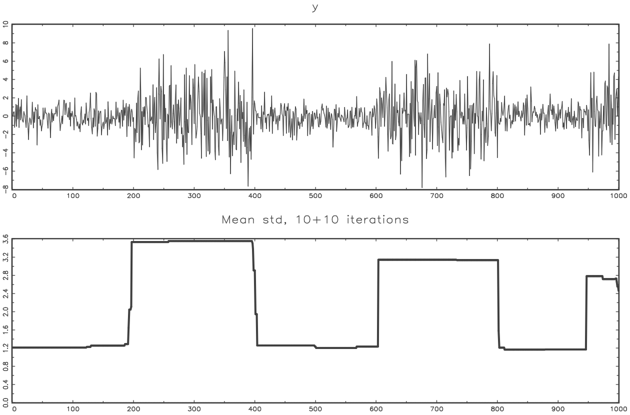

Figure: Breaks in variance

Application: US inflation

CPI inflation, 1951Q1-2004Q4 as an AR(1) with random breaks in intercept, autoregressive parameter, and variance (all can occur independently):

$$
\begin{aligned}
y_t &= c_t + b_t y_{t-1} + \sigma_t \sigma_e(K_{e,t}) e_t \\
c_t &= c_{t-1} + \sigma_c(K_{c,t}) u_t^c \\
b_t &= b_{t-1} + \sigma_b(K_{b,t}) u_t^b \\
log(\sigma_t^2) &= log(\sigma_{t-1}^2) + \sigma_v(K_{v,t}) v_t \\
p(K_t) &\equiv p(K_{m,t}, K_{v,t}) = p(K_{m,t}) p(K_{v,t}) \\
&= p(K_t | K_{s \neq t}).
\end{aligned}
$$

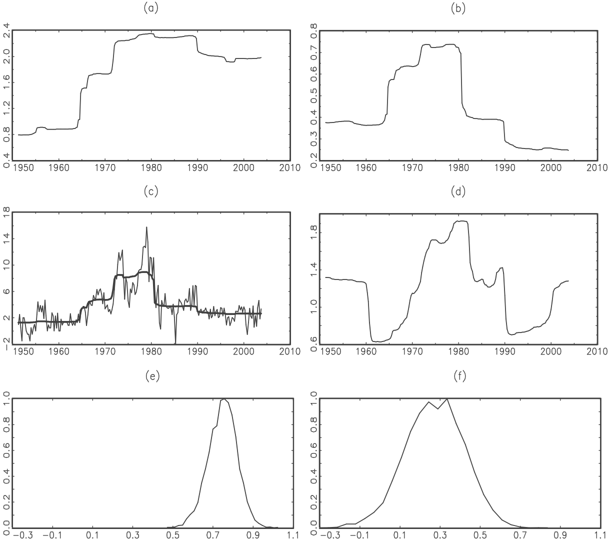

Figure: AR(1) model of U.S. inflation: (a) posterior mean of c_t (b) posterior mean of b_t (c) inflation and posterior median of $c_t/(1-b_t)$ (d) posterior mean of σ_t (e) posterior distribution of b_{1979Q1} (f) posterior distribution of b_{2004Q4}.

Work in progress or planned

- Shifts and jumps in stochastic volatility (with Dick van Dijk).
- Large simulation and forecasting exercise (with Massimiliano Marcellino).
- Large multivariate systems.

On priors

- Priors are important for inference if interventions are rare. Being uninformative on the frequency and size of breaks is not possible and should not be attempted.
- This is especially important if breaks are infrequent and close to the end of the sample.