Autoregressive Moving Average Infinite Hidden Markov-Switching Models

Luc Bauwens,[1,2] Jean-François Carpantier,[3] and Arnaud Dufays[4]

[1]Université catholique de Louvain, CORE
[2]SKEMA Business School
[3]University of Luxembourg, CREA
[4]Université Laval

Modeling and forecasting time series

Models with fixed parameters are unlikely to hold for long series.

Parameter changes:

- Abrupt or progressive?
  - Progressive changes: smooth transition, splines, TVP models...
  - Abrupt changes: Markov-switching (MS), change-point (CP).
  - Milestones:
    - ★ MS: Goldfeldt and Quandt (1973), Hamilton (1989).
    - ★ CP: Chib (1998).
- Frequent or rare?
- Global or partial?

  $\longrightarrow$ From Bauwens , Korobilis, Koop and Rombouts (2014) and some previous studies: infrequent and partial.

MS and CP: transition probability matrices

Markov-switching case (K recurrent regimes):

$$P_{MS} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \dots & p_{1K} \\ p_{21} & p_{22} & p_{23} & \dots & p_{2K} \\ \dots & \dots & \dots & \dots & \dots \\ p_{K1} & p_{K2} & p_{K3} & \dots & p_{KK} \end{pmatrix}$$

Change-point case (no recurrence):

$$\begin{pmatrix} p_{11} & 1 - p_{11} & 0 & \dots & \dots & 0 \\ 0 & p_{22} & 1 - p_{22} & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_{K-1,K-1} & 1 - p_{K-1,K-1} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Issues with MS and CP models

- MS or CP?
- How many states (regimes)?

$\longrightarrow$ These choices can be made by using the marginal log-likelihood (MLL),

BUT:

  - In case of path dependence (ARMA vs AR, GARCH vs ARCH), computations are heavy (Bauwens, Dufays and Rombouts: 2013).
  - Do all parameters change simultaneously? Usually imposed in MS and CP models, but superfluous parameters deteriorate forecasting performance.

IHMS

- Infinite hidden Markov-swtiching modeling framework:
  - Avoid choosing the number of regimes and the type of model by MLL: Assuming a possibly infinite number of regimes, thus encompassing any finite number.
  - Avoid several heavy estimations (one for each presumed number of regimes and type of model).

  Ferguson (1973) : DP (Bayesian NP analysis).

  Blackwell and MacQueen (1973): Pólya urn representation.

  Sethuraman (1994): stick-breaking representation.

  Teh et al. (2006): Hierarchical Dirichlet process (HDP).

  Fox et al. (2011): sticky HDP

- Used in several fields:
  - Visual recognition: Beal and Krishnamurthy (2006).
  - Genetics: Kivinen etl. (2007).
  - Economics: AR models: Song (2014), Jochmann (2015).
    Volatility: Jensen and Maheu (2010).

Contributions

- IHMS-ARMA model inference by MCMC.
  - Path dependence (due to MA component) issue solved by appropriate MH step.
  - Sampling of ARMA parameters by "Metropolis adapted Langevin algorithm".
- Different break dynamics for the conditional mean function parameters and for the variance.
- Steppingstone algorithm of Xie et al. (2011) for MLL computation (useful for comparing ARMA using different priors or with other models).
- Forecast comparisons on 18 macro series.

PLAN

- IHMS-ARMA MODEL
- MCMC
- EMPIRICAL STUDY

MS-ARMA WITH FINITE NUMBER OF REGIMES

The setting is similar to that of Hamilton (1989).

- For $t = 1, 2, \ldots, T$:

$$\begin{aligned} y_t &= \mu_{s_t} + \beta_{s_t} y_{t-1} + \phi_{s_t} \epsilon_{t-1} + \epsilon_t, \\ \epsilon_t &\sim N(0, \sigma^2_{s_t}). \end{aligned}$$

- $s_t \in \{1, 2, \ldots, K\}$ represents which regime is active at time period t and follows a first-order Markov chain with transition probability matrix $P_{MS} = (p_{ij})$.

- Two drawbacks:
  - The number of regimes K must be chosen before the estimation.
  - The four parameters must change simultaneously at each regime switch.

Dirichlet process mixture model/1

- Assuming breaks occur only in the variance of ϵ_t:

$$
\begin{aligned}
y_t &= \mu + \beta y_{t-1} + \phi\epsilon_{t-1} + \epsilon_t, \\
\epsilon_t &\sim N(0, \sigma_t^2), \\
\sigma_t^2 | G_0 &\sim G_0, \\
G_0 | \eta, H &\sim DP(\eta, H) \quad \text{(Dirichlet process).}
\end{aligned}
$$

  The DP is a 'non-parametric' prior on the variance of ϵ_t.

- From its Pólya urn representation (Blackwell and MacQueen, 1973), if we assume that at time t, only K different values $\tilde{\sigma}_i^2$ have been drawn, the DP prior implies :

$$
\sigma_t^2 | \sigma_1^2, \ldots, \sigma_{t-1}^2 \sim \sum_{i=1}^{K} \frac{n_i}{\eta + t - 1} \delta_{\tilde{\sigma}_i^2} + \frac{\eta}{\eta + t - 1} H.
$$

  The probability that $\sigma_t^2 = \tilde{\sigma}_i^2$ increases with the number of realizations that have already been assigned to regime i.

- This results highlights the time-varying nature of the variance.

Dirichlet process mixture model/2

- Sethuraman (1994) provides the 'stick-breaking' representation of a DP(η,H):

  "If $\{\pi_i\}_{i=1}^{\infty}$ and $\{\sigma_i^2\}_{i=1}^{\infty}$ are independent sequences of i.i.d. random variables defined as

$$\sigma_i^2 \sim H, \qquad \pi_i = \beta_i \prod_{l=1}^{i-1}(1-\beta_l) \text{ where } \beta_i \sim \text{Beta}(1,\eta),$$

  then $G_0 = \sum_{i=1}^{\infty} \pi_i \delta_{\sigma_i^2} \sim DP(\eta, H)$."

- $\{\pi_i\}_{i=1}^{\infty}$ defines a discrete distribution over the positive integers.
  NB: $\pi \sim \text{Stick}(\eta)$ will be used to denote such a distribution.

- The explicit form of G_0 highlights that the DP support is discrete.

Dirichlet process mixture model/3

- From the stick-breaking representation, it can be shown that the conditional predictive density of the DP mixture model is given by

$$f(y_t|y_{1:t-1}, \mu, \beta, \phi, \{\sigma_i^2\}_{i=1}^{\infty}, \{\pi_i\}_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \pi_i f_N(y_t|\mu + \beta y_{t-1} + \phi\epsilon_{t-1}, \sigma_i^2).$$

- This result shows that
  - the Dirichlet process helps to move to an infinite number of regimes,
  - but the transition probabilities to switch from one state to another are independent of time.
- Teh et al. (2006) were the first to restore the Markovian property in the state transitions by introducing the infinite hidden Markov-switching framework.
- Fox et al. (2011) developed the sticky-IHMS setting that copes with the high regime persistence typical in a time series context.

Complete sticky-IHMS-ARMA model

The sticky IHMS-ARMA model is defined as

$$
\begin{aligned}
y_t &= \mu_t + \beta_t y_{t-1} + \phi_t \epsilon_{t-1} + \epsilon_t, \\
\epsilon_t &\sim N(0, \sigma_t^2), \\
\psi_t \equiv \{\mu_t, \beta_t, \phi_t\} | \psi_{t-1}, G_{\psi_{t-1}} &\sim G_{\psi_{t-1}}, \\
G_{\psi_{t-1}} | G_0, \alpha_\psi, \kappa_\psi &\sim \text{DP}\left(\alpha_\psi + \kappa_\psi, \frac{\alpha_\psi G_0 + \kappa_\psi \delta_{\psi_{t-1}}}{\alpha_\psi + \kappa_\psi}\right), \\
G_0 | \eta_\psi, H_\psi &\sim \text{DP}(\eta_\psi, H_\psi), \\
\sigma_t^2 | \sigma_{t-1}^2, G_{\sigma_{t-1}^2} &\sim G_{\sigma_{t-1}^2}, \\
G_{\sigma_{t-1}^2} | G_1, \alpha_\sigma, \kappa_\sigma &\sim \text{DP}\left(\alpha_\sigma + \kappa_\sigma, \frac{\alpha_\sigma G_1 + \kappa_\sigma \delta_{\sigma_{t-1}^2}}{\alpha_\sigma + \kappa_\sigma}\right), \\
G_1 | \eta_\sigma, H_\sigma &\sim \text{DP}(\eta_\sigma, H_\sigma),
\end{aligned}
$$

where α_ψ, κ_ψ, η_ψ, α_σ, κ_σ, η_σ are positive parameters.

Stick breaking formulation of the model

s_t^{ψ} and s_t^{σ} are discrete random variables that can take any value in $\{1, 2, 3, \ldots\}$.

$$
\begin{aligned}
y_t &= \mu_{s_t^{\psi}} + \beta_{s_t^{\psi}} y_{t-1} + \phi_{s_t^{\psi}} \epsilon_{t-1} + \epsilon_t, \\
\epsilon_t &\sim N(0, \sigma^2_{s_t^{\sigma}}),
\end{aligned}
$$

$$
\begin{aligned}
s_t^{\psi} | s_{t-1}^{\psi}, \{p_i^{\psi}\}_{i=1}^{\infty} &\sim p^{\psi}_{s_{t-1}^{\psi}}, \\
p_i^{\psi} | \pi^{\psi}, \alpha_{\psi}, \kappa_{\psi} &\sim \text{DP}\left(\alpha_{\psi} + \kappa_{\psi}, \frac{\alpha_{\psi}\pi^{\psi} + \kappa_{\psi}\delta_i}{\alpha_{\psi} + \kappa_{\psi}}\right), \\
\pi^{\psi} | \eta_{\psi} &\sim \text{Stick}(\eta_{\psi}), \\
\psi_{s_t^{\psi}} \equiv \{\mu_{s_t^{\psi}}, \beta_{s_t^{\psi}}, \phi_{s_t^{\psi}}\} &\sim H_{\psi} \text{ (hierarchical truncated normal)}.
\end{aligned}
$$

NB: $E(p_i^{\psi} | \alpha_{\psi}, \pi^{\psi}, \kappa_{\psi}) = \frac{\alpha_{\psi}\pi^{\psi} + \kappa_{\psi}\delta_i}{\alpha_{\psi} + \kappa_{\psi}}$: self-transition probabilities are larger on average than probabilities of moving to another state if sticky parameter $\kappa_{\psi} > 0$.

Hierarchical prior distribution for ARMA parameters

For each regime i: $\{\mu_i, \beta_i, \phi_i\} \sim H_\psi \equiv \mathrm{N}(\bar{\mu}, \bar{\Sigma})\ \delta_{\{|\beta_i|<1, |\phi_i|<1\}}$

Hierarchical parameter: $\bar{\mu}$

$$\bar{\mu} \sim \mathrm{N}(\underline{\mu}, \underline{\Sigma})$$

$$\underline{\mu} = \{0, 0, 0\},\ \underline{\Sigma} = 0.1 I_3$$

Hierarchical parameter: $\bar{\Sigma}$

$$\bar{\Sigma}^{-1} \sim \mathrm{W}(\underline{V}, \underline{\nu})$$

$$\underline{V} = \frac{1}{5\underline{\nu}} I_3,\ \underline{\nu} = 5$$

$\Rightarrow$ Marginal prior on AR and MA coefficients: quasi uniform on (-1,+1).

Stick breaking formulation of the model

Similarly for the variance:

$$\begin{aligned} s_t^\sigma | s_{t-1}^\sigma, \{p_i^\sigma\}_{i=1}^\infty &\sim p_{s_{t-1}^\sigma}^\sigma, \\ p_i^\sigma | \pi^\sigma, \alpha_\sigma, \kappa_\sigma &\sim \mathrm{DP}\left(\alpha_\sigma + \kappa_\sigma, \frac{\alpha_\sigma \pi^\sigma + \kappa_\sigma \delta_i}{\alpha_\sigma + \kappa_\sigma}\right), \\ \pi^\sigma | \eta_\sigma &\sim \mathrm{Stick}(\eta_\sigma), \\ \sigma^2_{s_t^\sigma} &\sim H_\sigma \text{ (hierarchical inverse-gamma)}. \end{aligned}$$

Hierarchical prior distribution:

For each regime i: $\{\sigma_i^{-2}\} \sim H_\sigma \equiv \text{Gamma}(\bar{e}, \bar{f})$

Hierarchical parameter: $\bar{e}$

$\bar{e} \sim \text{Exp}(\underline{e_a})$

$\underline{e_a} = 2$

Hierarchical parameter: $\bar{f}$

$\bar{f}^{-1} \sim \text{Gamma}(\underline{f_a}, \underline{f_b})$

$\underline{f_a} = 10$, $\underline{f_b} = 1/5$

Conditional predictive densities

- Let $\Theta = \{\mu_i, \beta_i, \phi_i, \sigma_i\}_{i=1}^{\infty}$ and $F_{t-1} = \{y_{1:t-1}, \Theta, \{p_i^{\psi}\}_{i=1}^{\infty}, \{p_i^{\sigma}\}_{i=1}^{\infty}\}$.
- From the above representation of the IHMS-ARMA model, we can obtain the following predictive densities:

$$f(y_t | F_{t-1}, s_t^{\psi}, s_{1:t-1}^{\psi}, s_{1:t-1}^{\sigma}) = \sum_{j=1}^{\infty} p^{\sigma}_{s^{\sigma}_{t-1},j} f_N(y_t; \mu_{s_t} + \beta_{s_t} y_{t-1} + \phi_{s_t}\epsilon_{t-1}, \sigma_j^2),$$

  an infinite mixture of Normal distributions with time-varying probabilities.
  After integrating over the current mean parameters state value (s_t^{ψ}),

$$f(y_t | F_{t-1}, s_{1:t-1}^{\psi}, s_{1:t-1}^{\sigma}) = \sum_{i=1}^{\infty} p^{\psi}_{s^{\psi}_{t-1},i} \Big[\sum_{j=1}^{\infty} p^{\sigma}_{s^{\sigma}_{t-1},j} f_N(y_t | \mu_i + \beta_i y_{t-1} + \phi_i \epsilon_{t-1}, \sigma_j^2) \Big].$$

  showing that the model is equivalent to a MS-ARMA model with an infinite number of regimes for the mean function parameters and for the variance.
- This avoids the two drawbacks of the MS-ARMA model with finite number of states and forced simultaneous change of the mean function parameters and of the variance.

COMMENTS

- The same IHMS framework has been used in the economics literature:
  - Song (2014), Jochmann (2015) with AR models for macroeconomic series.
  - Jensen and Maheu (2010, 2013), 2014), Jin and Maheu (2014), Dufays (2012) for volatility models (GARCH, SV, RV).
- Our contributions:
  - We use two sticky IHMS processes.
  - For macro series: we add the MA component (facing the complication of path dependence).

Plan

- IHMS-ARMA MODEL
- MCMC
- EMPIRICAL STUDY

Dealing with the infinite number of regimes/1

- Could be done by applying the beam sampler of Van Gael et al. (2008). It augments the posterior with auxiliary variables that truncate the infinite number of states to a finite one, preserving the targeted posterior as invariant distribution.
- A simpler alternative: truncating the infinite sum to a large number of states L without embedding auxiliary random variables, a technique known as the *degree L weak limit approximation* (Ishwaran and Zarepour, 2002).

  Despite the truncation, if the chosen number L is large enough, the error is negligible, see Kurihara et al. (2007) and Fox et al. (2011).
- In the empirical study: we chose $L = 10$ (results with $L = 20$ are similar).

DEALING WITH THE INFINITE NUMBER OF REGIMES/2

Under the *degree L weak limit approximation*:

- Every row of each transition matrix is truncated to a finite Dirichlet distribution (denoted by 'Dir'), instead of being driven by a Dirichlet process.
- For the mean function parameters:
  $p_i^{\psi}|\pi^{\psi}, \alpha_{\psi}, \kappa_{\psi} \sim \text{DP}\left(\alpha_{\psi} + \kappa_{\psi}, \frac{\alpha_{\psi}\pi^{\psi} + \kappa_{\psi}\delta_i}{\alpha_{\psi} + \kappa_{\psi}}\right)$, is replaced by

  $p_i^{\psi} = \{p_{i1}^{\psi}, p_{i2}^{\psi}, ..., p_{iL}^{\psi}\}|\pi^{\psi}, \alpha_{\psi}, \kappa_{\psi} \sim$
  $\text{Dir}\left(\alpha_{\psi}\pi_1^{\psi}, \alpha_{\psi}\pi_2^{\psi}, ..., \alpha_{\psi}\pi_i^{\psi} + \kappa_{\psi}, ..., \alpha_{\psi}\pi_L^{\psi}\right)$, and

  $\pi^{\psi} \sim \text{Stick}(\eta_{\psi})$ by

  $\pi^{\psi} = \{\pi_1^{\psi}, \pi_2^{\psi}, ..., \pi_L^{\psi}\}|\eta_{\psi} \sim Dir\left(\frac{\eta_{\psi}}{L}, \frac{\eta_{\psi}}{L}, ..., \frac{\eta_{\psi}}{L}\right)$.
- And similarly for the variance...

Gibbs sampler blocks

1. $f(s^{\psi}_{1:T}|\Theta, P^{\psi}, s^{\sigma}_{1:T}, y_{1:T}) \quad \rightarrow$ MH-step (path dependence)
2. $f(P^{\psi}|\Theta, H_{Dir}, \pi^{\psi}, s^{\psi}_{1:T}, y_{1:T})$
3. $f(s^{\sigma}_{1:T}|\Theta, P^{\sigma}, s^{\psi}_{1:T}, y_{1:T})$
4. $f(P^{\sigma}|\Theta, H_{Dir}, \pi^{\sigma}, s^{\sigma}_{1:T}, y_{1:T})$
5. $f(\Theta|\bar{\mu}, \bar{\Sigma}, \bar{e}, \bar{f}, H_{Dir}, s^{\psi}_{1:T}, s^{\sigma}_{1:T}, y_{1:T}) \rightarrow$ MALA
6. $f(\bar{\mu}, \bar{\Sigma}, \bar{e}, \bar{f}|\Theta, H_{Dir}, s^{\psi}_{1:T}, s^{\sigma}_{1:T}, y_{1:T})$
7. $f(\pi^{\psi}, \pi^{\sigma}|\Theta, P^{\psi}, P^{\sigma}, H_{Dir}, s^{\psi}_{1:T}, s^{\sigma}_{1:T}, y_{1:T})$
8. $f(H_{Dir}|P^{\psi}, P^{\sigma}, \pi^{\psi}, \pi^{\sigma}, s^{\psi}_{1:T}, s^{\sigma}_{1:T}, y_{1:T})$

where $H_{Dir} = (\alpha_{\psi}, \alpha_{\sigma}, \kappa_{\psi}, \kappa_{\sigma}, \eta_{\psi}, \eta_{\sigma})$, and P^{ψ} and P^{σ} denote the truncated transition matrices ($L \times L$).

Except for steps 1 and 5, the conditional distributions are directly simulated (given that priors are partially conjugate).

Block 1: $f(s^{\psi}_{1:T}|\Theta, P^{\psi}, s^{\sigma}_{1:T}, y_{1:T})$

- Backward-forward algorithm (Rabiner, 1986) applicable if NO path dependence.
- We use an approximate MS-ARMA model not subjected to path dependence, and apply the BF algorithm to build a proposal used in a MH step.
- The approximate model replaces the unobserved lagged error by its conditional expectation (similar to approximate MS-GARCH model of Klaassen (2002))

$$\begin{aligned}\tilde{\epsilon}_{t-1,s_t^{\psi}} &:= E_{s^{\psi}_{t-1}}[\epsilon_{t-1}|y_{1:t-1}, s_t^{\psi}, \Theta, P^{\psi}, s^{\sigma}_{1:t-1}] \\ &= \sum_{i=1}^{L} \epsilon_{t-1}(i) f(s_{t-1} = i|y_{1:t-1}, s_t^{\psi}, \Theta, P^{\psi}, s^{\sigma}_{1:t-1}),\end{aligned}$$

where $\epsilon_{t-1}(s^{\psi}_{t-1}) = y_{t-1} - \mu_{s^{\psi}_{t-1}} - \beta_{s^{\psi}_{t-1}} y_{t-2} - \phi_{s^{\psi}_{t-1}} \tilde{\epsilon}_{t-2,s^{\psi}_{t-1}}$.

- A proposed $s^{\psi}_{1:T}$ is very likely to be rejected if it is drawn as one block from the approximate model. To ensure good MCMC mixing properties, we sample the state vector in small blocks of random sizes.

Block 5: Sampling the ARMA parameters

- Given the state vectors and $\{\mu_i, \beta_i, \phi_i\}_{i=1}^{L}$, the full conditional of distribution of $\{\sigma_i^2\}_{i=1}^{L}$ is a product of independent inverse-gamma distributions.
- For each block $\psi_i = \{\mu_i, \beta_i, \phi_i\}$ given all the rest, we use a normal proposal for a MH step (we adapt the Metropolis adapted Langevin algorithm):

$$\tilde{\psi}|\psi_i \quad \sim \quad N\Big(\psi_i + \frac{\gamma^2}{2} G^{-1}(\psi_i) \nabla \log f(\psi_i|D), \gamma^2 G^{-1}(\psi_i)\Big), \tag{1}$$

  where ψ_i is the previous accepted draw, $\log f(\psi_i|D)$ is the log of the full conditional, $-G(\psi_i)$ the Hessian , and γ is a tuning constant.
- Details in the paper.

Plan

- IHMS-ARMA MODEL
- MCMC
- EMPIRICAL STUDY

For the Mean:	$\eta_\psi \sim G(1,10)$	$\alpha_\psi + \kappa_\psi \sim G(1,10)$ $\rho_\psi = \frac{\kappa_\psi}{\alpha_\psi+\kappa_\psi} \sim Beta(\omega_{\text{MS/CP}}, 1)$
For the Variance:	$\eta_\sigma \sim G(1,10)$	$\alpha_\sigma + \kappa_\sigma \sim G(1,10)$ $\rho_\sigma = \frac{\kappa_\sigma}{\alpha_\sigma+\kappa_\sigma} \sim Beta(\omega_{\text{MS/CP}}, 1)$

Markov-switching type:	$\omega_{\text{MS/CP}} = 10$:	weak state persistence: $E(\rho_\psi) = 0.909$
Change-point type:	$\omega_{\text{MS/CP}} = 1000$:	strong state persistence: $E(\rho_\psi) = 0.999$

GDP growth rate 1947Q2-2014Q1 (T=268)

- IHMS-ARMA strongly dominate fixed parameter ARMA (with increases by at least 23 points of the MLL).
- IHMS models with prior implying weak regime persistence (MS-type) a posteriori more probable than IHMS models implying high regime persistence (CP-type), both for AR (0.82/0.18) and ARMA (0.77/0.23).
- IHMS-ARMA more probable than IHMS-AR: posterior probabilities of 0.87/0.13 (MS-prior) and 0.84/0.16 (CP-prior).
- Posterior probabilities of number of regimes:

IHMS-ARMA with MS-type prior

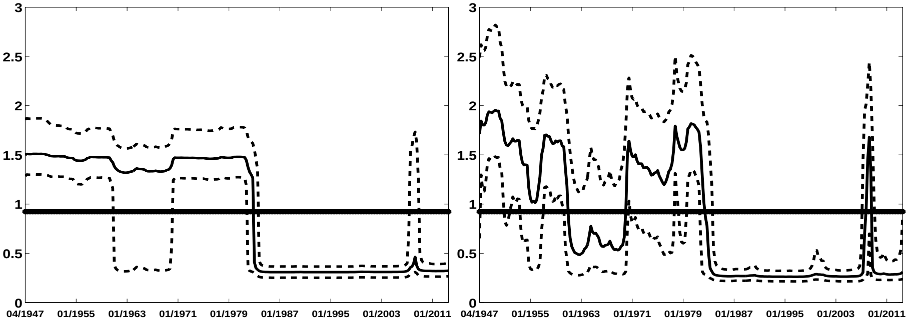
# Regimes	1	2	3	4	5	6	7	8	9
μ, β, ϕ	0.62	0.20	0.06	0.07	0.03	0.01	0	0	0
σ^2	0	0.06	0.19	0.26	0.23	0.15	0.07	0.03	0.01

IHMS-ARMA with CP-type prior

# Regimes	1	2	3	4	5	6	7	8	9
μ, β, ϕ	0.99	0.01	0.00	0	0	0	0	0	0
σ^2	0	0.80	0.16	0.03	0.01	0	0	0	0

GDP growth rate 1947Q2-2014Q1 (T=268)

Variance:



Left: CP-type prior. Right: MS-type prior.

Thick horizontal line: posterior median of the ARMA model with fixed parameters.

Thin continuous and dotted lines: IHMS-ARMA posterior median and the limits of the 70% posterior credible interval.

Forecasts from 1987Q1 to 2014Q1

Best forecasting model is IHMS-ARMA (CP prior).

CP and MS refer to the prior type of the IHMS hyper-parameters.

Forecast Horizons	1 quarter	2 quarters	1 year	2 years	3 years	4 years
			Average of Predictive Densities			
ARMA	0.33	0.30	0.28	0.27	0.27	0.27
IHMS-AR (CP)	0.46**	0.43**	0.39**	0.38**	0.37*	0.38*
IHMS-ARMA (CP)	0.47**	0.44**	0.40**	0.39**	0.38*	0.38*
IHMS-AR (MS)	0.43**	0.39**	0.36**	0.34**	0.33*	0.33*
IHMS-ARMA (MS)	0.44**	0.41**	0.37**	0.34**	0.33**	0.33*
			Continuously Ranked Probability Score			
ARMA	0.37	0.42	0.47	0.50	0.51	0.51
IHMS-AR (CP)	0.34**	0.38**	0.42**	**0.44****	**0.44****	**0.44****
IHMS-ARMA (CP)	**0.33****	**0.36****	**0.41****	**0.44****	**0.45****	0.45*
IHMS-AR (MS)	0.35**	0.39**	0.43**	0.46**	0.47**	0.47*
IHMS-ARMA (MS)	0.34**	0.37**	0.42**	0.46**	0.47**	0.47
			Mean Squared Forecast Error			
ARMA	0.40	0.53	0.68	0.78	0.83	0.82
IHMS-AR (CP)	0.39	0.49**	0.60**	0.67*	0.70	0.68
IHMS-ARMA (CP)	0.37**	0.45**	0.56**	0.65*	0.69	0.68
IHMS-AR (MS)	0.40	0.52	0.64*	0.70*	0.73	0.73
IHMS-ARMA (MS)	0.38**	0.47**	0.60**	0.69*	0.73	0.73

CRPS: (average) distance between predictive CDF and empirical cdf

Inflation rate Feb. 1959-Nov. 2012 (T=646)

- IHMS-ARMA strongly dominate fixed parameter ARMA (with increases by at least 62 points of the MLL).
- IHMS models with prior implying weak regime persistence (MS-type) a posteriori more probable than IHMS models implying high regime persistence (CP-type), both for AR (0.95/0.05) and ARMA (0.90/0.10).
- IHMS-ARMA more probable than IHMS-AR: posterior probabilities of 0.85/0.15 (MS-prior) and 0.73/0.27 (CP-prior).
- Posterior probabilities of number of regimes:

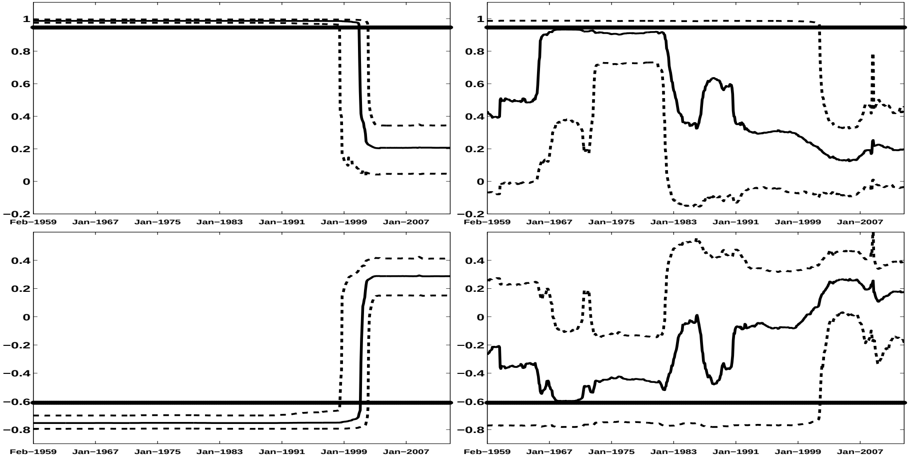
IHMS-ARMA with MS prior

# Regimes	1	2	3	4	5	6	7	8	9
μ, β, ϕ	0	0.08	**0.28**	0.27	0.20	0.13	0.03	0.01	0.00
σ^2	0	0.11	0.22	**0.26**	0.20	0.13	0.06	0.02	0.00

IHMS-ARMA with CP prior

# Regimes	1	2	3	4	5	6	7	8	9
μ, β, ϕ	0	**0.70**	0.29	0.01	0	0	0	0	0
σ^2	0	**0.87**	0.12	0.01	0.00	0.00	0	0	0

Top: AR(1) parameter. Bottom: MA(1).



Left: CP-type prior. Right: MS-type prior.

Forecasts from April 1991 to Nov. 2012

Best forecasting model depends on loss function.

CP and MS refer to the prior type of the IHMS hyper-parameters.

Forecast Horizons	1 month	2 months	4 months	8 months	12 months	16 months
			Average of Predictive Densities			
ARMA	0.15	0.14	0.14	0.13	0.12	0.12
IHMS-AR (CP)	0.16**	0.15**	0.15**	0.14**	0.14**	0.14**
IHMS-ARMA (CP)	0.16**	0.15**	0.15**	0.14**	0.13**	0.13**
IHMS-AR (MS)	0.16**	0.15**	0.15**	0.14**	0.14**	0.14**
IHMS-ARMA (MS)	0.16**	0.16**	0.15**	0.14**	0.13**	0.13**
			Continuously Ranked Probability Score			
ARMA	1.19	1.28	1.24	1.34	1.37	1.33
IHMS-AR (CP)	1.13**	1.22*	1.19	1.25*	1.29**	**1.27**
IHMS-ARMA (CP)	1.15*	1.24**	1.20**	1.28**	1.33	1.30
IHMS-AR (MS)	**1.12****	**1.21***	**1.18***	**1.24***	**1.28****	1.28
IHMS-ARMA (MS)	1.14**	1.22**	**1.18****	1.27**	1.31**	**1.27***
			Mean Squared Forecast Error			
ARMA	5.52	6.60	6.52	6.93	7.24	6.53
IHMS-AR (CP)	5.00*	6.09*	5.84*	6.15	6.55*	6.26
IHMS-ARMA (CP)	5.19	6.33	6.08*	6.56	7.13	6.62
IHMS-AR (MS)	4.94**	6.04*	5.85*	6.17	6.51*	6.33
IHMS-ARMA (MS)	5.26	6.32**	5.98*	6.35*	6.78**	6.12*

CRPS: (average) distance between predictive CDF and empirical cdf

18 QUARTERLY SERIES: 1959Q1-1990Q3-2011Q3

Summary of forecast results for 18 series:

- IHMS-ARMA yields better forecasts for a majority of series. Conclusion more robust for APD criterion than RMSFE.
- For some of these: improvement is strong.
- When fixed parameter ARMA is best, the improvements over IHMS are small.

18 QUARTERLY SERIES: 1959Q1-1990Q3-2011Q3

1. Real Gross Domestic Product
2. Personal Income
3. Real Personal Consumption Expenditures
4. Personal Consumption Expenditures: Chain-type Price Index
5. Real Gross Private Domestic Investment
6. Business Sector: Output Per Hour of All Persons
7. Real Imports of Goods and Services
8. Real Exports of Goods and Services
9. Real Change in Private Inventories
10. Real Government Consumption Expenditures and Gross Inv.
11. Compensation of Employees: Wages and Salary Accruals
12. Net Corporate Dividends
13. Personal Saving
14. Real Disposable Personal Income
15. Gross Domestic Product: Implicit Price Deflator
16. Nonfarm Business Sector: Unit Labor Cost
17. Private Residential Fixed Investment
18. Gross Saving

Predictive performance of IHMS-ARMA with respect to ARMA

Criterion: Average of Predictive Densities

Forecast Horizons	One quarter	Two quarters	One year	Two years
1	34.65 (CP)	33.01 (CP)	32.65 (CP)	33.57 (CP)
2	2.40 (MS)	-0.06 (ARMA)	1.07 (MS)	1.16 (MS)
3	11.11 (CP)	10.62 (CP)	8.67 (CP)	6.45 (CP)
4	0.62 (MS)	1.94 (MS)	4.04 (MS)	3.80 (MS)
5	18.65 (CP)	17.87 (CP)	17.83 (CP)	17.16 (CP)
6	8.72 (CP)	9.25 (CP)	9.34 (CP)	8.26 (CP)
7	57.72 (CP)	47.59 (CP)	41.33 (CP)	43.18 (CP)
8	62.21 (CP)	57.40 (CP)	57.18 (CP)	58.57 (CP)
9	-3.46 (ARMA)	-3.44 (ARMA)	-2.15 (ARMA)	-1.13 (ARMA)
10	7.73 (MS)	6.53 (MS)	5.07 (CP)	5.05 (CP)
11	3.09 (MS)	2.39 (MS)	2.23 (MS)	-0.05 (ARMA)
12	18.58 (MS)	0.19 (MS)	-6.74 (ARMA)	-6.55 (ARMA)
13	-0.18 (ARMA)	-0.35 (ARMA)	-1.18 (ARMA)	0.01 (MS)
14	11.40 (MS)	5.96 (MS)	4.85 (MS)	5.26 (MS)
15	1.04 (MS)	1.61 (MS)	2.09 (MS)	2.07 (MS)
16	11.21 (CP)	12.07 (CP)	10.38 (CP)	6.45 (CP)
17	46.82 (CP)	41.36 (CP)	37.43 (CP)	34.28 (CP)
18	2.50 (MS)	3.03 (MS)	1.45 (MS)	0.46 (MS)
Perc. ARMA	11 %	17 %	17 %	17 %

MS (CP) means IHMS-ARMA with MS (CP) prior.

18 QUARTERLY SERIES: 1959Q1-1990Q3-2011Q3

Predictive performance of IHMS-ARMA with respect to ARMA

Criterion: RMSFE

Forecast Horizons	One quarter	Two quarters	One year	Two years
1	-3.36 (MS)	-3.62 (MS)	-0.82 (MS)	-1.67 (CP)
2	5.15 (ARMA)	-0.92 (MS)	0.50 (ARMA)	-0.60 (MS)
3	-2.82 (CP)	-4.24 (CP)	-0.36 (CP)	0.94 (ARMA)
4	-4.60 (MS)	-5.49 (MS)	-6.78 (MS)	-4.27 (MS)
5	-9.22 (MS)	-1.27 (MS)	1.31 (ARMA)	1.20 (ARMA)
6	-1.13 (MS)	0.74 (ARMA)	0.66 (ARMA)	0.13 (ARMA)
7	-29.56 (CP)	-3.75 (MS)	3.06 (ARMA)	0.52 (ARMA)
8	-35.41 (CP)	6.01 (ARMA)	0.84 (ARMA)	-0.04 (MS)
9	3.40 (ARMA)	7.30 (ARMA)	-0.99 (MS)	-2.38 (MS)
10	-3.33 (MS)	-3.29 (MS)	-3.03 (MS)	-1.43 (MS)
11	-1.02 (MS)	-2.65 (MS)	-3.80 (MS)	-2.30 (MS)
12	34.71 (ARMA)	6.87 (ARMA)	4.32 (ARMA)	-0.74 (CP)
13	-8.82 (MS)	-13.36 (MS)	-7.46 (CP)	-3.48 (CP)
14	0.56 (ARMA)	0.65 (ARMA)	0.42 (ARMA)	-1.56 (CP)
15	-0.37 (MS)	-1.86 (MS)	-3.47 (MS)	-3.85 (MS)
16	-2.77 (MS)	-2.64 (CP)	-6.03 (CP)	-6.73 (CP)
17	-13.65 (CP)	-16.67 (CP)	-9.62 (CP)	0.25 (ARMA)
18	0.99 (ARMA)	-1.60 (MS)	0.69 (ARMA)	1.53 (ARMA)
Perc. ARMA	28 %	28 %	44 %	33 %

MS (CP) means IHMS-ARMA with MS (CP) prior.