Appendix to "Bayesian Mixtures of Autoregressive Models" published in the Journal of Computational and Graphical Statistics

Sally Wood, Ori Rosen and Robert Kohn

Appendix A: Details of the sampling scheme

Stage I:

1. Initialize all the parameters.

2. Draw the lag p from the multinomial distribution

$$p(p|\boldsymbol{y}^*, r, \boldsymbol{z}) \propto c^{-r(p+1)/2}|X_p'X_p|^{r/2}|\prod_{j=1}^{r}|X_p'Z_jX_p + c^{-1}X_p'X_p|^{-1/2}b_j^{-a_j},$$

where $Z_j = \mathrm{diag}(z_{tjpr}, t = P+1, \ldots, n)$. Note that z_{tjpr} takes on the same value (zero or one) for all $t \in \{1 + (s-1)L, \ldots, sL\}$. The expressions for a_j and b_j are $a_j = \frac{1}{2}\sum_{t=P+1}^{n} z_{jtpr} + \alpha$ and $b_j = \frac{1}{2}\boldsymbol{y}^* M_j \boldsymbol{y}^* + \beta$, where $M_j = Z_j - Z_jX_p(X_p'Z_jX_p + c^{-1}X_p'X_p)^{-1}X_p'Z_j$.

3. Let $p_{sjpr} = \prod_{t=1+(s-1)L}^{sL} p(y_t|\boldsymbol{x}_{t-1}; \boldsymbol{\phi}_{jpr}, \sigma_{jpr}^2)$. Draw the indicators from

$$p(z_{sjpr} = 1|y_t, \boldsymbol{x}_{t-1}; \boldsymbol{\phi}_{pr}, \boldsymbol{\delta}_{pr}, \boldsymbol{\sigma}_{pr}^2) = \frac{\pi_{sjpr}\, p_{sjpr}}{\sum_{h=1}^{r} \pi_{shpr}\, p_{sjpr}},$$

$j = 1, \ldots, r$, $t \in \{1 + (s-1)L, \ldots, sL\}$, $s = 1, \ldots, S$, where $\boldsymbol{x}_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$.

4. Draw $\boldsymbol{\delta}_{pr}$ from

$$p(\boldsymbol{\delta}_{pr}|\boldsymbol{z}_{pr}) = \prod_{s=1}^{S}\prod_{j=1}^{r} \pi_{sjpr}^{z_{sjpr}} \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{j=1}^{r}\boldsymbol{\delta}_{jpr}'\boldsymbol{\delta}_{jpr}\right)$$

via a Metropolis-Hastings step.

5. For $j = 1, \ldots, r$, draw $\boldsymbol{\phi}_{jpr}$ from the multivariate normal distribution with mean vector $(X_p' Z_j X_p + c^{-1} X_p' X_p)^{-1} X_p' Z_j \boldsymbol{y}^*$ and variance-covariance vector $\sigma_{jpr}^2 (X_p' Z_j X_p + c^{-1} X_p' X_p)^{-1}$.

6. For $j = 1, \ldots, r$, draw σ_{jpr}^2 from $IG(\frac{1}{2} \sum_{s=1}^{S} \sum_{t=1+(s-1)L}^{sL} z_{tjpr} + \alpha, \frac{1}{2} \boldsymbol{y}^{*'} M_j \boldsymbol{y}^* + \beta)$. For identifiability, we constrain the σ_{jpr}^2's to be ordered.

Stage II:

As described in Section 3.2, r and p are first drawn from discrete uniform distributions. The parameter vectors $\boldsymbol{\phi}_{pr}$, $\boldsymbol{\delta}_{pr}$ and $\boldsymbol{\nu}_{pr}$ are then drawn via a Metropolis-Hastings step. The target distribution is

$$p(r, p, \boldsymbol{\phi}_{pr}, \boldsymbol{\delta}_{pr}, \boldsymbol{\nu}_{pr} | \boldsymbol{y}) \propto \prod_{t=1}^{n} \sum_{j=1}^{r} \pi_{tjpr} \, p(y_t | \boldsymbol{x}_{t-1}; \boldsymbol{\phi}_{jpr}, \sigma_{jpr}^2) p(r) p(p) p(\boldsymbol{\phi}_{pr}) p(\boldsymbol{\delta}_{pr}) p(\boldsymbol{\nu}_{pr}) \, ,$$

where $p(r)$, $p(p)$, $p(\boldsymbol{\phi}_{pr})$, $p(\boldsymbol{\delta}_{pr})$ and $p(\boldsymbol{\nu}_{pr})$ are the prior distributions. As described in Section 2, the priors on the σ_{jpr}^2s are independent $IG(a, b)$, and accordingly, the priors on the ν_{jpr} are independent with probability density function $p(\nu_{jpr}) = \frac{b^a}{\Gamma(a)} \exp(-a\nu_{jpr} - be^{-\nu_{jpr}})$, obtained from the transformation $\nu = \log \sigma^2$. To describe the reversible jump step, let $q(r^{(n)}, r^{(p)}, \boldsymbol{\eta}^{(n)} | r^{(c)}, p^{(c)}, \boldsymbol{\eta}^{(c)})$, where $\boldsymbol{\eta} = (\boldsymbol{\phi}', \boldsymbol{\delta}', \boldsymbol{\nu}')'$, be a transition probability function that moves the chain from $(r^{(c)}, p^{(c)} \boldsymbol{\eta}^{(c)})$ to $(r^{(n)}, p^{(n)}, \boldsymbol{\eta}^{(n)})$. The superscript (c) denotes the current value in the chain. The proposal density is given by $q(r^{(n)}, p^{(n)} | r^{(c)}, p^{(c)}) q(\boldsymbol{\eta}^{(n)} | r^{(n)}, p^{(n)}, \boldsymbol{\eta}^{(c)})$, i.e., new values $r^{(n)}$ and $p^{(n)}$ of r and p, respectively, are proposed, and conditional on $r^{(n)}$ and $p^{(n)}$, $\boldsymbol{\eta}^{(n)}$ is proposed. As the density $q(\boldsymbol{\eta}^{(n)} | r^{(n)}, p^{(n)}, \boldsymbol{\eta}^{(c)})$, we use $f_{\phi}(\boldsymbol{\phi}; \hat{\boldsymbol{\phi}}_{p^{(n)} r^{(n)}}, \hat{\Sigma}_{\boldsymbol{\phi}_{p^{(n)} r^{(n)}}}) \times f_{\delta}(\boldsymbol{\delta}; \hat{\boldsymbol{\delta}}_{p^{(n)} r^{(n)}}, \hat{\Sigma}_{\boldsymbol{\delta}_{p^{(n)} r^{(n)}}}) \times f_{\nu}(\boldsymbol{\nu}; \hat{\boldsymbol{\nu}}_{p^{(n)} r^{(n)}}, \hat{\Sigma}_{\boldsymbol{\nu}_{p^{(n)} r^{(n)}}})$, where f_{ϕ}, f_{δ} and f_{ν} are multivariate normal densities, and the means and covariance matrices are as described in Section 3.1. The acceptance probability is

$$\min \left\{ 1, \frac{p(r^{(n)}, p^{(n)}, \boldsymbol{\eta}^{(n)} | \boldsymbol{y})}{p(r^{(c)}, p^{(c)}, \boldsymbol{\eta}^{(c)} | \boldsymbol{y})} \frac{q(r^{(c)}, p^{(c)} | r^{(n)}, p^{(n)}) q(\boldsymbol{\eta}^{(c)} | r^{(c)}, p^{(c)}, \boldsymbol{\eta}^{(n)})}{q(r^{(n)}, p^{(n)} | r^{(c)}, p^{(c)}) q(\boldsymbol{\eta}^{(n)} | r^{(n)}, p^{(n)}, \boldsymbol{\eta}^{(c)})} \right\} \, .$$

Appendix B: Simulation results

k	$L = 10$			$L = 5$		
	25%	50%	75%	25%	50%	75%
	129.9	162.4	197.2	42.5	134.2	200.5
1	147.9	184.0	224.5	114.1	146.5	221.6
	168.8	204.1	258.1	134.8	200.5	245.1
	-39.2	79.1	172.2	-92.1	-8.8	150.8
2	27.3	130.1	222.3	-37.4	40.5	226.4
	98.5	182.8	244.5	-1.1	150.9	284.3
	58.8	136.6	170.8	-8.1	55.1	152.0
3	115.7	150.2	175.1	16.2	96.9	189.5
	137.2	170.8	186.0	50.3	147.9	199.1
	-91.1	28.1	104.2	-107.6	-23.6	87.7
4	-14.8	75.6	123.3	-71.0	23.7	150.0
	36.8	104.7	175.0	-23.2	95.0	191.4
	7.6	114.0	147.0	-22.6	12.9	112.5
5	65.9	139.0	157.3	-10.3	62.7	169.2
	123.8	147.0	184.0	21.9	118.2	208.8

Table B.1: Percentiles of $100\Delta \log(|KL|)$ for the k-step-ahead forecasts (middle entry for each k), $k = 1, \ldots, 5$ and 95% bootstrap confidence limits (first and third lines for each value of k) for these percentiles; left: segment length, $L = 10$; right: $L = 5$.

s	25%	50%	75%	s	25%	50%	75%
	132.7	207.4	291.4		185.5	300.6	360.1
5	170.0	255.8	328.6	34	212.5	335.7	410.5
	209.7	292.1	373.3		314.7	369.7	434.1
	121.3	214.1	304.9		170.1	244.7	297.5
10	177.8	268.9	343.4	40	198.7	285.9	326.5
	223.2	307.1	389.7		259.5	302.1	362.2
	155.9	237.0	321.2		82.1	93.8	103.6
16	181.0	287.1	360.8	48	89.7	99.1	106.1
	240.0	323.1	408.3		94.5	104.9	108.4
	167.4	260.2	336.8		125.5	137.2	146.1
22	193.1	305.1	378.0	49	129.0	142.4	149.2
	263.0	339.6	425.0		138.7	146.7	150.9
	180.8	282.9	350.7		159.6	179.0	192.5
28	201.9	319.9	394.8	50	173.9	188.3	197.9
	296.1	354.8	437.2		183.7	194.3	202.3

Table B.2: Percentiles of $100\Delta\log(L2)$ for the log spectral densities in ten different segments (middle entry for each s) and 95% bootstrap confidence limits (first and third lines for each value of s) for these percentiles.

k	25%	50%	75%	k	25%	50%	75%
	18.4	57.5	84.8		44.3	74.0	113.0
1	35.2	73.9	97.3	6	64.1	100.1	131.1
	58.7	86.0	126.6		80.6	113.0	161.9
	35.4	69.2	100.9		42.0	74.3	111.2
2	53.9	85.6	119.1	7	63.1	95.9	133.3
	69.1	101.1	143.7		76.7	112.8	158.5
	31.3	65.3	101.4		41.3	72.3	118.7
3	51.2	83.9	119.5	8	59.1	93.7	130.5
	67.1	107.2	137.8		73.2	118.7	156.5
	31.4	64.5	108.3		38.4	68.7	108.7
4	55.1	88.8	122.1	9	60.9	94.7	132.9
	64.5	110.5	144.4		69.2	113.3	154.1
	37.5	69.8	110.0		38.9	68.7	109.9
5	63.7	90.8	128.9	10	63.6	97.2	132.3
	72.4	113.9	158.6		75.6	110.9	153.6

Table B.3: Percentiles of $100\Delta\log(|KL|)$ for the k-step-ahead forecasts (middle entry for each k), $k = 1, \ldots, 10$ and 95% bootstrap confidence limits (first and third lines for each value of k) for these percentiles.

4

s	25%	50%	75%	s	25%	50%	75%
	-86.4	-45.8	-5.7		-65.8	-41.3	-14.4
5	-62.9	-24.2	49.0	60	-56.8	-24.9	4.7
	-38.3	-1.8	98.8		-38.5	-14.4	19.3
	47.2	86.5	105.0		7.5	58.8	109.2
15	74.2	94.7	119.0	70	36.6	83.2	146.2
	88.9	105.0	129.1		61.9	109.2	168.9
	49.1	87.0	118.7		6.9	53.9	119.6
25	75.1	108.3	134.2	75	35.1	101.0	156.3
	88.9	120.1	157.8		62.2	134.3	173.1
	34.0	73.3	102.6		28.9	55.3	88.2
35	61.1	90.8	116.5	85	40.5	72.4	117.4
	73.3	102.7	147.7		62.2	97.1	172.8
	24.4	45.8	68.8		40.7	80.8	120.0
45	39.0	61.9	91.4	95	55.3	101.8	158.3
	47.8	75.2	98.3		85.7	122.7	216.3

Table B.4: Percentiles of $100\Delta\log(L2)$ for the log spectral densities in ten different segments (middle entry for each s) and 95% bootstrap confidence limits (first and third lines for each value of s) for these percentiles.

Appendix C: Addition to Section 5.2

To address the question raised by the referee regarding whether our conclusion in Section 5.2 would change if the true model were AR(10) whose characteristic polynomial had complex roots, we generated data from such a model and present some results here. Although only one realization is used, these results are typical of data generated from such a model. The generated data appear in Figure C.1. The true autoregressive parameter values and estimates from two MCMC schemes, one for $R = 1$ and the other for $R = 2$, are given in Table C.1. The mixing functions for the MCMC scheme where $R = 2$ are presented in Figure C.2. Table C.1 shows that the parameter estimates of the one-component mixture are almost identical to the parameter estimates of the first component of the two-component mixture. The parameter estimates of the second component of the two-component mixture are very different. The ϕ's are not significantly different from zero, while the extremely large value of σ^2 of the second component suggests that the purpose of the second component is to increase the noise of the model. However, the weight attached to the second component is very small, except at the boundary of the data where this weight increases to 0.2. This low weight means that the spectral density estimates from the two MCMC schemes, which appear in Figure C.3, are almost identical; only in the last segment is there any discernible difference, and this difference is very small.

Figure C.4 shows the posterior distribution of the number of lags when $R = 1$ (panel (a)) and $R = 2$ (panel (b)), along with the posterior distribution of the number of components (panel (c)).
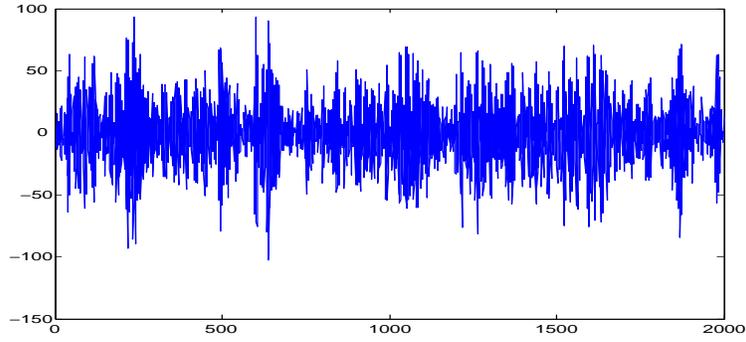
Figure C.1: Data generated from an AR(10) model with complex roots. The values of the autoregressive coefficients are given in Table C.1.

Parameter	True Value	Posterior Mean with $R = 1$ $j = 1$	Posterior Mean with $R = 2$ $j = 1$	$j = 2$
ϕ_0	0.00	-0.04	-0.04	-0.53
ϕ_1	2.15	2.16	2.15	-0.19
ϕ_2	-5.27	-5.31	-5.31	-0.51
ϕ_3	6.95	7.04	7.03	-0.14
ϕ_4	-9.64	-9.84	-9.83	-0.33
ϕ_5	8.85	9.06	9.05	-0.25
ϕ_6	-8.45	-8.72	-8.71	-0.56
ϕ_7	5.30	5.48	5.47	1.40
ϕ_8	-3.50	-3.65	-3.65	-1.59
ϕ_9	1.22	1.28	1.28	1.29
ϕ_{10}	-0.51	-0.54	-0.54	-0.39
σ^2	1.00	1.023	1.023	118.21

Table C.1: Posterior means of the parameters from two MCMC schemes, one where the maximum number of components, $R = 1$ and the other for $R = 2$.
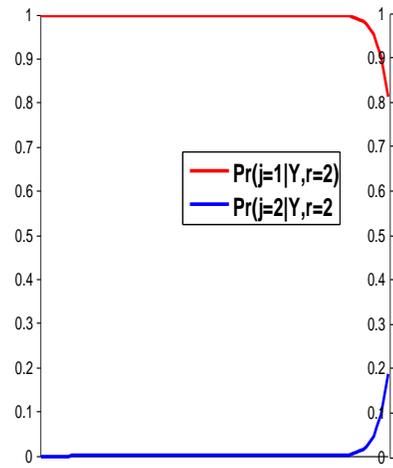
Figure C.2: Plot of the mixing functions for a mixture of two components.
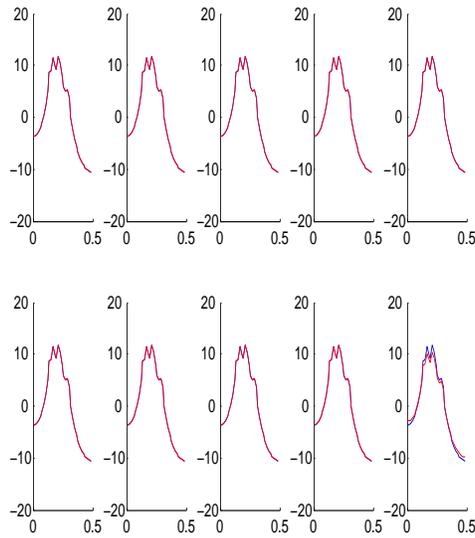
Figure C.3: Plot of the log spectral densities for $R = 1$ (blue) and $R = 2$ (red).
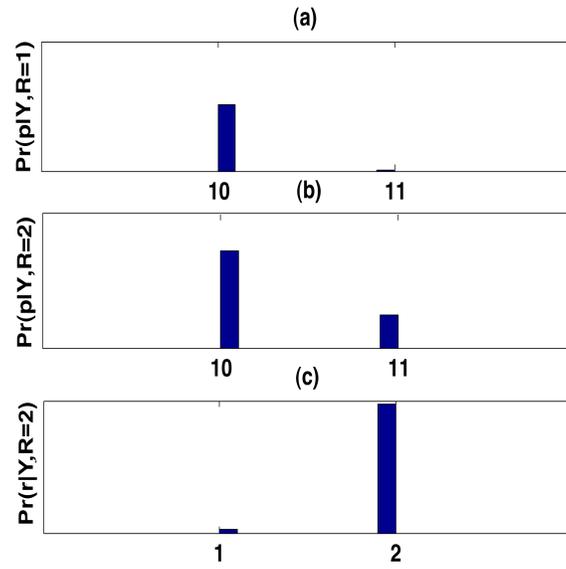
Figure C.4: Posterior distributions: (a) $Pr(p|Y, R = 1)$, (b) $Pr(p|Y, R = 2)$ and (c) $Pr(r|Y, R = 2)$.