Bayesian AR model with Laplace innovations. Results and further proposals.

Jan Šindelář

Department of Adaptive Systems, UTIA Academy of Sciences of the Czech Republic

Department of Probability and Mathematical Statistics, Charles University

e-mail: jenik.sindelar@gmail.com

Modelling time series

$$Y = (Y_1, Y_2, \dots, Y_T) \tag{1}$$

Assumptions:

- All quantities have density.
- Relation between data is described by model $f(Y_t|\Theta, \mathcal{F}_{t-1})$ and time-independent parameters Θ

$$f(Y_t|\Theta, \mathcal{F}_{t-1}) \Leftarrow Y_{t+1} = \Theta\Phi_t + \Sigma W_{t+1} \tag{2}$$

- Prior distribution $f(\Theta|\mathcal{F}_0)$ describes beliefs - conjugate form.

In a model with constant parameters, we use Bayesian data update

$$f(\Theta|Y_t, \mathcal{F}_{t-1}) = \frac{f(Y_t|\Theta, \mathcal{F}_{t-1})f(\Theta|\mathcal{F}_{t-1})}{\int_\Omega f(Y_t|\Theta, \mathcal{F}_{t-1})f(\Theta|\mathcal{F}_{t-1})d\Theta} \tag{3}$$

In a model with Gaussian innovations and GiW (NiG) prior the estimation has two important properties

- Exponential form of density – transforms multiplication into summation of exponents
- Quadratic (polynomial) form in the exponent conserves form when summed with other quadratic form (polynomial of same order)

in a model with Laplace innovations and conjugate prior the first property holds, while the second one fails

In a model with constant parameters, we use Bayesian data update

$$f(\Theta|Y_t,\mathcal{F}_{t-1}) = \frac{f(Y_t|\Theta,\mathcal{F}_{t-1})f(\Theta|\mathcal{F}_{t-1})}{\int_\Omega f(Y_t|\Theta,\mathcal{F}_{t-1})f(\Theta|\mathcal{F}_{t-1})d\Theta} \tag{3}$$

In a model with Gaussian innovations and GiW (NiG) prior the estimation has two important properties

- Exponential form of density – transforms multiplication into summation of exponents
- Quadratic (polynomial) form in the exponent conserves form when summed with other quadratic form (polynomial of same order)

in a model with Laplace innovations and conjugate prior the first property holds, while the second one fails

In a model with constant parameters, we use Bayesian data update

$$f(\Theta|Y_t, \mathcal{F}_{t-1}) = \frac{f(Y_t|\Theta, \mathcal{F}_{t-1})f(\Theta|\mathcal{F}_{t-1})}{\int_{\Omega} f(Y_t|\Theta, \mathcal{F}_{t-1})f(\Theta|\mathcal{F}_{t-1})d\Theta} \tag{3}$$

In a model with Gaussian innovations and GiW (NiG) prior the estimation has two important properties

- Exponential form of density – transforms multiplication into summation of exponents
- Quadratic (polynomial) form in the exponent conserves form when summed with other quadratic form (polynomial of same order)

in a model with Laplace innovations and conjugate prior the first property holds, while the second one fails

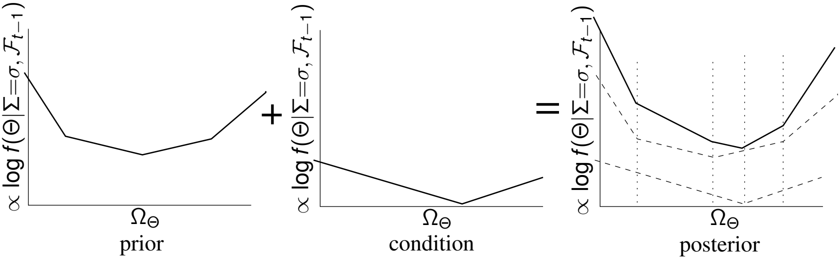


Figure: The principal of update of logarithmized conditional probability density of parameters $f(\Theta|\Sigma, \mathcal{F}_{t-1})$ in a model with Laplace noise via Bayes rule in one dimension. The logarithm transforms multiplication into addition.

and Metropolis-Hastings or

$$J_t = \sum_{P_j \in \mathcal{C}} \frac{\mathrm{vol}(P_j)\Gamma(\nu_t - n - 1)}{2^{\nu_t} n!} \sum_{k=1}^{n+1} \frac{1}{H_k^{\nu_t - n - 1}} \prod_{l=1, k \neq l}^{n+1} \frac{1}{H_l - H_k} \tag{4}$$

where H_k is value of exponent at point $P_k^0 \in \mathcal{C}$.

TABLE I

EMPIRICAL VARIANCES OF SOME ALTERNATIVE LOCATION ESTIMATORS[a]

(Sample Size 20)

tail heaviness - kurtosis →

Estimators	Distributions: Normal	10% 3σ[b]	10% 10σ[c]	Laplace	Cauchy	
Mean	1.00	1.88	11.54	2.10	12,548.0	Gauss
10% trimmed mean	1.06	1.31	1.46	1.60	7.3	
25% trimmed mean	1.20	1.41	1.47	1.33	3.1	
Median	1.50	1.70	1.80	1.37	2.9	Laplace
Gastwirth[d]	1.23	1.45	1.51	1.35	3.1	
Trimean[e]	1.15	1.37	1.48	1.43	3.9	

[a] Abstracted from Exhibit 5 in Andrews, *et al.* [3].
[b] Gaussian Mixture: $.9\Phi(1)+.1\Phi(3)$.
[c] Gaussian Mixture: $.9\Phi(1)+.1\Phi(10)$.
[d] $\tilde{\beta}=.3\beta^*(1/3)+.4\beta^*(1/2)+.3\beta^*(2/3)$, where $\beta^*(\theta)$ is the θth sample quantile.
[e] $\tilde{\beta}=1/4\beta^*(1/4)+1/2\beta^*(1/2)+1/4\beta^*(3/4)$.

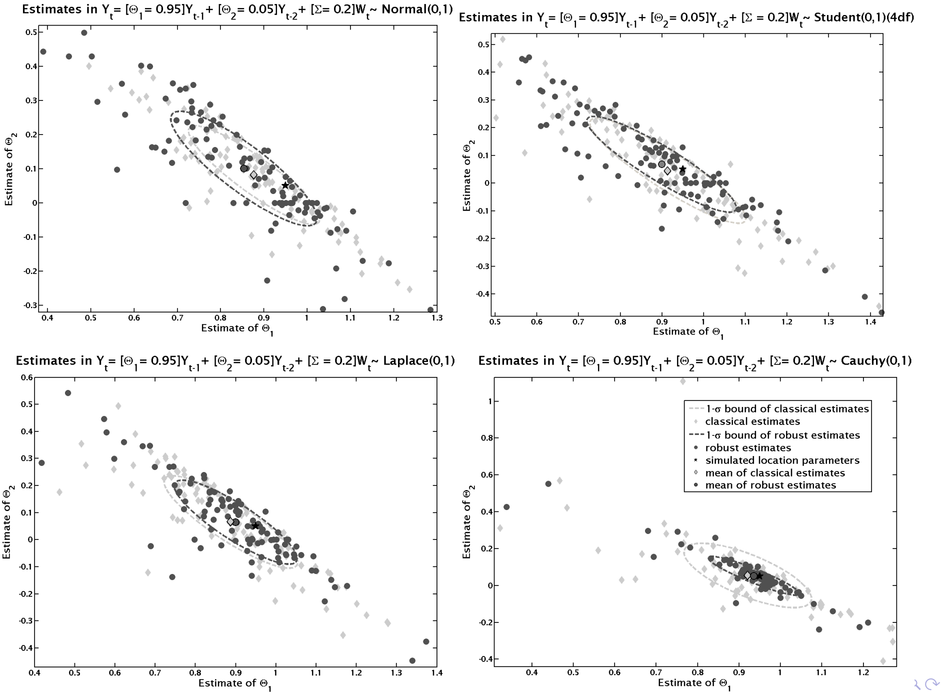
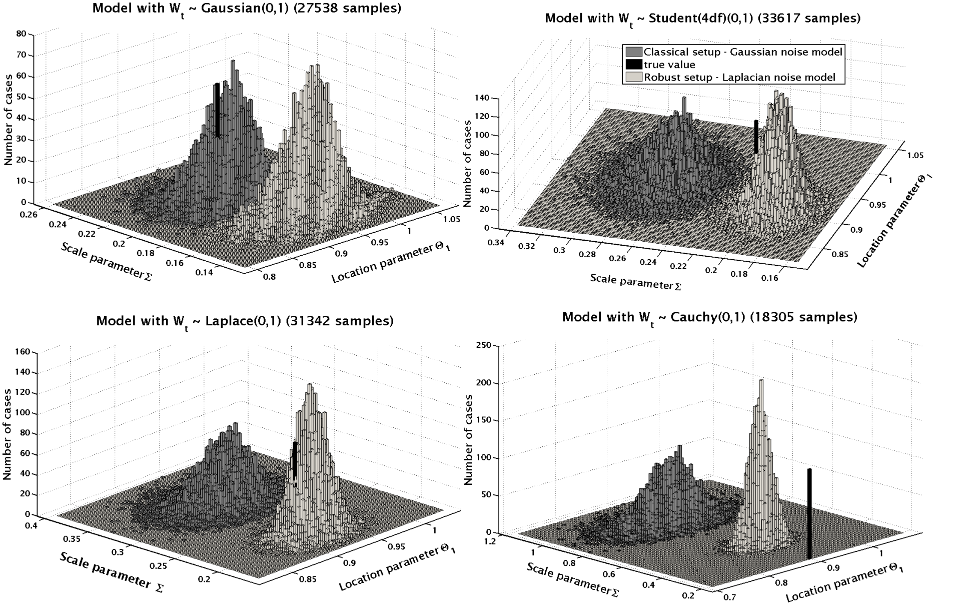

Estimates in Y_t = [Θ_1 = 0.95]Y_t-1 + [Θ_2 = 0.05]Y_t-2 + [Σ = 0.2]W_t ~ Normal(0,1)
Estimates in Y_t = [Θ_1 = 0.95]Y_t-1 + [Θ_2 = 0.05]Y_t-2 + [Σ = 0.2]W_t ~ Student(0,1)(4df)
Estimates in Y_t = [Θ_1 = 0.95]Y_t-1 + [Θ_2 = 0.05]Y_t-2 + [Σ = 0.2]W_t ~ Laplace(0,1)
Estimates in Y_t = [Θ_1 = 0.95]Y_t-1 + [Θ_2 = 0.05]Y_t-2 + [Σ = 0.2]W_t ~ Cauchy(0,1)
Estimate of Θ_1
Estimate of Θ_2
1-σ bound of classical estimates
classical estimates
1-σ bound of robust estimates
robust estimates
simulated location parameters
mean of classical estimates
mean of robust estimates

Table: Descriptive statistics of maximum likelihood estimates of location parameters in model $y_t = (\alpha_1 = 0.95)y_{t-1} + (\alpha_2 = 0.05)y_{t-2} + 0.2e_t$ for different noises and two expected models.

model\real	$\mathcal{N}(0,1)$		$t_4(0,1)$	
	$\mathrm{mean}(\hat{\alpha}_1, \hat{\alpha}_2)$	$\mathrm{cov}(\hat{\alpha}_1, \hat{\alpha}_2)$	$\mathrm{mean}(\hat{\alpha}_1, \hat{\alpha}_2)$	$\mathrm{cov}(\hat{\alpha}_1, \hat{\alpha}_2)$
Gaussian	$\begin{pmatrix} .8771 \\ .0820 \end{pmatrix}$	$\begin{pmatrix} .0229 & -.0196 \\ -.0196 & .0208 \end{pmatrix}$	$\begin{pmatrix} .9136 \\ .0448 \end{pmatrix}$	$\begin{pmatrix} .0350 & -.0322 \\ -.0322 & .0357 \end{pmatrix}$
Laplacian	$\begin{pmatrix} .8535 \\ .1011 \end{pmatrix}$	$\begin{pmatrix} .0285 & -.0247 \\ -.0247 & .0281 \end{pmatrix}$	$\begin{pmatrix} .8988 \\ .0682 \end{pmatrix}$	$\begin{pmatrix} .0321 & -.0281 \\ -.0281 & .0297 \end{pmatrix}$

model\real	$\mathcal{L}(0,1)$		$\mathcal{C}(0,1)$	
	$\mathrm{mean}(\hat{\alpha}_1, \hat{\alpha}_2)$	$\mathrm{cov}(\hat{\alpha}_1, \hat{\alpha}_2)$	$\mathrm{mean}(\hat{\alpha}_1, \hat{\alpha}_2)$	$\mathrm{cov}(\hat{\alpha}_1, \hat{\alpha}_2)$
Gaussian	$\begin{pmatrix} .8876 \\ .0661 \end{pmatrix}$	$\begin{pmatrix} .0275 & -.0247 \\ -.0247 & .0290 \end{pmatrix}$	$\begin{pmatrix} .9207 \\ .0550 \end{pmatrix}$	$\begin{pmatrix} .0236 & -.0193 \\ -.0193 & .0303 \end{pmatrix}$
Laplacian	$\begin{pmatrix} .8998 \\ .0652 \end{pmatrix}$	$\begin{pmatrix} .0232 & -.0205 \\ -.0205 & .0239 \end{pmatrix}$	$\begin{pmatrix} .9354 \\ .0532 \end{pmatrix}$	$\begin{pmatrix} .0121 & -.0106 \\ -.0106 & .0110 \end{pmatrix}$



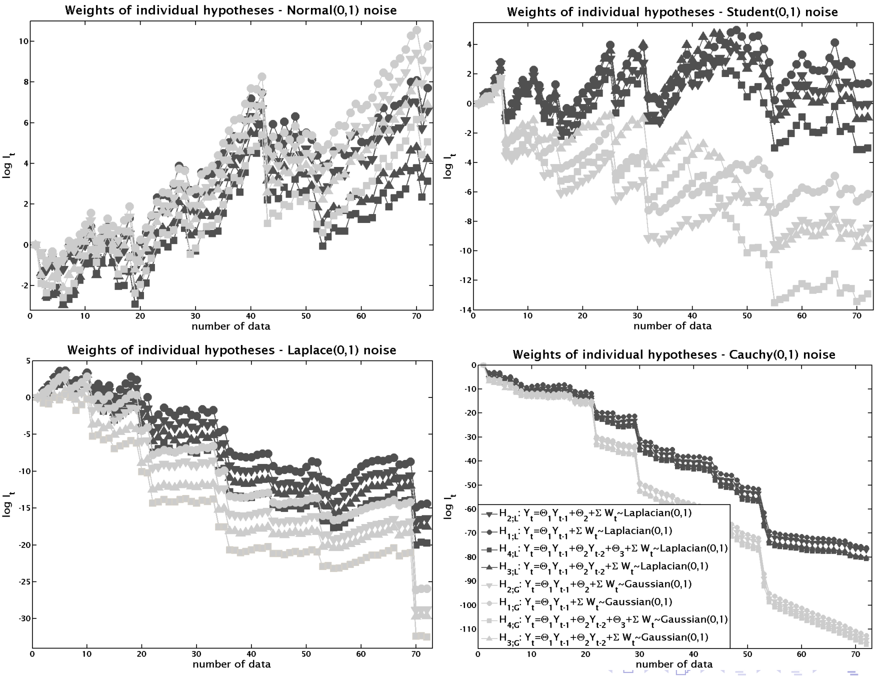
Model with W_t ~ Gaussian(0,1) (27538 samples)
Number of cases
Scale parameter Σ
Location parameter Θ_1
Model with W_t ~ Student(4df)(0,1) (33617 samples)
Classical setup - Gaussian noise model
true value
Robust setup - Laplacian noise model
Number of cases
Scale parameter Σ
Location parameter Θ_1
Model with W_t ~ Laplace(0,1) (31342 samples)
Number of cases
Scale parameter Σ
Location parameter Θ_1
Model with W_t ~ Cauchy(0,1) (18305 samples)
Number of cases
Scale parameter Σ
Location parameter Θ_1

$$
\begin{aligned}
H_1 : Y_t &= \Theta_1 Y_{t-1} + \sigma W_t \\
H_2 : Y_t &= \Theta_1 Y_{t-1} + \Theta_2 + \sigma W_t \\
H_3 : Y_t &= \Theta_1 Y_{t-1} + \Theta_2 Y_{t-2} + \sigma W_t \\
H_4 : Y_t &= \Theta_1 Y_{t-1} + \Theta_2 Y_{t-2} + \Theta_3 + \sigma W_t
\end{aligned}
$$

$$
R : Y_t = [\Theta_1 = 0.9]\, Y_{t-1} + [\Sigma = 0.2]\, W_t
$$

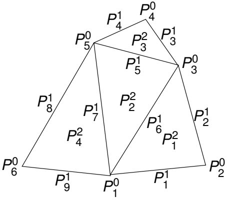
real\hypo	$H_1{+}W_t{\sim}\mathcal{L}$	$H_2{+}W_t{\sim}\mathcal{L}$	$H_3{+}W_t{\sim}\mathcal{L}$	$H_4{+}W_t{\sim}\mathcal{L}$	$H_1{+}W_t{\sim}\mathcal{N}$	$H_2{+}W_t{\sim}\mathcal{N}$	$H_3{+}W_t{\sim}\mathcal{N}$	$H_4{+}W_t{\sim}\mathcal{N}$
$\mathcal{N}^{a}(0,1)$	0.486	0.013	0.164	0.004	0.29	0.001	0.043	$5.3{\cdot}10^{-7}$
$\mathcal{N}^{b}(0,1)$	0.191	0.007	0.006	0.002	0.63	0.013	0.093	0.002
$t_4(0,1)$	0.88	0.026	0.092	0.003	$5.4{\cdot}10^{-6}$	$3.4{\cdot}10^{-9}$	$2.8{\cdot}10^{-8}$	$1.5{\cdot}10^{-11}$
$\mathcal{L}(0,1)$	0.876	0.048	0.073	0.004	$2.6{\cdot}10^{-7}$	$3{\cdot}10^{-10}$	$9.1{\cdot}10^{-11}$	$7.9{\cdot}10^{-13}$
$\mathcal{C}(0,1)$	0.87	0.096	0.031	0.004	$7.0{\cdot}10^{-18}$	$8.1{\cdot}10^{-21}$	$4.5{\cdot}10^{-20}$	$5.8{\cdot}10^{-23}$

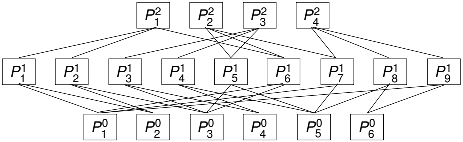
[a,b] The choice of hypotheses for data from model with Normally distributed noise depends on the chosen prior distribution. We therefore give results for a prior specified by density $f(\Lambda|\mathcal{F}_0)$ the result is given in line $\mathcal{N}^a(0,1)$ and also a result with prior specified by $f(\Lambda|\mathcal{F}_1)$ (posterior after one update) the result given in line $\mathcal{N}^b(0,1)$.



Weights of individual hypotheses - Normal(0,1) noise
Weights of individual hypotheses - Student(0,1) noise
Weights of individual hypotheses - Laplace(0,1) noise
Weights of individual hypotheses - Cauchy(0,1) noise
$\log l_t$
number of data
$H_{2;L}$: $Y_t=\Theta_1 Y_{t-1}+\Theta_2+\Sigma\ W_t$~Laplacian(0,1)
$H_{1;L}$: $Y_t=\Theta_1 Y_{t-1}+\Sigma\ W_t$~Laplacian(0,1)
$H_{4;L}$: $Y_t=\Theta_1 Y_{t-1}+\Theta_2 Y_{t-2}+\Theta_3+\Sigma\ W_t$~Laplacian(0,1)
$H_{3;L}$: $Y_t=\Theta_1 Y_{t-1}+\Theta_2 Y_{t-2}+\Sigma\ W_t$~Laplacian(0,1)
$H_{2;G}$: $Y_t=\Theta_1 Y_{t-1}+\Theta_2+\Sigma\ W_t$~Gaussian(0,1)
$H_{1;G}$: $Y_t=\Theta_1 Y_{t-1}+\Sigma\ W_t$~Gaussian(0,1)
$H_{4;G}$: $Y_t=\Theta_1 Y_{t-1}+\Theta_2 Y_{t-2}+\Theta_3+\Sigma\ W_t$~Gaussian(0,1)
$H_{3;G}$: $Y_t=\Theta_1 Y_{t-1}+\Theta_2 Y_{t-2}+\Sigma\ W_t$~Gaussian(0,1)

Complexity:





$$N_n(r,t) = \sum_{p=n-r}^{n} \binom{p}{n-r} \binom{t}{p}$$

$$\lim_{t\to\infty} N_n(r,t) < t^n \quad \forall r \in \{0,1,\ldots,n\}$$

Thank you for your attention.