Flexible modeling of conditional distributions using smooth mixtures of asymmetric student t densities

Feng Li
(*joint with Mattias Villani and Robert Kohn*)

Department of Statistics
Stockholm University

September 16, 2009

Outline of the talk

1. **Mixture distributions**
2. **ME, SMR and SAGM models**
3. **Smooth mixture of asymmetric student's t densities**
4. **Application to daily S&P 500 returns**

Mixture distributions

- For a given x, a mixture distribution $p(y|x)$ is a finite mixture

$$\sum_{k=1}^{K} \omega_k f_k\left(y_i|\theta_k\right),\ i=1,...,n.$$

- Latent variable formulation for MCMC

$$\Pr\left(s_i=k\right)=\omega_k$$
$$y_i|\left(s_i=k\right)\sim f_k\left(y_i|\theta_i\right)$$

- Two-block Gibbs sampler
  - Sample $s=(s_1,...,s_n)$ conditional on $(\theta_1,...,\theta_k)$.
  - Sample each θ_k conditional on the allocation s.
- A smooth mixture model is a finite mixture density with weights that are smooth function of the covariates, e.g

$$\omega_k\left(x\right)=\frac{\exp\left(x'\gamma_k\right)}{\sum_{r=1}^{K}\exp\left(x'\gamma_r\right)}$$

Stockholm University

ME, SMR and SAGM models

- Mixture-of-Experts (ME) (Jacobs *et al.* (1991))
  - A mixture of regressions where the mixing probabilities are functions of covariates.
  - Flexibly model the mean regression and frequently used in the machine learning literature.
  - The components are often linear homoscedastic regressions or even constant functions.
  - *simple-and-many* approach.
- Smoothly Mixing Regression (SMR) (Geweke & Keane (2007))
  - A generalization of the ME model for regression density estimation
  - Fail to fit heteroscedastic data even with a very large number of components
- Smooth Adaptive Gaussian Mixtures (SAGM) (Villani *et al.* (2008))
  - A smooth finite mixture of Gaussian densities with the mixing probabilities.
  - The **mixing probabilities**, the **components means** and **components variances** modeled as functions of the covariates.
  - Bayesian variable selection are in all three sets of covariates.
  - *complex-but-few* approach — Enough flexibility is used within the mixture components so that the number of components can be kept to a minimum.

Stockholm University

Smooth mixture of asymmetric student's t densities

The model

- The split-t density is

$$c \cdot \kappa\left(\mu, \phi, v\right) I\left(y \leq \mu\right) + c \cdot \kappa\left(\mu, \lambda\phi, v\right) I\left(y > \mu\right),$$

  where $\kappa\left(\mu, \phi, v\right) = \left(\frac{v}{v + \frac{(y-\mu)^2}{\phi^2}}\right)^{(v+1)/2}$ is the kernel of student t density and c is the normalization constant.
- Each of the four parameters μ, ϕ, λ and ν are connected to covariates as

$$\begin{aligned} \mu &= \beta_{\mu 0} + x_t' \beta_\mu \\ \ln \phi &= \beta_{\phi 0} + x_t' \beta_\phi \\ \ln \lambda &= \beta_{\lambda 0} + x_t' \beta_\lambda \\ \ln v &= \beta_{v 0} + x_t' \beta_v \end{aligned}$$

  but any smooth link function can equally well be used in the MCMC methodology.
- This make it possible e.g. to have the degrees of freedom smoothly varying over covariate space; to capture skewness and excess kurtosis with the components.
- *Common* components if $\beta_\mu = \beta_\phi = \beta_\lambda = \beta_v$, else *separate* components.

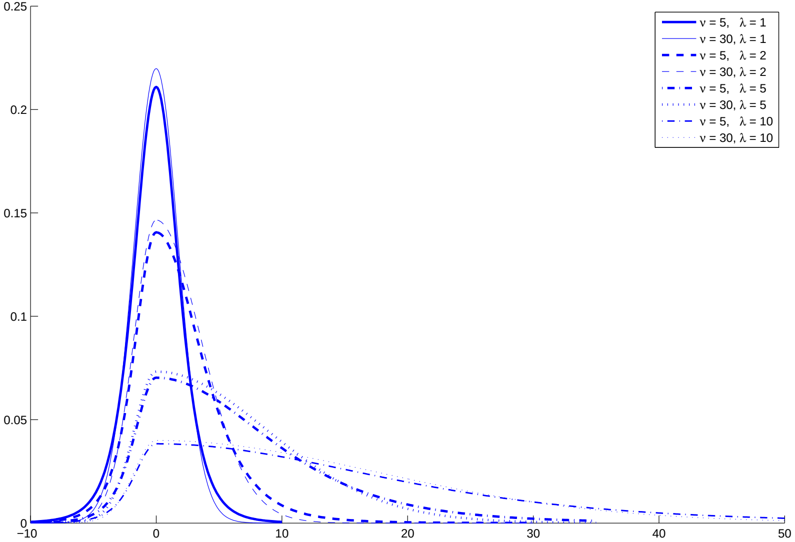

Figure: Graphical display of the split-t density with location parameter $\mu = 0$ and scale parameter $\lambda = 1.8$.

Smooth mixture of asymmetric student's t densities

Discussion — Why not over-fit?

- The prior
  - We use an easy specified prior, $\beta|\mathcal{I} \sim N(0, \tau_\beta^2 I)$, where $\mathcal{I}$ is the covariate indicators.
  - We investigate the sensitivity of the posterior inferences and model comparison with respect to τ_β.
  - One can use the g-prior $\beta \sim N(0, \tau_\beta^2 (X'X)^{-1})$ (Zellner, 1986) which is less appealing in a mixture context.
- Variable selection (details in next page)
  - Investigate the importance of covariates.
  - More efficient.
- Automatically add components to make each component simpler.
- Evaluating the out-of-sample log predictive density score(LPDS) – details in "model comparison" .

- In a general regression model, the likelihood function is $p\left(y|\beta\right)=\prod_{i=1}^{n}p\left(y_i|\phi_i\right)$ where $k\left(\phi_i\right)=x_i'\beta$ (link function).
- We need first two derivatives of $\ln p(y_i|\phi_i)$ with respect to ϕ_i.
- We do Bayesian variable selection within MCMC.
  - Set up variable selection indicator $\mathcal{I}=(I_1,...,I_n)$ where $I_i=1$ indicates X_i are in the model and $I_i=0$ means $\beta_i=0$.
  - Sample β and I by using finite-step Newton's method. We only iterate a few steps($\leq$3).
  - Dimension might change here. But exploits that $k(\phi_i)=x_i'\beta$ always has the same dimension (Villani *et at.* 2008).

Smooth mixture of asymmetric student's t densities

Model comparison

- Why not marginal likelihood?
  - The key quantity is Bayesian model comparison is the marginal likelihood.
  - The marginal likelihood is sensitive to the choice of prior, which is especially true when the prior is not very informative (Kass, 1993).
- We use B-fold cross-validation of the log predictive density score(LPDS)
  - $B^{-1} \sum_{b=1}^{B} \ln p\left(\tilde{y}_b | \tilde{y}_{-b}, x\right)$
  - Compute the LPDS for ME, SMR, SAGM and our split model with different components.
  - Compare the differences of LPDS.

Application to daily S&P 500 returns
The data

- Response variable: Daily returns from S&P 500 index.
- Covariates
  - **LastDay, LastWeek, LastMonth**, Moving average of returns from the previous one, five and 20 trading days respectively.
  - **CloseAbs80, CloseAbs95**, Geometrically declining average of past returns $(1-\varphi)\sum_{s=0}^{\infty}\varphi^s|y_{t-2-s}|$ with φ of .80 and .95 respectively.
  - **CloseSqr80, CloseSqr95**, The square root of $(1-\varphi)\sum_{s=0}^{\infty}\varphi^s y_{t-2-s}^2$ with φ of .80 and .95 respectively.
  - **MaxMin80, Maxmin95**, Information of volatility – $(1-\varphi)\sum_{s=0}^{\infty}\varphi^s\left(\ln p_{t-1-s}^{(h)} - \ln p_{t-1-s}^{(l)}\right)$ with φ of .80 and .95 respectively.
- The models are estimated using 4646 trading days from 1990-Jan-01 to 2008-May-29(before finical crisis).
- The models are evaluated out-of-sample on the 199 trading days from 2008-May-30 to 2009-Mar-13(finical crisis period).

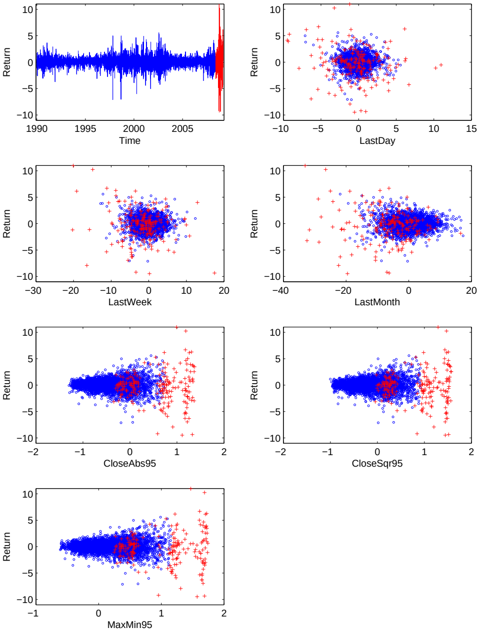

Figure: Time series plot of Return(up-left) and scatter plots of Return against a covariate(others) for S&P500 (1990-Jan-01 – 2009-Mar-13).

Application to daily S&P 500 returns

Results

- A normalized residuals is defined as $\Phi^{-1}(F(y_t))$, where $F(y_t)$ is the cumulative predictive distribution. If the model is correct, the normalized residuals should be *iid* $N(0,1)$.
- The LPDS is reported for different models.
- Posterior summary of the one-component split-t model

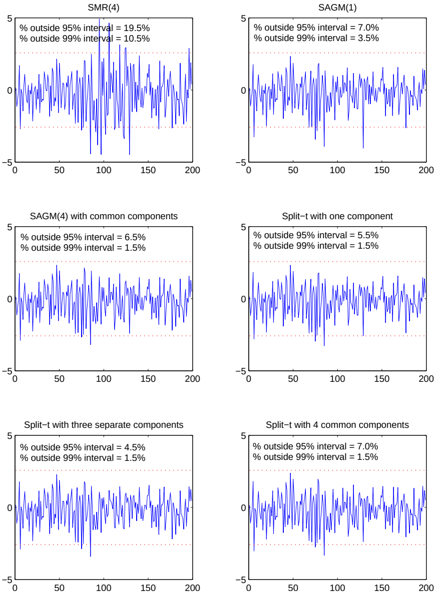

Figure: The 199 normalized residuals in the evaluation sample over time and the 99% probability intervals under the N(0, 1).

Model	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	Max n.s.e.
SMR	−1044.78	−638.89	−505.74	−487.11	−489.19	0.98 (3)
+ Skew	−540.91	−525.07	−513.85	−506.68	−506.13	0.82 (2)
+ DF	−544.00	−518.71	−498.93	−500.14	−494.29	0.89 (1)
+ Skew + DF	−530.86	−504.63	−498.03	−498.83	−496.87	0.88 (5)
SAGM Common	−477.73	−473.10	−473.12	−470.30	−472.86	0.26 (2)
+ Skew	−474.18	−467.29	−468.75	−467.93	−467.22	0.35 (4)
+ DF	−474.74	−472.92	−470.51	−469.40	−468.87	0.34 (4)
+ Skew + DF	**−472.37**	−468.92	−469.30	−466.21	**−465.86**	0.53 (4)
SAGM Separate		−469.21	−469.50	−470.53	−471.02	0.49 (3)
+ Skew		−468.48	−466.93	−467.48	−468.02	0.58 (4)
+ DF		−469.08	−469.24	**−462.03**	−467.78	0.72 (5)
+ Skew + DF		**−466.84**	**−462.56**	−462.47	−474.58	0.74 (5)
GARCH(1,1)	−479.03					
t-GARCH(1,1)	−477.39					

Table: Evaluating the out-of-sample log predictive density score (LPDS)

Parameters	Mean	Stdev	Post.Incl.	IF
		Location μ		
Const	0.084	0.019	–	9.919
		Scale ϕ		
Const	0.402	0.035	–	7.125
LastDay	-0.190	0.120	0.036	0.903
LastWeek	**-0.738**	**0.193**	**0.985**	**18.519**
LastMonth	**-0.444**	**0.086**	**0.999**	**4.133**
CloseAbs95	0.194	0.233	0.035	1.445
CloseSqr95	0.107	0.226	0.023	2.715
MaxMin95	**1.124**	**0.086**	**1.000**	**6.012**
CloseAbs80	0.097	0.153	0.013	–
CloseSqr80	0.143	0.143	0.021	–
MaxMin80	-0.022	0.200	0.017	–
		Degrees of freedom ν		
Const	2.482	0.238	–	5.708
LastDay	0.504	0.997	0.112	2.899
LastWeek	**-2.158**	**0.926**	**0.638**	**5.463**
LastMonth	0.307	0.833	0.089	5.560
CloseAbs95	0.718	1.437	0.229	3.020
CloseSqr95	1.350	1.280	0.279	2.758
MaxMin95	1.130	1.488	0.222	6.564
CloseAbs80	0.035	1.205	0.101	2.789
CloseSqr80	0.363	1.211	0.112	3.330
MaxMin80	-1.672	1.172	0.254	4.178
		Skewness λ		
Const	-0.104	0.033	–	10.423
LastDay	-0.159	0.140	0.027	1.170
LastWeek	-0.341	0.170	0.135	8.909
LastMonth	-0.076	0.112	0.016	–
CloseAbs95	-0.021	0.096	0.008	–
CloseSqr95	-0.003	0.108	0.006	–
MaxMin95	0.016	0.075	0.008	–
CloseAbs80	0.060	0.115	0.009	–
CloseSqr80	0.059	0.111	0.010	–
MaxMin80	0.093	0.096	0.013	–

Table: Posterior summary of the one-component split-t model

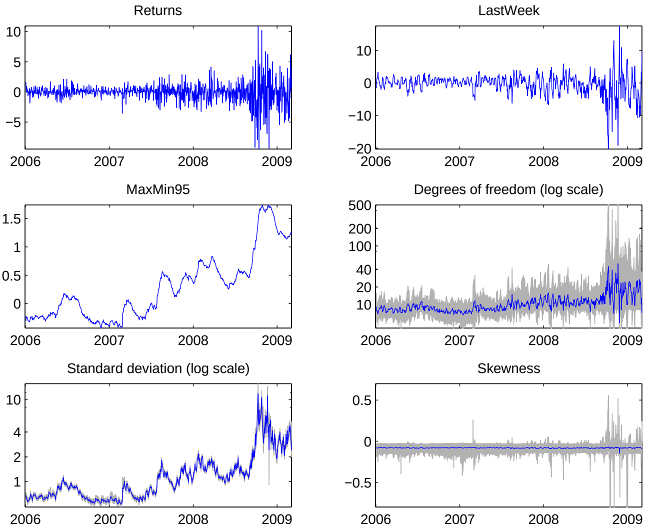

Figure: Time series plot of the posterior median and 95% probability intervals for some moments of the return distribution. The posterior distribution is based on the full sample up to March 13, 2009.

Thank you!