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

Robert Kohn (joint with Mattias Villani and Feng Li)

UNSW

Aim

- Aim: A model for the conditional density $p(y|x)$, where y is univariate and x is a possibly high-dimensional vector of covariates.
- Density estimation problem, but there is a new density for every x.
- Flexibility 1: $p(y|x)$ should be a flexible density for any given x. Finite mixtures.
- Flexibility 2: $p(y|x)$ should be flexible across the covariate space. Smooth mixtures.

Outline of the talk

- Smooth mixtures
- Smooth Adaptive Gaussian Mixtures (SAGM)
- Smooth mixture of asymmetric Student t densities
- Application to daily SP500 returns.

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

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

- Latent variable formulation for MCMC:

$$Pr(s_i = k) = \omega_k$$

$$y_i|(s_i = k) \sim f_k(y_i|\theta_k).$$

- Two-block Gibbs sampler: i) Sample $s = (s_1, ..., s_n)$ conditional on $\theta_1, ..., \theta_k$. ii) Sample each θ_k conditional on the allocation s.

Smooth Mixtures of Normals

- $p(y|x)$ can be modelled as a regression with mixture of Gaussian errors. Same shape of the density for all x. Not flexible enough.
- More flexibility if mixing probabilities ω_k depend on x.
- Example: **Smoothly Mixing Regressions (SMR)** in Geweke and Keane (2007, JoE)

$$p(y|x) = \sum_{k=1}^{K} \omega_k(x) \cdot N\left[y|\mu_k(x), \sigma_k^2\right]$$

- The **mixing function** $\omega_k(x)$ can be modelled as multinomial logit

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

- Smooth mixtures gives a probabilistic partitioning of the covariate space.
- SMRs can generate heteroscedasticity, but not enough ...

Smooth Adaptive Gaussian Mixtures (SAGM)

- SAGM (Villani, Kohn and Giordani, 2009, JoE) extends the SMR in three ways:
  - **heteroscedastic components**: also the σ_k^2 depend on x
  - **variable selection** in mean, variance and mixing function
  - **spline components** with special variable selection prior for mixtures (additive and surfaces)
- General **SAGM** model

$$p(y|x) = \sum_{k=1}^{K} \pi_k(z) \cdot N\left[y|v'\alpha_k, exp(w'\delta_k)\right],$$

  where z, v and w are subsets or transformations of the covariates in x.
- Simulations and real data show that having **heteroscedastic components (i.e. σ_k^2 depending on x) can be crucial for heteroscedastic data**. We obtain a better, more interpretable, out-of-sample fit with substantially less components. Variable selection is very useful, especially when x is high-dimensional.

Taming the Beast

- These models are typically over-parametrized. Why don't we over-fit?
- The **prior**.
- Focus on **out-of-sample performance**.
- **Variable selection**. The MCMC algorithm automatically removes unnecessary covariates.
- **Automatic selection of spline knots**. Knots in component k are automatically 'deleted' in regions where $\omega_k(x)$ is small.
- **Self-adjustment**. Adding components makes all components simpler.
- We may simplify the model using e.g. a **common variance function**: $\sigma_k^2(w) = \sigma_k^2 \cdot exp(w'\delta)$ or common variable selection indicators (a covariate is either in all components or in none of them).

Smooth mixtures of asymmetric student t densities

- Extending SAGM: mixture components are asymmetric student t densities. The next step in the **Complex-and-Few** approach.
- Split-t (Geweke, 1989)

$$f(y) \propto \begin{cases} t(y|\mu,\phi,\nu) & \text{if } y \leq \mu \\ t(y|\mu,\lambda\phi,\nu) & \text{if } y > \mu \end{cases},$$

  where $t(y|\mu,\phi,\nu)$ denotes the student t density with v degrees of freedom and variance $\phi\nu/(\nu-2)$. Continuous at μ.
- All four parameters are linked to covariates (possibly splines)

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

- Bayesian variable selection in all four split-t parameters and in the mixing function $\omega(x)$.

Inference - Finite-Step Newton Proposals

- Consider the likelihood function in the following general regression model

$$p(y|\beta) = \prod_{i=1}^{n} p(y_i|\phi_i)$$

where the parameters ϕ_i are connected to covariates through a link function (e.g. linear, log)

$$k(\phi_i) = x_i'\beta.$$

- Metropolis-Hastings is typically simple: use Newton method to obtain the posterior mode of β and the Hessian at the mode. Multivariate student-t distribution proposal. Only need first two derivatives of $\ln p(y_i|\phi_i)$ wrt ϕ_i.

- What if the model has two sets of regression coefficients, β and δ? How to sample from the full conditionals $p(\beta|\delta, y)$ and $p(\delta|\beta, y)$?
- Naive time-consuming way: Newtons method in each updating step.
- Smarter way: Finite-step Newton. Don't iterate all the way to the mode, a few steps are enough. But set up the MH accept prob right!
- Bayesian variable selection: Sample β and variable selection indicators jointly with MCMC ($I_j = 0$, means $\beta_j = 0$).
- How to propose β conditional on I? Finite-step Newton with variable dimension. Exploits that $k(\phi_i) = x_i'\beta$ always has the same dimension, and $k(\phi_{ic}) = x_i'\beta_c$ and $k(\phi_{ip}) = x_i'\beta_p$ are expected to be quite close.
- Bottom line: This is a general method for simulating from an arbitrary number of full conditional posteriors, and simultaneously do variable selection in each parameter block. Efficient. Fast.

The Distribution of S&P500 Returns

- Our data: daily returns from the S&P500 index $[y_t = 100 \ln(p_t / p_{t-1})]$ during
  - Estimation: January 1, 1990 - May 28, 2008.
  - Evaluation: May 29, 2008 - March 13, 2009.
- Geweke and Keane (2007) applied an SMR with two covariates: $x_{1t} = y_{t-1}$, and

$$x_{2t} = (1 - \varphi) \sum_{s=0}^{\infty} \varphi^s \left| y_{t-2-s} \right| .$$

- We add seven additional covariates including LastWeek, LastMonth and MaxMin

$$(1 - \varphi) \textstyle\sum_{s=0}^{\infty} \varphi^s (\ln p_{t-1-s}^{(h)} - \ln p_{t-1-s}^{(l)}),$$

  where $p_t^{(h)}$ and $p_t^{(l)}$ are the highest and lowest prices at day t.
- Model evaluation using out-of-sample LPDS

$$LPDS = \sum_{t \in \mathcal{T}} \ln p(y_t | x_t).$$

S&P500 - Model evaluation. May 29, 2008 - March 13, 2009

Model	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
SMR	-1044.78	-638.89	-505.74	-487.11	-489.19
+ Skew	-540.91	-525.07	-513.85	-506.68	-506.13
+ DF	-544.00	-518.71	-498.93	-500.14	-494.29
+ Skew + DF	-530.86	-504.63	-498.03	-498.83	-496.87
SAGM Common	-477.73	-473.10	-473.12	-470.30	-472.86
+ Skew	-474.18	-467.29	-468.75	-467.93	-467.22
+ DF	-474.74	-472.92	-470.51	-469.40	-468.87
+ Skew + DF	$-\mathbf{472.37}$	-468.92	-469.30	-466.21	$-\mathbf{465.86}$
SAGM Separate		-469.21	-469.50	-470.53	-471.02
+ Skew		-468.48	-466.93	-467.48	-468.02
+ DF		-469.08	-469.24	$-\mathbf{462.03}$	-467.78
+ Skew + DF		$-\mathbf{466.84}$	$-\mathbf{462.56}$	-462.47	-474.58
GARCH(1,1)	-479.03		t-GARCH(1,1)	-477.39	

Posterior inferences for the one-component model

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	–