Bayesian Nonparametric Calibration and Combination of Predictive Distributions

Roberto Casarin[1] Tilmann Gneiting[2] Francesco Ravazzolo[3]

[1]University of Venice

[2]Heidelberg Institute for Theoretical Studies

[3]Norges Bank and BI Norwegian Business School

Wien, 16th January, 2014

Calibration and **Combination**

GneBalRaf07(JRSSB), GneRan10 (JRSSB) and GneRan13 (EJS)

Linear combination model

- Let $F_{jt}(y)$, $j = 1, \ldots, M$, be a set of predictive cdfs from different models, for a variable of interest y_t, and conditional on the information set, $\mathcal{F}_{t-1}$, available at time $t-1$.
- Let $\Delta_{[0,1]^M}$ be the standard M-dimensional simplex, that is $\Delta_{[0,1]^M}$ $= \{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M) \mid \sum_{i=1}^{M} \omega_i = 1,\ \omega_i \geq 0,\ i = 1, \ldots, M\}$
- Our combination model is

$$H(y_t|\boldsymbol{\omega}) = \sum_{i=1}^{M} \omega_i F_{it}(y_t) \tag{1}$$

with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M) \in \Delta_{[0,1]^M}$ the combination weights.

Calibration and Combination

Probabilistic calibrated combination

- Let $g : [0, 1] \mapsto [0, 1]$ be a calibration function
- A (combination and) calibration model is

$$F_t(y_t) = g\left(H(y_t|\omega)\right), \tag{2}$$

In terms of densities

$$f_t(y_t) = c\left(H(y_t|\omega)\right) h(y_t|\omega), \tag{3}$$

where c is the first order derivative of g and

$$h(y_t|\omega) = \sum_{i=1}^{M} \omega_i f_{it}(y_t) \tag{4}$$

A beta calibration model

GneRan13 suggests to choose g as the incomplete beta function $B_{\alpha,\beta}(z)$, that is also the cdf of a beta distribution, $\mathcal{B}e(\alpha,\beta)$, with parameters $\alpha > 0$, $\beta > 0$.

A beta calibration model

$$f_t(y_t) = f_{\alpha,\beta}\left(H(y_t|\boldsymbol{\omega})\right) h(y_t|\boldsymbol{\omega}), \tag{5}$$

where α, β and $\boldsymbol{\omega}$ are the parameters of the probabilistic calibrated combination model, and with $f_{\alpha,\beta}(x)$ the pdf of a beta distribution, $\mathcal{B}e(\alpha,\beta)$, evaluated at x.

Motivating example - Multimodality (1)

Data generating process

Data are generated from

$$y_t \overset{i.i.d.}{\sim} p_1\mathcal{N}(-2, 0.25) + p_2\mathcal{N}(0, 0.25) + p_3\mathcal{N}(2, 0.25), \quad t = 1, \dots, 1000,$$

Predictive Models

Two Gaussian ($\mathcal{N}(\mu, \sigma^2)$) predictive models:

1. $(\mu, \sigma^2) = (-1, 1)$
2. $(\mu, \sigma^2) = (2, 1)$

We denote the pdf and cdf with$\varphi(x|\mu, \sigma^2)$ and $\Phi(x|\mu, \sigma^2)$, respectively.

Motivating example - Multimodality (2)

We compare the following alternative models

1. Ideal model (I)

$$f(y) = p_1\varphi(y|-2, 0.25) + p_2\varphi(y|2, 0.25) + p_3\varphi(y|2, 0.25),$$

2. Non-calibrated model (NC)

$$f(y|\boldsymbol{\theta}) = \omega\varphi(y|-1, 1) + (1-\omega)\varphi(y|2, 1), \quad \boldsymbol{\theta} = \omega, \omega = 0.5$$

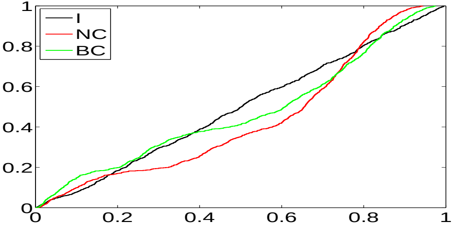
3. Beta calibration model (BC)

$$f(y|\boldsymbol{\theta}) = f_{\alpha,\beta}\left(H(y|\omega)\right) h(y|\omega), \quad \boldsymbol{\theta} = (\alpha, \beta, \omega)$$

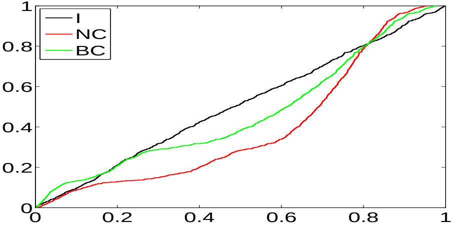
where $H(y|\omega) = \omega\Phi(y|-1, 1) + (1-\omega)\Phi(y|2, 1)$ and $h(y|\omega) = \frac{dH}{dy}$.

Motivating example - Multimodality (3)

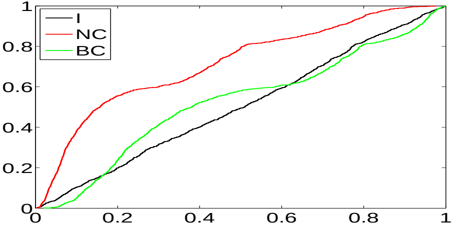
$\mathbf{p} = (1/5, 1/5, 3/5)$



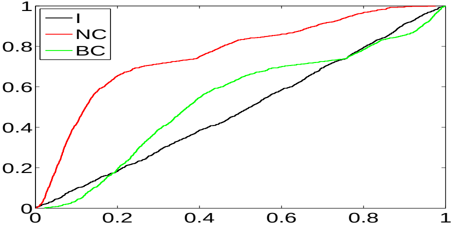
$\mathbf{p} = (1/7, 1/7, 5/7)$



$\mathbf{p} = (3/5, 1/5, 1/5)$

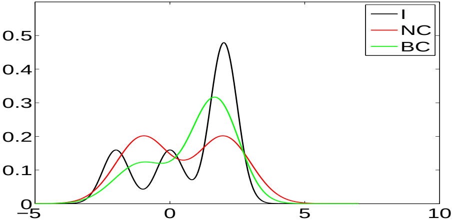


$\mathbf{p} = (5/7, 1/7, 1/7)$

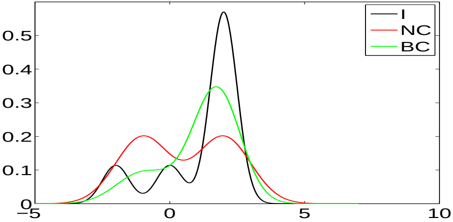


Motivating example - Multimodality (4)

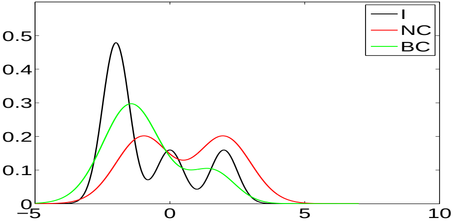
$\mathbf{p} = (1/5, 1/5, 3/5)$



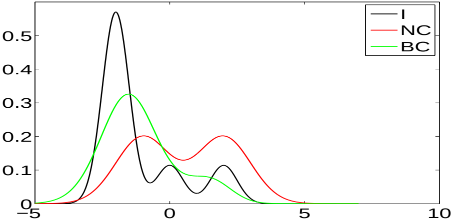
$\mathbf{p} = (1/7, 1/7, 5/7)$



$\mathbf{p} = (3/5, 1/5, 1/5)$



$\mathbf{p} = (5/7, 1/7, 1/7)$



Motivating example - Heavy tails (1)

Data generating process

We assume that the data are generated by the following mixture of the two Student-t distributions, i.e.

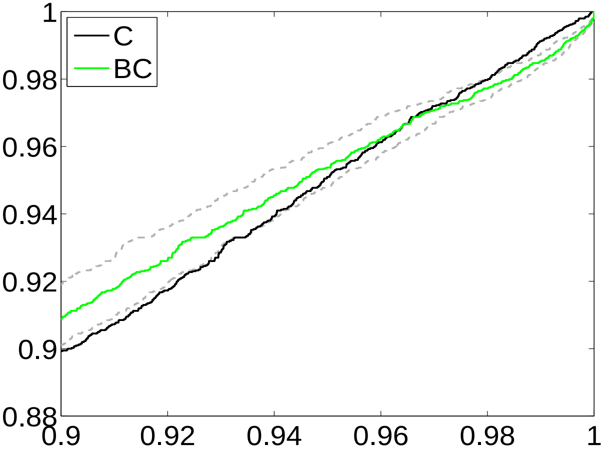
$$y_t \overset{i.i.d.}{\sim} \frac{1}{2}\mathcal{T}(-1,1,6) + \frac{1}{2}\mathcal{T}(2,1,6), \quad t = 1,\ldots,3000,$$

where $\mathcal{T}(\mu,\sigma,\nu)$ denotes a Student-t distribution with location, scale and degrees of freedom parameters μ, σ and ν respectively.

Model set

We assume that the predictive distribution is obtained from the combination of the two normal distributions given in the previous example: $\mathcal{N}(-1,1)$ and $\mathcal{N}(2,1)$. The I and BC models are defined as in the previous example.

Motivating example - Heavy tails (2)



Contribution

Our contribution is:

1. Proposing a Bayesian approach to density calibration and combination.
2. Proposing a Bayesian non–parametric approach to density calibration and combination:
   - Finite mixture of beta densities.
   - Infinite mixtures of beta densities.
3. Developing an efficient algorithm for posterior computation.
4. Providing evidence of better calibration on two well known datasets.

Bayesian non–parametric literature

Modelling issues

- Clustering and heavy tails: JenMah10 (JoE), Gri10 (JoFE), TadKot09 (BA), RodTer08 (BA)
- Clustering changes over time: GriSte11 (JoE) and Tad11 (JASA)
- Clustering changes over space: BasCasLei13 (JoE)

Computational methods

- Posterior approximation: Esc94 (JASA) and EscWes95 (JASA).
- Slice sampling: Wal07 (CoSta), HatNicWal11 (CSDA) and KalGriWal11 (StaCo).
- Retrospective sampling: Pap08 (WP), GriSte11 (JoE).
- Particle learning: Tad11 (JASA).

A general combination and calibration model

We extend the existing calibration model and propose

Beta mixture of densities

$$f_t(y_t) = \sum_{k=1}^{K} w_k f_{\alpha_k,\beta_k}\left(H(y_t|\boldsymbol{\omega}_k)\right) h(y_t|\boldsymbol{\omega}_k), \tag{6}$$

where $\mathbf{w} = (w_1, \ldots, w_K) \in \Delta_{[0,1]^K}$ are the mixture probabilities, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ are the beta calibration parameters and $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K)$ is the set of component-specific linear combination weights $\boldsymbol{\omega}_k = (\omega_{1k}, \ldots, \omega_{Mk}) \in \Delta_{[0,1]^M}$.

Special case

Assume a common linear pooling scheme ($\omega_{ik} = \omega_i$, $k = 1, \ldots, K$, $i = 1, \ldots, M$), and set $\alpha_k = k$ and $\beta_k = K - k + 1$. Then one obtains a special beta mixture density function called Bernstein density (e.g., see PetWas02 (JRSSB))

Consistency result

Any bounded function f on the $[0, 1]$ interval can be approximated by a Bernstein density, that is

$$\lim_{K\to\infty} \left(\sup_{y\in[0,1]} \left| \sum_{k=1}^{K} w_{k,K}(f) f_{\alpha_k,\beta_k}(y) - f(y) \right| \right) = 0,$$

where $w_{k,K}(f) = \int_{((k-1)/K)}^{k/K} f(x)dx$ (e.g., see PetWas02 (JRSSB) for an application to Bernstein prior)

Inference issues

Bernstein densities

- weight estimation and the truncation of the number of components
- parameter restriction, one could expect to have the same accuracy with a smaller number of components, without restrictions (RobRou02)

Beta mixtures

- loss of parameter parsimony
- choice of the number of components

Random Beta mixtures

- a compromise among flexibility and parsimony
- we propose random beta mixtures
- we apply Dirichlet process prior

Bayesian finite beta mixtures (1)

Centered parameterization

Let $\mu = \alpha/(\alpha+\beta)$ and $\phi = \alpha+\beta$ (e.g., see BilCas11 (SNDE) and CasDalLei12(BA)), the density of the beta is

$$f_{\mu,\phi}(z) = B(\mu\phi,(1-\mu)\phi)^{-1} z^{\mu\phi-1}(1-x)^{(1-\mu)\phi-1}\mathbb{I}_{[0,1]}(z), \qquad (7)$$

Easier interpretation of the calibration parameters: μ represents the level of the combination cdf at which the beta calibration is centred.

Bayesian calibration model

$$f_t(y_t|\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k f_{\mu_k,\phi_k}\left(H(y_t|\boldsymbol{\omega}_k)\right) h(y_t|\boldsymbol{\omega}_k), \qquad (8)$$

where $K < \infty$

Bayesian finite beta mixtures (2)

Prior distributions

$$
\begin{aligned}
\mu_k &\sim \mathcal{B}e(\xi_{1\mu}, \xi_{2\mu}),\ k = 1, \ldots, K && (9)\\
\phi_k &\sim \mathcal{G}a(\xi_{1\phi}, \xi_{2\phi}),\ k = 1, \ldots, K && (10)\\
\boldsymbol{\omega} &\sim \mathcal{D}ir(\xi_\omega, \ldots, \xi_\omega) && (11)\\
\mathbf{w} &\sim \mathcal{D}ir(\xi_w, \ldots, \xi_w) && (12)
\end{aligned}
$$

Alternative specification

See RobRou02 (WP) and BouZioMon06 (StaCo) for alternative priors distributions on μ_k and ϕ_k to avoid flat and the bimodal shapes.

Bayesian infinite beta mixtures (1)

A calibration model

$$f_t(y_t|\boldsymbol{\theta}) = f_{\mu,\phi}\left(H(y_t|\boldsymbol{\omega})\right) h(y_t|\boldsymbol{\omega}), \tag{13}$$

where $\boldsymbol{\theta} = (\mu, \phi, \boldsymbol{\omega})$, with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M)$.

Prior distribution

Assume a nonparametric prior for $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta} \sim G(\boldsymbol{\theta})$ where

$$G \sim DP(\psi, G_0) \tag{14}$$

and $DP(\psi, G_0)$ denotes a Dirichlet process (DP) (Ferguson73) with concentration parameter ψ and base measure G_0.

Bayesian infinite beta mixtures (2)

Stick-breaking representation

Following Set94

$$G(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k \delta_{\boldsymbol{\theta}_k}(d\boldsymbol{\theta})$$

with random weights

$$w_k = v_k \prod_{l=1}^{k-1} (1 - v_l) \tag{15}$$

where v_l are i.i.d. $\mathcal{B}e(1, \varphi)$ and the atoms θ_k are i.i.d. from the base measure G_0.

Base measure

$$\mathcal{B}e(\xi_\mu, \xi_\mu)\mathcal{G}a(\xi_\phi/2, \xi_\phi/2)\mathcal{D}ir(\nu_1, \ldots, \nu_M). \tag{16}$$

Bayesian infinite beta mixtures (3)

Infinite mixture representation

$$f_t(y_t|G) = \int f_t(y_t|\boldsymbol{\theta})dG(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (17)$$
$$= \sum_{k=1}^{\infty} w_k f_{\mu_k,\phi_k}\left(H(y_t|\boldsymbol{\omega}_k)\right) h(y_t|\boldsymbol{\omega}_k)$$

Interpretation

- A combination of **local linear pooling** models, with different combination weights and beta-calibration parameters.
- **Local calibration** functions for different parts of the predictive pdf.
- The component-specific model weights indicates the **contribution** of each predictive model to the different part of the predictive support.

Bayesian infinite beta mixtures (4)

Properties

- The number of components has prior distribution (Ant74)

$$P(K = k|\psi, T) = \frac{T!\Gamma(\psi)}{\Gamma(\psi + T)} z_{Tk} \psi^k \tag{18}$$

  with $z_{Tk} = |s_{Tk}|$ where s_{Tk} is the signed Stirling number (AbrSte72, p. 824).
- The dispersion hyper-parameter ψ is driving the prior expected number of parameters.
- The results of the posterior inference are usually presented for different values of ψ.
- It also possible to assume a prior for ψ (EscWes95).

Bayesian inference (1)

Data augmentation

Slice sampling variables u_t, $t = 1, 2, \ldots, T$, i.i.d. standard uniform

$$f_t(y_t, u_t|G) = \sum_{k=1}^{\infty} \mathbb{I}_{\{u_t < w_k\}} f_{\mu_k,\phi_k}\left(H(y_t|\boldsymbol{\omega}_k)\right) h(y_t|\boldsymbol{\omega}_k) \qquad (19)$$

Complete data likelihood

$$L(Y, U|G) = \prod_{t=1}^{T} \sum_{k \in A_t} f_{\mu_k,\phi_k}\left(H(y_t|\boldsymbol{\omega}_k)\right) h(y_t|\boldsymbol{\omega}_k), \qquad (20)$$

where $Y = (y_1, \ldots, y_T)$, $U = (u_1, \ldots, u_T)$, $A_t = \{k|u_t < w_k\}$. Note that $N_t = Card(A_t)$, that is the number of components of the infinite sum, is finite when conditioning on the slice variables (finite mixture representation of the infinite mixture model).

Bayesian inference (2)

Data augmentation

Allocation variables, d_t, $t = 1, \ldots, T$, with $d_t \in A_t$

Complete data likelihood

$$L(Y, U, D|G) = \prod_{t=1}^{T} \mathbb{I}_{\{u_t < w_{d_t}\}} f_{\mu_{d_t}, \phi_{d_t}}(H(y_t|\omega_{d_t}))\, h(y_t|\omega_{d_t}) \quad (21)$$

where $D = (d_1, \ldots, d_T)$.

Bayesian inference (3)

Joint posterior distribution

$$\pi(U, D, V, \Theta, \psi|Y) \propto \prod_{t=1}^{T} \mathbb{I}_{\{u_t < w_{d_t}\}} f_{\mu_{d_t},\phi_{d_t}}\left(H(y_t|\boldsymbol{\omega}_{d_t})\right) h(y_t|\boldsymbol{\omega}_{d_t}) \quad (22)$$

$$\cdot \prod_{k\geq 1} (1-v_k)^{\psi-1} \mu_k^{\xi_\mu -1} (1-\mu_k)^{\xi_\mu -1} \phi_k^{\xi_\phi/2} \exp\{-\xi_\phi \phi_k/2\} \prod_{i=1}^{M} \omega_{ik}^{\nu/2-1}.$$

Joint posterior distribution

Joint exact sampling is not easy. This calls for possibly efficient sampling procedures: collapsed and blocked Gibbs sampling (Wal07 and KalWal11).

Posterior approximation (1)

Definitions

- $\mathcal{D}_k = \{t = 1, \ldots, T | d_t = k\}$: set of indexes of the observations allocated to the k-th component of the mixture.
- $\mathcal{D} = \{k | \mathcal{D}_k \neq \emptyset\}$: set of indexes of the non-empty mixture components.
- $D^* = \sup \mathcal{D}$: number of stick-breaking components used.

Gibbs sampling

Generate sequentially from

1. $\pi(\Theta | U, D, V, Y, \psi)$
2. $\pi(V, U | \Theta, D, Y, \psi)$
3. $\pi(D | \Theta, V, U, Y, \psi)$
4. $\pi(\psi | Y)$.

Posterior approximation (2)

Further remarks

- Sampling all the infinite elements of Θ and V is not needed, since only the elements in the full conditional pdfs of D are needed (KalWal11).
- The maximum number of atoms and stick-breaking components to sample is $N^* = \max\{t = 1, \ldots, T | N_t^*\}$, where N_t^* is the smallest integer such that $\sum_{j=1}^{N_t^*} w_j > 1 - u_t$.

Collapsed Gibbs

Sample from the joint $\pi(V, U | \Theta, D, Y, \psi)$ by:

1. splitting $V = (V^*, V^{**})$, where $V^* = (v_1, \ldots, v_{D^*})$ and $V^{**} = (v_1, \ldots, v_{N^*})$,
2. collapsing the Gibbs by sampling from $\pi(V^* | \Theta, D, Y, \psi)$ and $\pi(U | V^*, \Theta, D, Y, \psi)$ and then from $\pi(V^{**} | V^*, U, \Theta, D, Y, \psi)$.

Posterior approximation - Full conditionals of V^* and U

Full conditional of V^* given (D, Θ, Y, ψ)

The element of V^* have full conditionals

$$\pi(v_k|D, Y) \propto (1 - v_k)^{\psi + b_k - 1} v_k^{a_k} \tag{23}$$

$k \leq D^*$, that are the pdfs of a $\mathcal{B}e(a_k + 1, b_k + \psi)$ with $a_k = \sum_{t=1}^{T} \mathbb{I}_{\{d_t = k\}}$ and $b_k = \sum_{t=1}^{T} \mathbb{I}_{\{d_t > k\}}$.

Full conditional of U given (V, D, Θ, Y, ψ)

It is the uniform

$$\pi(u_t|V, D, Y) \propto \frac{1}{w_{d_t}} \mathbb{I}_{\{u_t < w_{d_t}\}} \tag{24}$$

for $t = 1, \ldots, T$.

Posterior approximation - Full conditional of V^{**}

Full conditional of V^{**} given $(V^*, U, D, \Theta, Y, \psi)$

The element of V^{**} have full conditionals

$$\pi(v_k|U, D, Y) \propto (1 - v_k)^{\psi-1} \tag{25}$$

$k = D^* + 1, \ldots, N^*$, that are the pdf of a $\mathcal{B}e(1, \psi)$.

Posterior approximation - Full conditional of Θ

Full conditional of Θ given (U, D, V, Y, ψ)

Sample from

$$\pi(\boldsymbol{\theta}_k | U, D, V, Y) \propto \prod_{t \in \mathcal{D}_k} f_{\mu_k,\phi_k}(H(y_t|\boldsymbol{\omega}_k))(h(y_t|\boldsymbol{\omega}_k)) \tag{26}$$

$$\cdot \mu_k^{\xi_\mu - 1}(1-\mu_k)^{\xi_\mu - 1} \phi_k^{\xi_\phi/2} \exp\{-\xi_\phi \phi_k/2\} \prod_{i=1}^{M} \omega_{ik}^{\nu/2-1} \mathbb{I}_{\{\boldsymbol{\omega} \in \Delta_{[0,1]^M}\}}$$

for $k \in \mathcal{D}$, and from the prior G_0 for $k \notin \mathcal{D}$.

Sampler for $\boldsymbol{\theta}_k$

We iterate two MH chains with the following target distributions:

1. $\pi(\mu_k, \phi_k | \boldsymbol{\omega}_k, U, D, V, Y, \psi)$
2. $\pi(\boldsymbol{\omega}_k | \mu_k, \phi_k, U, D, V, Y, \psi)$.

Posterior approximation - Full conditionals of D and ψ

Full conditional of D given (V, U, Θ, Y, ψ)

Sample from

$$\pi(d_t|V, U, Y) \propto \mathbb{I}_{\{u_t < w_{d_t}\}} f_{\mu_{d_t}, \phi_{d_t}} \left(H(y_t|\boldsymbol{\omega}_{d_t})\right) h(y_t|\boldsymbol{\omega}_{d_t}), \qquad (27)$$

with $d_t \in \{1, \ldots, N_t^*\}$

Full conditional of ψ given Y

$$\pi(\psi|K, T) \propto B(\psi, T)\psi^{K+c-1} \exp\{-d\psi\}\mathbb{I}_{\psi \in (0,\infty)}, \qquad (28)$$

depends only on the number of observations T and the number of mixture components N^*. (MH step)

Simulated data - Multimodality

Calibration models

1. Two-component beta mixture calibration model (BMC1) with common linear pooling

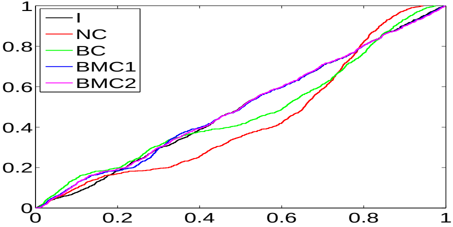
$$f(y|\boldsymbol{\theta}) = \left(w f_{\alpha_1,\beta_1}\left(H(y|\omega)\right) + (1-w) f_{\alpha_2,\beta_2}\left(H(y|\omega)\right)\right) h(y|\omega),$$

2. Two-component beta mixture calibration model (BMC2) with component-specific linear pooling

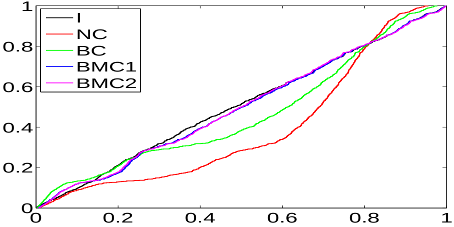
$$f(y|\boldsymbol{\theta}) = w f_{\alpha_1,\beta_1}\left(H(y|\omega_1)\right) h(y|\omega_1) + (1-w) f_{\alpha_2,\beta_2}\left(H(y|\omega_2)\right) h(y|\omega_2),$$

Simulated data - Multimodality (finite mix)

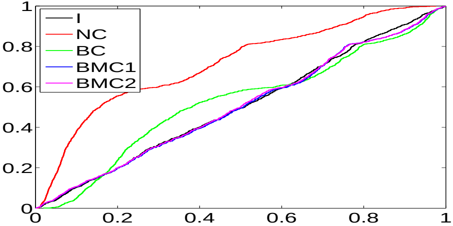
$\mathbf{p} = (1/5, 1/5, 3/5)$



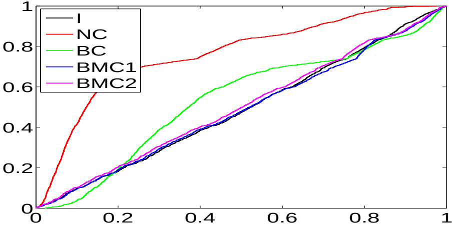
$\mathbf{p} = (1/7, 1/7, 5/7)$



$\mathbf{p} = (3/5, 1/5, 1/5)$

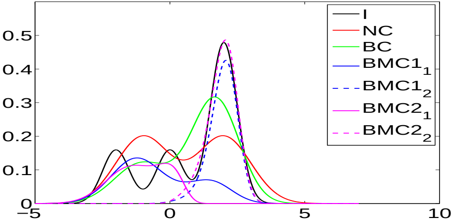


$\mathbf{p} = (5/7, 1/7, 1/7)$

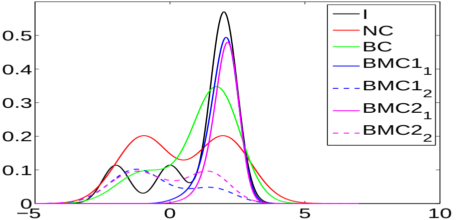


Simulated data - Multimodality (finite mix)

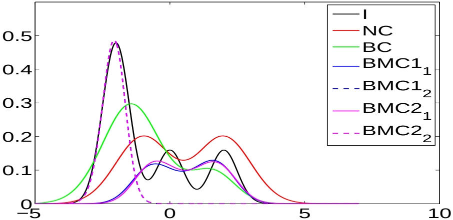
$\mathbf{p} = (1/5, 1/5, 3/5)$



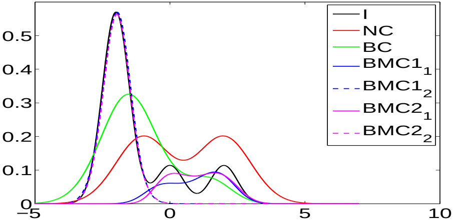
$\mathbf{p} = (1/7, 1/7, 5/7)$



$\mathbf{p} = (3/5, 1/5, 1/5)$



$\mathbf{p} = (1/7, 1/7, 5/7)$

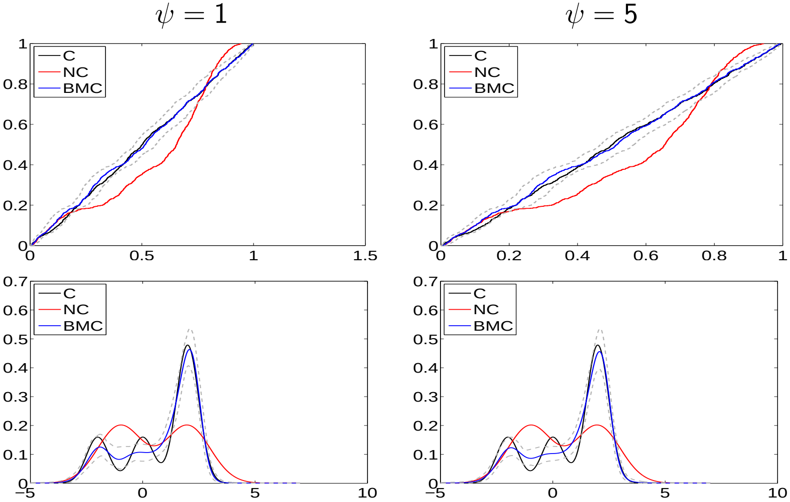


Simulated data - Multimodality (finite mix)

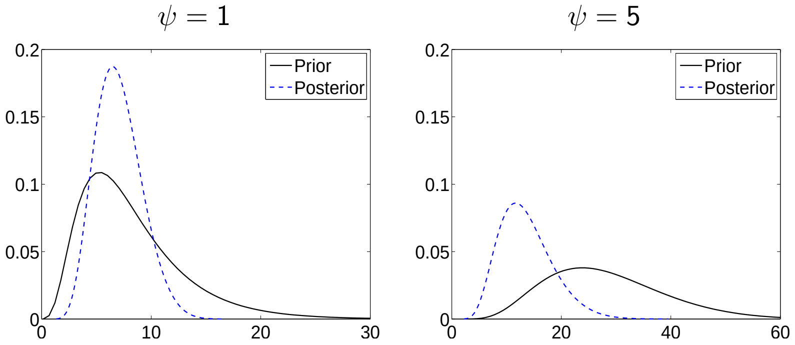
$\mathbf{p} = (1/5, 1/5, 3/5)$							
	w	α_1	β_1	α_2	β_2	ω_1	ω_2
BC	1.00	0.97	1.50			0.20	
BMC1	0.47	0.99	2.04	11.67	4.64	0.45	
BMC2	0.36	0.94	27.48	22.19	4.87	0.04	0.67

$\mathbf{p} = (1/7, 1/7, 5/7)$							
	w	α_1	β_1	α_2	β_2	ω_1	ω_2
BC	1.00	1.04	1.47			0.17	
BMC1	0.35	0.93	1.78	12.18	4.35	0.51	
BMC2	0.44	0.87	2.08	17.71	5.09	0.29	0.54

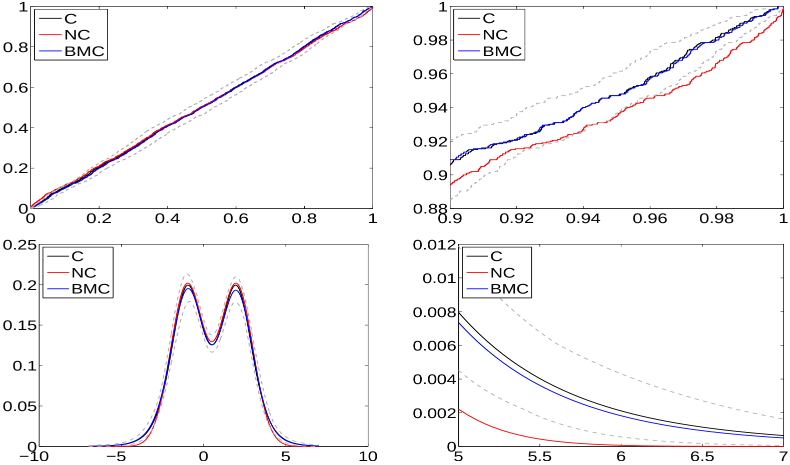
Simulated data - Multimodality (infinite mix)



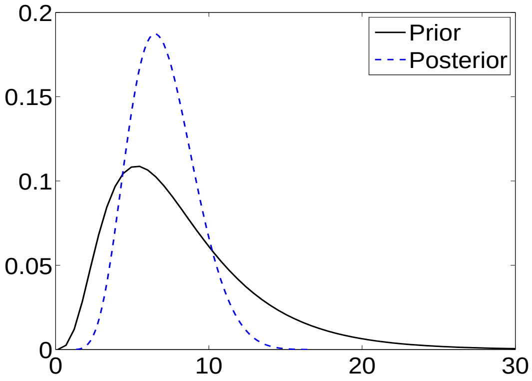
Simulated data - Multimodality (infinite mix)



Simulated data - Heavy tails (infinite mix)



Simulated data - Heavy tails (infinite mix)



Financial data

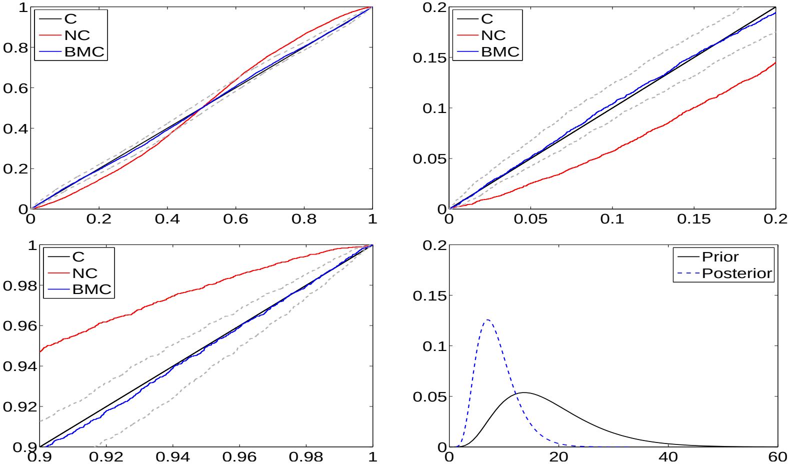
Data

- S&P500 daily percent log returns data from 3 January 1972 to 31 December 2008, (GewAmi11 (JoE), GewAmi10 (IJF), MitKap13 (JoE).
- December 15, 1975 to December 31, 2006: in-sample calibration
- January 3, 2007 to December 31, 2008: out-of-sample analysis, (we extend evidence in GewAmi11 (JoE), GewAmi10 (IJF) by including the Great Financial Crisis.

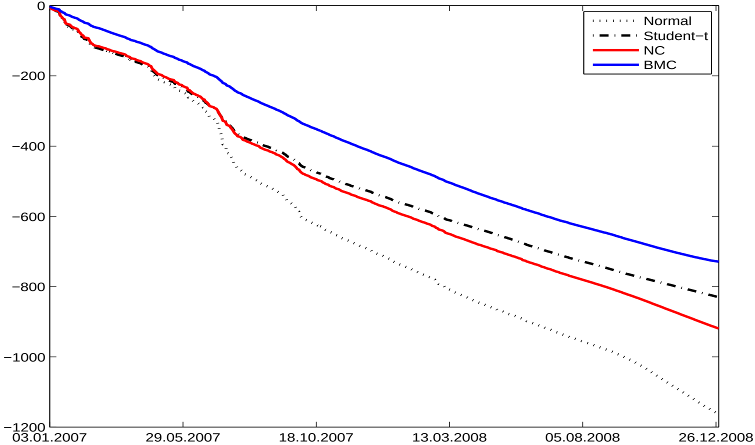
Models

- Models: Gaussian GARCH(1,1) and Student-t GARCH(1,1) (ML estimate sequentially, window size of 1250 trading days, one-day-ahead density forecasts.
- Non-calibrated (NC): linear pooling with recursive log score weights.

Financial data



Financial data



Cumulative log-scores for the Normal GARCH model (Normal), Student-t GARCH model (Student-t), linear pooling (NC) and beta mixture calibration (BMC).

Wind speed data (1)

Data

- 50 ensemble member predictions of wind speed at 10-meter above the ground, obtained from the global ensemble prediction system of the European Centre for Medium-Range Weather Forecasts (ECMWF) (LerTho13 (Tel)).
- Daily maximum of each ensemble member at the Frankfurt location.
- One day ahead forecasts are given by the maximum over lead times. Daily maximum wind speed is the maximum over the 24 hours.

Wind speed data (2)

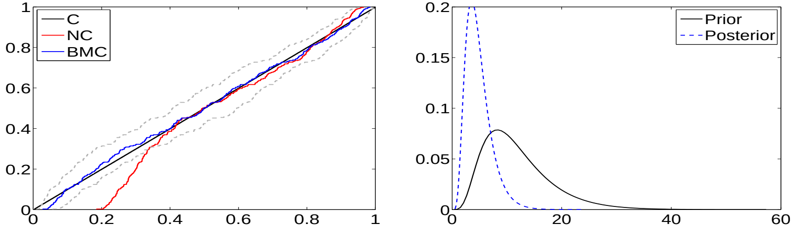
Data

- verification period: August 9, 2010 to April 30, 2011.
- training period: February 1, 2010 to 30 April 2011 (see GneRafWesGol05 (MWR), ThoGne2010 (JRSSA) and ThoJoh2012 (MWR))
- model selection and forecasts: May 1, 2010 to 8 August 2010.

Models

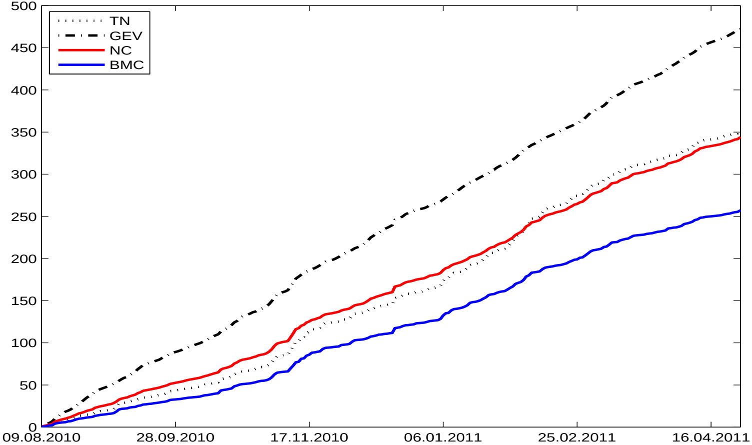
- truncated normal distribution (TN) and the generalized extreme value distribution (GEV).
- TN estimated by CRPS and GEV estimated by ML

Forecast ensemble data



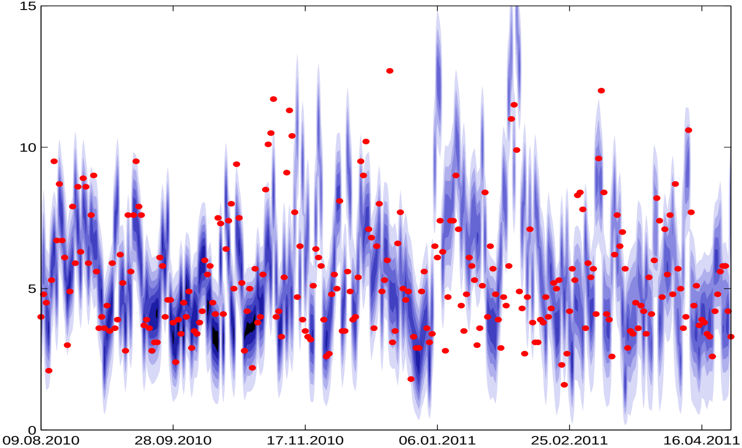
Maximum wind speed at the station of the Frankfurt airport. Left: PITs (left top) of the combination models C, NC, BMC, and BMC 99% HPD (gray). Right: prior and posterior number of mixture components.

Wind speed data



Cumulative CRPS for the truncated normal (TN), the the generalized extreme value distribution (GEV), linear pooling (NC) and beta mixture calibration (BMC)

Wind speed data



Fanchart of the BMC model and observations (red points) over the sample period from August 9, 2010 to 30 April 2011.

Conclusion

- A new Bayesian nonparametric approach to predictive density calibration accounting for parameter uncertainty.
- We build on the predictive density calibration and combination framework of GneBalRaf07 and GneRan13 and propose the use of infinite mixtures of Beta densities for the calibration.
- Each component of the beta mixture is calibrating different parts of the predictive cdf and is using local combination weights.
- On simulated data, stock returns and wind speed data, our Bayesian infinite Beta mixture model provides well calibrated and accurate density forecasts.

Thank you!