Master’s Degree programme – Second Cycle (*D.M. 270/2004*)
in Economics, Models and Methods in Economics and Management

Final Thesis

Bayesian Calibration of Generalised Pools of Predictive Distributions

Supervisor
Ch. Prof. Roberto Casarin

Co-supervisor
Ch. Prof. Andrea Giacomelli

Graduand
Giulia Mantoan
Matriculation Number 832360

Academic Year
2014 / 2015

Master Thesis

Bayesian Calibration of Generalised Pools of Predictive Distributions

Giulia Mantoan

832369@stud.unive.it

Ca' Foscari University of Venice
Department of Economics
Models and Methods in Economics and Management

Supervisor
Roberto Casarin
Co-supervisor
Andrea Giacomelli

October, 21 2015

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Laurea Magistrale
in
Models and Methods in Economics and Management

This thesis is dedicated to my parents Antonello e Agnese and my sister Elisa,
to encourage me to take all the challenges the life offers me with quiet and determination,
to Amedeo, first of my supporter, to share with me the uncertainties of the future,
to Ana and Laura, to have made Venice and Paris my home,
finally to my mentor Marco Miglierina, to be always critical and present.

Questa tesi dedicata ai miei genitori Antonello e Agnese e a mia sorella Elisa, per incoraggiarmi a cogliere le sfide che la vita mi offre con calma e determinazione,
ad Amedeo, primo dei miei sostenitori, per condividere con me le incertezze del futuro,
ad Ana e Laura per aver reso Venezia e Parigi la mia casa,
infine al mio mentor Marco Miglierina, per essere sempre critico e presente.

I wish to thank the committee members who were more than generous with their expertise and precious time. A special thanks to Prof. Roberto Casarin, my thesis supervisors, for his countless hours of reflecting, reading, discussion and most of all patience throughout the entire process of the creation of this essay.

I would like to acknowledge and thank prof. Francesco Ravazzolo, for his attention, support and discussion.

Contents

List of Tables

List of Figures

Abstract

Decision-makers often consult different experts to build a reliable forecast on some uncertain variable of interest. Combining more opinions and calibrating them to maximise the forecast accuracy is consequently a crucial issue in several economic problems. A Bayesian approach was applied to derive a combined and calibrated density function using random calibration functionals and random combination weights. In particular, it compares the application of linear, harmonic and logarithmic pooling in the Bayesian combination approach. The three combination schemes, i.e. linear, harmonic and logarithmic, are studied in simulation examples with multimodal densities. The simulation examples show that in a beta-mixture calibration framework the three combination schemes are substantially equivalent.

Chapter 1

Introduction

Decision-makers often consult experts for reliable forecast about some uncertain future outcome. Expert opinion has been used in a more or less systematic way in many fields: weather forecast, aerospace program, military intelligence, nuclear energy and in policy analysis. In economic field, experts' forecasts are often combined to produce estimates on the basis of past performance and the observed values of some exogenous variables. The forecast can be expressed in terms of future realisation, and in this case it is referred to Point Forecast, or in terms of probabilities of the future values (full distribution) of the variable, defined as Probabilistic Forecast.

Combining different experts' forecasts or predictive cumulative distribution functions is a critical issue in order to construct a single consensus forecast representing the experts' advice. Among the first papers on forecasting with more predictions, Barnard (1963), who consider air passenger data, and Roberts (1965), who introduced a distribution that is essentially a weighted average of the posterior distributions of two models which is similar to the result of a Bayesian Model Averaging (BMA) procedure. See Hoeting et al. (1999) for a review on BMA with an historical perspective. Nowadays, the literature on combination of point forecasts has reached a relatively mature state dating back to papers such as Bates and Granger (1969). Timmermann (2006) provides an extensive summary of the literature and the success of the forecast combinations in the economic field. However, the literature on density forecasting and on density combinations from different experts has emerged only recently, see Corradi and Swanson (2006), Mitchell and Hall (2005), Hall and Mitchell (2007) and Wallis (2005) for a survey. There are two elementary choices in combining predictive densities from many experts. One is the method of aggregation or functional form of combining. The other is the construction of the weights attached to the individual density forecasts. Possibile methods of aggregation are described in an early review of Genest and Zidek (1986). The linear pooling, proposed by Stone (1961) has been used almost exclusively in empirical applications on density forecast gains, see Ranjan and Gneiting (2010) and Geweke and Amisano (2011). Starting from these pooling schemes, the traditional pools are generalised by Billio et al. (2013) and Fawcett et al. (2013).

Moreover to evaluate the accuracy of the final experts' advice, the experts must be calibrated. The calibration is a measurement process to evaluate how good is the expert assessment: an expert is well-calibrated if the subjective probability mass function (on density function) agree with the sample distribution of the realisations of the unknown variable in the long run.
Bassetti et al. (2015) introduce a Bayesian approach to predictive density calibration and combination through the use of random calibration functionals and random weights. Extending Ranjan and Gneiting (2010) and Gneiting and Ranjan (2011), they propose both finite beta and infinite beta mixture for the calibration. For combination they apply a local linear pool.

Some of the most used Survey of Professional Forecasters (SPF) both point forecast both probability forecast are: the ASA-NBER with surveys on inflation, components of consumption, government spending, the FED Philadelphia and the ECB mainly on GDP, unemployment rate and inflation for different areas of interest, and other national institutions such as the Reserve Bank of India, the Monetary Authority of Singapore and the National Bank of Poland, just to mention a few.

In this work, a beta mixture approach is proposed to combine and calibrate prediction functions and compare linear, harmonic and logarithmic pooling in the application of the Bayesian approach. Relative to Bassetti et al. (2015), the number of beta components is fixed to the family of generalized linear combination schemes (i.e. harmonic and logarithmic) proposed in DeGroot et al. (1995). The Bayesian nonparametric algorithm in Bassetti et al. (2015) is extended to the new combination methods. The effects of the three schemes are studied in simulation examples with multimodal densities.

The results show that three combination schemes are substantially equivalent in a beta-mixture calibration framework. The remainder of the paper is organised as follow. Section 2 introduces linear, harmonic and logarithmic combination models and the notion of calibration. Section 3 discusses Bayesian inference for the calibrated combination models. In Section 4 the results of the simulation exercises are shown.

Chapter 2

Combination and Calibration

2.1 Subjective probability distribution

The word “opinion” used in the introduction, will refer to a collection of statements (in this case numerical) expressing the expert’s degree of belief about the world. Most of the solutions to aggregation problems proposed that each expert encoded his opinion as a Subjective Probability Distribution. If Ω denoted the collection of mutually exclusive statements about the world, a probability measure P, assigns a number $0 \leq P \leq 1$ to each subset E of Ω, according to the degree to which this is believed to contain the fixed, but unobserved realisation $\omega \in \Omega$. Generally, P is constructed in such a way that:

$$P\left(\bigcup_{i=1}^{N} E_i\right) = \sum_{i=i}^{N} P(E_i) \tag{2.1}$$

where the E_i’s are mutually disjoint subset of Ω. The condition 2.1 is referred to as *countable σ-additivity*. De Finetti (1970) advocated the use of *finite σ-additivity* (i.e., $N < \infty$) instead of *countable σ-additivity* (i.e., $N = +\infty$) property of probability measure P; in this work the framework is the σ-additivity, for further generalisation please refer to Casarin et al. (2015b).

Subjective probability distributions may be expresses as probability measures, densities or mass functions. In the framework of probabilistic forecasting treated here, and for a real-valued outcome, a probabilistic forecast can be represented in the form of a predictive cumulative distribution function (hereafter predictive cdf), which might be discrete, mixed discrete-continuous or continuous with a predictive density function (hereafter predictive pdf).

2.2 Combination Model

Probability distribution is expression of the expert's subjective belief, which is based on a prior experience the individual has had with the problem at hand. Thus, their current opinions may differ because they do not collect same information and they do not interpret data in the same way. In this case a method to combine the different sources of information is needed. Suppose to have K σ-algebra $\mathcal{A}_1, \cdots, \mathcal{A}_K$ representing different information sets, and a sequence of predictive cumulative distribution functions $F_1, \cdots, F_K$, for the variable of interest Y, and that a summary of F_k's is required. Probability distribution, as mentioned above, is expression of that expert's subjective beliefs, which is based on a prior experience the individual has had with the problem at hand. Thus, their current opinions may differ because they do not collect same information and they do not interpret data in the same way.

An ideal strategy to combine predictive cumulative distribution functions may be combine information sets to issue the conditional distribution of the observation Y given the σ-algebra$(\mathcal{A}_1, \cdots, \mathcal{A}_K)$ generated by the information sets $(\mathcal{A}_1, \cdots \mathcal{A}_K)$. However information sets are not known in practice. The solution is to model the conditional distribution of the observation Y given the σ-algebra generated by the predictive cumulative distribution functions. This result will be appreciable later talking about calibration methods.

Following the notation Gneiting and Ranjan (2013), it's defined a parametric family of combination formulas, with parameter $\boldsymbol{\theta}$, as a mapping:

$$H(\cdot|\boldsymbol{\theta}) : \mathcal{F}^K \to \mathcal{F}, (F_1, \cdots, F_K) \mapsto H((F_1, \cdots, F_K)|\boldsymbol{\theta})$$

with $\boldsymbol{\theta} \in \Theta$ a parameter, $\mathcal{F}_k$, $k = 1, \cdots, K$, a sequence of cumulative distribution function, where $\mathcal{F}$ is a suitable space of distribution and Θ is a parameter space.

Here, adopting approach of DeGroot et al. (1995), it were employed three types of pooling schemes, denoted with $H_m(y|\omega)m = 1, 2, 3$, special cases of the generalised linear form:

$$g(H_m(y|\boldsymbol{\omega})) = \sum_{k=1}^{K} \omega_k g(F_k(y))$$

where g is a continuous monotonic function and $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_k)'$ is a vector of combination weights, with $\sum_{k=1}^{K}(\omega_1, \cdots, \omega_K)$ and $\omega_k \geq 0$. If g is differentiable with inverse g^{-1} and the cdfs $F_k, k = 1, \cdots, K$, admits pdf

$f_k, k = 1, \cdots, K$ then the generalised combination model con be re-written in terms of probability density function h_m as

$$h_m = \frac{1}{g'(H_m(y|\boldsymbol{\omega}))} \sum_{k=1}^{K} \omega_k g'(F_k(y)) g(f_k(y)) \tag{2.2}$$

where g' denotes the first derivate of g. The three cases considered in this paper are:

1. Linear opinion pool (m=1) i.e. g(x)=x

$$H_1(y|\boldsymbol{\omega}_k) = \sum_{k=1}^{K} \omega_k F_k(y) \tag{2.3}$$

2. Harmonic opinion pool (m=2) i.e. g(x)=1/x

$$H_2(y|\boldsymbol{\omega}_k) = \left(\sum_{k=1}^{K} \omega_k F_k(y)^{-1} \right)^{-1} \tag{2.4}$$

3. Logarithmic opinion pool (m=3) i.e g(x)=log(x)

$$H_3(y|\boldsymbol{\omega}_k) = \prod_{k=1}^{K} F_k(y)^{\omega_k} \tag{2.5}$$

with densities functions:

1. Linear opinion pool (m=1) i.e. g(x)=x

$$h_1(y|\boldsymbol{\omega}_k) = \sum_{k=1}^{K} \omega_k f_k(y) \tag{2.6}$$

2. Harmonic opinion pool (m=2) i.e. g(x)=1/x

$$h_2(y|\boldsymbol{\omega}_k) = H_2(y|\boldsymbol{\omega})^2 \sum_{k=1}^{K} \omega_k F_k(y)^{-2} f_k(y) \tag{2.7}$$

3. Logarithmic opinion pool (m=3) i.e g(x)=log(x)

$$h_3(y|\boldsymbol{\omega}_k) = H_3(y, \boldsymbol{\omega}) \prod_{k=1}^{K} \omega_k F_k(y)^{-1} f_k(y) \tag{2.8}$$

$$\tag{2.9}$$

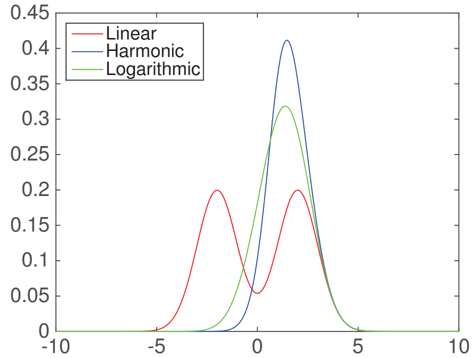


Figure 2.1: Density Function for linear (H_1, red line), harmonic (H_2, blue line), and logarithmic (H_3, green line) combination models. Combination weight $\omega = 0.5$

where $f_k(y)$ corresponds to the probability density function of $F_k(Y)$.
To conclude, an example of predictive pdfs is presented in figure 2.1, in order to appreciate the difference among the three types of combination schemes.

The predictive functions to be combined are:

$$F_1 \sim \mathcal{N}(2,1), \qquad\qquad F_2 \sim \mathcal{N}(-2,1)$$

where $\mathcal{N}(\mu, \sigma)$ is the normal distribution with location μ and scale σ.

At first look, linear combination model is able to generate multimodal pdfs, whereas harmonic and logarithmic models generate unimodal pdf with certain degrees of skewness depending on the value of the combination weights (see Fig. 2.2).

Most of literature on the issue, characterises different types of combination formulas wherever they satisfy or not some particular conditions. The strong and weak setwise properties of McConway (1981), the zero preservation property by Bacharach (1975) or the independence preservation property by Laddaga (1977). Such weights mean models have found many ad hoc applications but they raise serious problems with respect to accountability, neutrality and empirical control. For these reasons, it is preferred to use the perspective of Hora (2010) and taking into account calibration and dispersion.

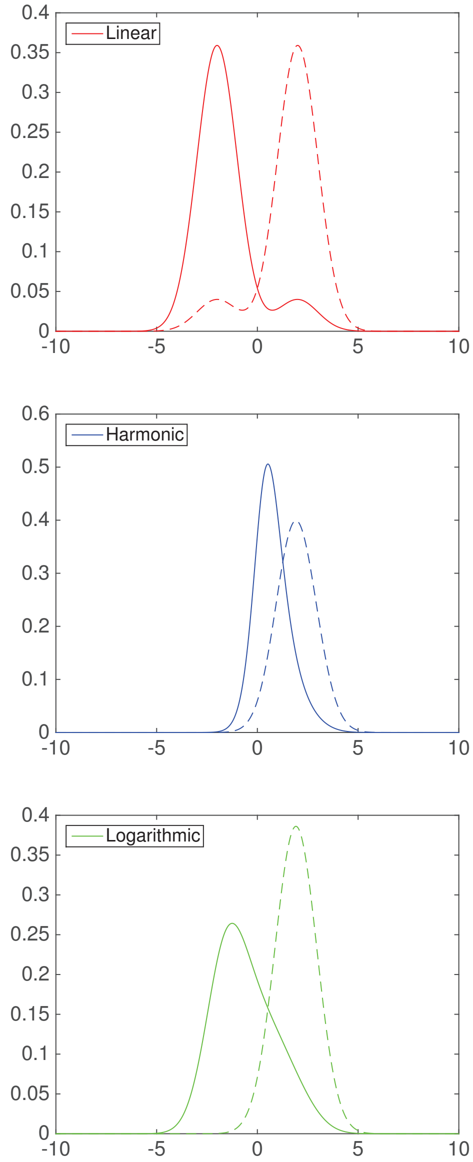


Figure 2.2: Combination densities for the three schemes (different rows) for $\omega = 0.9$ (solid lines) and $\omega = 0.1$ (dashed lines).

2.3 Calibration Model

The calibration issue states what means for F_k to be a "good" predictive distribution function for the empirical data Y.

Dawid (1984) introduced the criterion of *complete calibration* for comparing prequential probabilities $F_k = P(Y_k = 1|Y_1, \cdots, Y_k - 1)$ with the binary outcomes Y_k. This criterion requires that the averages of the F_k and of the Y_k converges to the same limit. The validity of this criterion is justified by the fact that the above property holds with probability one, so that its failure discredits F_k. In the case of continuous quantities Y_k, Dawid (1982) apply the Rosenblatt (1952) concept of Probability Integral Transform (PIT); which is a random variable described by:

$$Z_f = F_k(Y_k)$$

where F_k is the predictive cumulative distribution function. In the case of continuous Y_k (and the continuous F), $Y_k \sim F_k$ then Z_k is standard uniform as shown in the following: $P(Z_k \leq z) = P(F_k(Y_k) \leq z) = P(Y_k \leq F_k^{-1}(z) = F_k(F_k^{-1}(z))) = z$,which is the cumulative distribution function of a standard uniform.

In summary, the PIT is the value that the predictive cumulative distribution function attains at the observation; the PIT takes values in the unit interval, and so the possible values of its variance are constrained to the closed interval $[a, \frac{1}{4}]$. A variance of $\frac{1}{12}$ corresponds to a uniform distribution.

Gneiting and Ranjan (2013) generalised the complete calibration criterion used by Dawid (1984) applying it on non-binary outcomes Y. As mentioned at the beginning of this section, a useful tool for combining predictive distribution function is the conditional distribution of the observation Y given the σ-algebra generated by the predictive cumulative distribution functions, $F_1, \cdots, F_K$, or by the combination formula:

$$G(y|F_1, \cdots, F_K) = \psi(H(F_1(y), \cdots, F_K(y)))|\boldsymbol{\theta} \tag{2.10}$$

almost surely.
where $\psi(x)$ is a map from $\mathcal{F}$ to $\mathcal{F}$. This is a modified version of the auto-calibration property given in Tsyplakov (2011) for $\psi(x) = x$. Here we assume that the calibration is obtain by a distortion, through ψ, of a combination scheme H. Following the combination schemes given in the previous section, the relationship between calibration and combination is given by the following composition of functions

$$G_m(y|\boldsymbol{\omega}) = (\psi \circ \varphi^{-1}) \left(\sum_{k=1}^{K} \omega_k \varphi(F_k(y)) \right) \tag{2.11}$$

where $(\psi o \varphi^{-1})(x) = \psi(g^{-1}(x))$.
In particular this must satisfy some requirements on the PITs dispersion: the aggregation method is flexibly dispersive or exchangeably dispersive in the sense analysed by Gneiting and Ranjan (2013):

1. The combination formula is flexibly dispersive if for the class $\mathcal{F}$ of fixed, non-random cdf, for all $F_0 \in \mathcal{F}$ and $F_1, \cdots, F_k \in \mathcal{F}$, $\mathcal{L}(Y) = F_0$ then $H(F_{-1}, \cdots, F_k)$ is a neutrally dispersed forecast (i.e. $varZ_H = 1/2$)

2. The combination formula is exchangeably flexible dispersive if for the class $\mathcal{F}$ of fixed, non-random cdf, for all $F_0 \in \mathcal{F}$ and $F_1, \cdots, F_k \in \mathcal{F}$, $\mathcal{L}(Y) = F_0$ then $H(F_{-1}, \cdots, F_k)$ is anonymous
(i.e. $H(F_{\pi(1)}, \cdots, F_{\pi(K)}) = H(F_1, \cdots, F_K)$) and neutrally dispersed forecast (i.e. $varZ_H = 1/2$)

In a nutshell, aggregation methods has to be sufficiently flexible to accommodate situations typically encountered in practice. In the next part, a possible solution to the problem of choosing the combination and calibration scheme will be described.

2.3.1 Beta mixture calibration and combination model

Introduced by Ranjan and Gneiting (2010) and generalised in Gneiting and Ranjan (2013), the beta transformation of the pooling operator H takes the form:

$$G_m(y|\boldsymbol{\theta}) = B_{\alpha,\beta}(H_m(y|\boldsymbol{\omega})) \tag{2.12}$$

where $B_{\alpha,\beta}$ denotes the cumulative distribution function of the beta distribution with parameters $\alpha > 0$ and $\beta > 0$ and $H_m(y|\boldsymbol{\omega})$ is one of the combination formulae defined by 2.3, 2.4, 2.5. Moreover consider that $\alpha > 1$ and $\beta > 1$ reduces the beta-transformed pool in the beginning pooling operator. If $F_1, \cdots, F_K$ admits Lebesgue densities, the previous can be written in terms of aggregated probability density functions (pdf):

$$g(y|\boldsymbol{\theta}) = (h_m(y|\boldsymbol{\omega}))b_{\alpha,\beta}(H_m(y|\boldsymbol{\omega})) \tag{2.13}$$

where h_m is defined by equations 2.6, 2.7, 2.8 and $b_{\alpha,\beta}$ is the pdf of the beta distribution. Bassetti et al. (2015) interprets the beta transformation as a parametric calibration function which acts on the combination of $F_1, \cdots, F_K$ with weights ω_k, $k = 1, \cdots, K$.

Furthermore equations 2.12 and 2.13 are generalised proposing the use of a mixture of beta calibration and combination models:

$$G_m(y|\boldsymbol{\theta}) = \sum_{j=1}^{J} \rho_j B_{\alpha_j,\beta_j}(H_m(y|\boldsymbol{\omega}_j)) \tag{2.14}$$

and

$$g_m(y|\boldsymbol{\theta}) = \sum_{j=1}^{J} \rho_j (h_m(y|\boldsymbol{\omega}_j)) b_{\alpha_j,\beta_j}(H_m(y|\boldsymbol{\omega}_j)) \tag{2.15}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\rho})$, comprises $\boldsymbol{\alpha} = \alpha_1, \cdots, \alpha_J$ and $\boldsymbol{\beta} = \beta_1, \cdots, \beta_J$, the beta calibration parameters, $\boldsymbol{\omega} = \omega_{1j}, \cdots, \omega_{Kj}$ the vector of combination weighs and $\rho = (\rho_1, \cdots, \rho_J)$ the vector of the beta mixture weights.

In conclusion, a simulation example is reported to illustrate the effect of beta combination and calibration model on predicting realisations of the variable of interest x. Consider:

$$x_i \sim \mathcal{N}(0, 1)$$

for $i = [1, 1000]$ and two predictive cdfs:

$$F_1 \mapsto \mathcal{N}(0.5, 1)$$
$$F_2 \mapsto \mathcal{N}(0, 3)$$

the first one errs in predicting the mean of the distribution, the second one errs in predicting the variance. Here we not pay attention to the combination formula that generates the two predictive functions. In the figure 2.3, which represent cdfs of PITs, the difference among the two errors types is evident: errors in mean are displayed by a cdf that overestimate (or underestimate, depending on error sign) the “true” cumulative density function; while errors in variance appear as an underestimate in the left side of the distribution, and an overestimate in the right side, the discontinuity point corresponds at the mean, in which the two line intersect.

Now let apply the beta transformation to each predictive function separately in order to appreciate the effect of the procedure. In figure 2.4 are reported the initial predictive function, their beta transformation, the simulated “true realisations”.

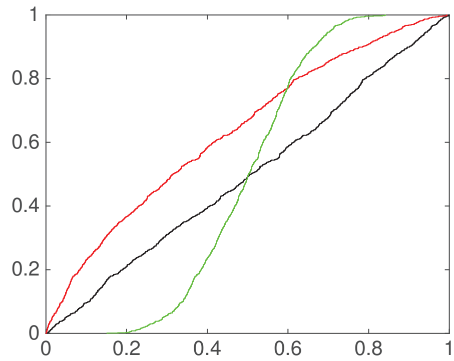


Figure 2.3: PITs cumulative density functions generated by (H_1, red line) and (H_2, green line) and CDF of simulated realisation of variable of interest x.

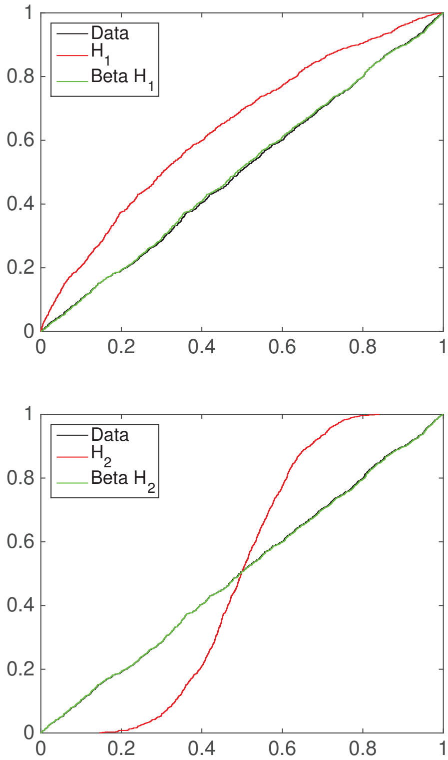


Figure 2.4: First row: beta calibration of F_1 with estimated $\alpha = 0.773$ and $\beta = 1.352$. Second row: beta calibration of F_2 with estimated $\alpha = 7.485$ and $\beta = 7.477$.

Chapter 3

Bayesian Inference

Before proceeding in the presentation of the Bayesian inference setting, a different parametrisation of the problem has been proved being more convenient in various papers involving beta distributions (see, e.g, Robert and Rousseau (2002), Bouguila et al. (2006), Casarin et al. (2012), Billio and Casarin (2011) and Casarin et al. (2015b)). The technique consists in thinking the standard Beta distribution of x with ($\alpha > 0$ and $\beta > 0$) in terms of gamma distribution. Then, the reparameterised beta density function:

$$b^*_{\mu,\nu}(x) = \frac{\Gamma(\nu)}{\Gamma(\mu\nu)\Gamma((1-\mu)\nu)} x^{\mu\nu-1}(1-x)^{(1-\mu)\nu-1}\mathbb{I}_{[0,1]}(x),$$

where $\mu = \alpha/(\alpha+\beta)$, $\nu = \alpha+\beta$ and $\Gamma(\cdot)$ denoted the gamma distribution.

The aim of this section is to provide an estimation procedure for a combined and calibrated model, in which the cumulative prediction functions F_{kt} are aggregated in a single cumulative distribution function (F_t) for the subsequent realisation y_{T+1}. To handle this issue, consider a unit prediction horizon, where the training set is composed by predictive cdfs $F_{1t}, \cdots, F_{Kt}$ based on the information available in $t-1$ along with the respective realisations y_t.

Let consider the following reparameterised cdf and pdf function of beta mixture combination and calibration model in equations 2.14 and 2.15:

$$G_{mt}(y_t) = \sum_{j=1}^{J} \rho_j B^*_{\mu_k,\nu_k}(H_m(y_t|\boldsymbol{\omega}_k))$$

$$g_{mt}(y_t) = \sum_{j=1}^{J} \rho_j h_m(y_t|\boldsymbol{\omega}_k) b^*_{\mu_k,\nu_k}(H_m(y_t|\omega_k))$$

for $t = 1, \cdots, T$, $m = 1, \cdots, 3$ represent the type of combination employed, $k = 1, \cdots, K$ the predictive distributions and $j = 1, \cdots, J$ the number of beta mixtures. Moreover, the parameter $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_K) \in (0, 1)$, $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_K) \in (0, \infty)$ $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_J) \in \Delta_K^J$ and $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_K) \in \Delta_J$, are collected in a single parameter vector $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\omega}, \boldsymbol{\rho})$.

Thus, the bayesian approach in Bassetti et al. (2015) assumes:

$$\begin{aligned}
\boldsymbol{\mu}_k &\sim \mathcal{B}e(\xi_{\mu 1}, \xi_{\mu 2}) \\
\boldsymbol{\nu}_k &\sim \mathcal{G}a(\xi_{\nu 1}, \xi_{\nu 2}), \\
\boldsymbol{\omega}_k &\sim \mathcal{D}ir(\xi_{\omega 1}, \cdots, \xi_{\omega_J}), \\
\boldsymbol{\rho} &\sim \mathcal{D}ir(\xi_{\rho 1}, \cdots, \xi_{\rho_J}),
\end{aligned}$$

for $k = 1, \cdots, K$ where $\mathcal{B}e(\alpha, \beta)$ is a Beta Distribution with density proportional to $x^{\alpha-1}(1-x)^{\beta-1}$, $\mathcal{G}a(\gamma, \delta)$ is a Gamma distribution with density proportional to $x^{\gamma} exp\{-\delta x\}$ for $x > 0$, and $\mathcal{D}ir(\epsilon_1, \cdots, \epsilon_J)$ is a Dirichlet distribution with density proportional to $\prod_{j=1}^{J} x_j^{\epsilon_j - 1}$.

The complete data likelihood is:

$$L(Y, D|\boldsymbol{\theta}) = \prod_{t=1}^{T} \prod_{j=1}^{J} \rho_k h_m(y_y|\boldsymbol{\omega}_j) b^*_{\mu_k,\nu_k}(H_m(y_t|\boldsymbol{\omega}_j))^{d_{jt}}$$

where d_{jt} was introduced as allocation variable following the data augmentation issue of Frühwirth-Schnatter (2006), $Y = (y_1, \cdots, y_T)$ and
$D = (d_{11}, \cdots, d_{J1}, \cdots, d_1T, \cdots, d_JT)$. The implied joint distribution of D and $\boldsymbol{\theta}$ given the observation y:

$$\pi(D, \boldsymbol{\theta}|Y) \propto g(\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\omega}) \prod_{j=1}^{J} \rho_j^{\xi_\rho + T_j - 1} \prod_{t \in \mathcal{D}_j} h_m(y_t|\boldsymbol{\omega}_k) b^*_{\mu_j,\nu_j}(H_m(y_t|\boldsymbol{\omega}_j)),$$

where $g(\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\omega})$ corresponds to the prior density of the parameters, and $\mathcal{D}_j = t = 1, \cdots, T|d_{jt} = 1$.

To sample from the joint distribution, a Gibbs sampler is created to draws iteratively from $\pi(D|\boldsymbol{\theta}, Y)$, $\pi(\boldsymbol{\mu}, \boldsymbol{\nu}|\boldsymbol{\rho}, \boldsymbol{\omega}, D, Y)$, $\pi(\boldsymbol{\omega}|\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\rho}, D, Y)$, $\pi(\boldsymbol{\rho}|\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\omega}, D, Y)$. The output of the algorithm is $\boldsymbol{\theta}_i = \boldsymbol{\rho}_i, \boldsymbol{\mu}_i, \boldsymbol{\nu}_i, \boldsymbol{\omega}_i$ for $i = 1, \cdots, I$, where I is the number of iteration in the Gibbs sampler, which can be used to approximate the cdf posterior predictive distribution at time $T + 1$, $G_{mT+1}(Y_{T+1}$ as follow,

$$\hat{G}_{mT+1}(Y_{T+1} = \frac{1}{I - I_0} \sum_{i=I_0+1}^{I} \sum_{j=1}^{J} \rho_{ji} B^*_{\mu_{ki},\nu_{ki}} \left(\varphi^{-1} \left(\sum_{k=1}^{K} \omega_{jki} \varphi(F_{T+1}(y_{T+1})) \right) \right).$$

where I_0 is a set of burn-in MCMC samples. In the next chapter, we consider the two-component beta mixtures, i.e. the case of $J = 2$. The Gibbs sampler was replace by a global MH sampler, with target distribution:

$$\begin{aligned}\pi(\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\omega}, \boldsymbol{\rho}|Y) \propto \prod_{t=1}^{T} &\rho h_m(y_t|\boldsymbol{\omega}_1) b^*_{\mu_1,\nu_1}(H_m(y_t|\boldsymbol{\omega}_1)) \\ &+ (1-\rho) h_m(y_t|\boldsymbol{\omega}_2) b^*_{\mu_2,\nu_2}(H_m(y_t|\boldsymbol{\omega}_2)) \mu^{\xi_\mu - 1}(1-\mu)^{\xi_\mu - 1} \\ &\cdot \nu^{\xi_\nu - 1} exp\{-\xi_\nu \nu\} \omega^{\xi_\omega - 1}(1-\omega)^{\xi_\omega - 1} \rho^{\xi_\rho - 1}(1-\rho)^{\xi_\rho - 1}. \quad (3.1)\end{aligned}$$

Chapter 4

Empirical Analysis

4.1 Simulation study

In this simulation study we focus on multimodal true distributions. A random variable $Y_i,\ i = 1, \cdots, I$ is simulate from a mixture of three normal distributions. We denote with $F(y|\mu, \sigma)$ the cdf of the distribution $\mathcal{N}(\mu, \sigma)$ and the data generating process (DPG) is specified as:

$$Y_i \overset{i.i.d.}{\sim} p_1\mathcal{N}(-2, 2) + p_2\mathcal{N}(0, 2) + p_3\mathcal{N}(2, 2), \tag{4.1}$$

where $\boldsymbol{p} = (p_1, p_2, p_3) \in \Delta_3$ and Δ_k denotes the k-dimensional standard simplex.

Moreover, we assume that the set of predictive models includes the following two normal distribution $\mathcal{N}(-1, 1)$ and $\mathcal{N}(0.5, 3)$. The distributions of the combination schemes compared in our simulation experiment are:

1. The Equally Weighted Model (EW)

$$H_1(y, \omega) = \omega F(y| - 1, 1) + (1 - \omega)F(y|0.5, 3),$$

$$H_2(y, \omega) = \frac{\omega}{F(y| - 1, 1)} + \frac{(1 - \omega)}{F(y|0.5, 3)},$$

$$H_3(y, \omega) = exp\{\omega log(F(y| - 1, 1)) + (1 - \omega)log(F(y|0.5, 3))\},$$

where ω is the combination weight equal to 1/2. H_1, H_2, H_3 Corresponds to equations 2.3, 2.4, 2.5 for linear, harmonic and logarithmic pool, respectively, when $k = 2$.

2. The beta-transformed model (BC1)

$$G_m(y|\theta) = B_{\alpha,\beta}(H_m(y|\boldsymbol{\omega}))$$

where $\theta = (\alpha, \beta, \omega)$, $H_m(y|\boldsymbol{\omega})$ is defined by equations 2.3, 2.4, 2.5 and $h_m(Y, \omega_k)$ is defined by equation 2.6, 2.7 and 2.8

3. The two-component beta mixture model (BC2)

$$G_m(y|\theta) = \rho B_{\alpha_1,\beta_1}(H_m(y|\boldsymbol{\omega}_1)) + (1-\rho) B_{\alpha_2,\beta_2}(H_m(y|\boldsymbol{\omega}_2)),$$

where $\theta = (\rho, \alpha_1, \alpha_2, \beta_1, \beta_2, \boldsymbol{\omega}_1', \boldsymbol{\omega}_2')$ and $H_m(Y, \boldsymbol{\omega})$ are the same as in the BC1 model.

The posterior approximation is based on a set of 100,000 MCMC iterations after a burn-in period of 50,000 iterations. An example of MCMC output is given in Appendix.
The posterior means of BC1 and BC2 parameters (represented by the vector $\boldsymbol{\theta}$), of i.i.d. 1,000 observations each, are reported in Table 4.1 for the linear combination models, in Table 4.2 for the harmonic combination models, and in Table 4.3 for the logarithmic combination models, according to p_i. In these tables, α and β stands for the parameters of the beta distribution in the BC1 model and in the first component of the BC2 model, while the second component of BC2 is referred to α^* and β^*.

Generally, BC2 models build more flexible predictive cdf: in most of the cases presented, the BC1 models do not take into account the first predictive distribution function (F_1), while BC2 weights more the first one than the second predictive cdf, with few exceptions. Comparing pooling schemes, no clear tendency appear form the tables.

Figure 4.1 displays a comparison through PITs of linear, harmonic and logarithmic pools when those are combined with the equally weighted model and the 45 degree line, which represent the PITs plot for the unknown ideal model: as it is possible to see, the EW model does not behave in the same way for the different pool schemes: indeed, in the first two cases it is preferable to adopt for a logarithmic pool, while in the last two cases, the logarithmic one is the worst, and the nearest to the 45 degree line is the linear pool. The decision than depend on the nature of the data to be forecasted.

A graphical inspection of PIT cumulative density functions of the three models are proposed to compare them with the simulated data to be predicted seen in the first columns of Fig. 4.2, 4.3 and 4.4. In all the experiments the

Table 4.1: Parameters estimates in the linear combination model for different choices of the mixture probabilities $\boldsymbol{p_i}$ of the data generating process.

P	(1/5, 1/5, 3/5)		(1/7, 1/7, 5/7)		(3/5, 1/5, 1/5)		(5/7, 1/7, 1/7)	
θ	BC1	BC2	BC1	BC2	BC1	BC2	BC1	BC2
α	0.755	3.293	0.921	6.970	0.461	0.452	0.496	0.650
β	0.642	0.953	0.639	0.937	0.816	3.744	0.812	0.876
ω	0.015	0.191	0.000	0.500	0.256	0.925	0.342	0.230
$\alpha*$		0.692		0.665		0.550		0.707
$\beta*$		3.093		0.713		0.827		13.033
$\omega*$		0.150		0.233		0.063		0.315
ρ		0.697		0.512		0.215		0.806

Table 4.2: Parameters estimates in the harmonic combination model for different choices of the mixture probabilities $\boldsymbol{p_i}$ of the data generating process.

P	(1/5, 1/5, 3/5)		(1/7, 1/7, 5/7)		(3/5, 1/5, 1/5)		(5/7, 1/7, 1/7)	
θ	BC1	BC2	BC1	BC2	BC1	BC2	BC1	BC2
α	0.744	7.026	0.906	7.775	0.416	0.383	0.457	0.457
β	0.634	0.878	0.632	1.013	0.755	0.827	0.747	0.778
ω	0.042	0.529	0.024	0.456	0.363	0.734	0.507	0.511
$\alpha*$		0.615		0.665		3.720		0.462
$\beta*$		0.929		0.651		1.133		0.734
$\omega*$		0.380		0.302		0.093		0.474
ρ		0.453		0.415		0.824		0.456

Table 4.3: Parameters estimates in the logarithmic combination model for different choices of the mixture probabilities $\boldsymbol{p_i}$ of the data generating process.

P	(1/5, 1/5, 3/5)		(1/7, 1/7, 5/7)		(3/5, 1/5, 1/5)		(5/7, 1/7, 1/7)	
θ	BC1	BC2	BC1	BC2	BC1	BC2	BC1	BC2
α	0.751	7.062	0.917	6.514	0.441	2.587	0.469	2.180
β	0.639	0.950	0.640	0.966	0.764	1.109	0.753	0.869
ω	0.018	0.517	0.000	0.431	0.370	0.031	0.465	0.411
$\alpha*$		0.578		0.645		0.367		0.515
$\beta*$		0.823		0.680		0.875		2.770
$\omega*$		0.426		0.379		0.843		0.423
ρ		0.484		0.510		0.274		0.389

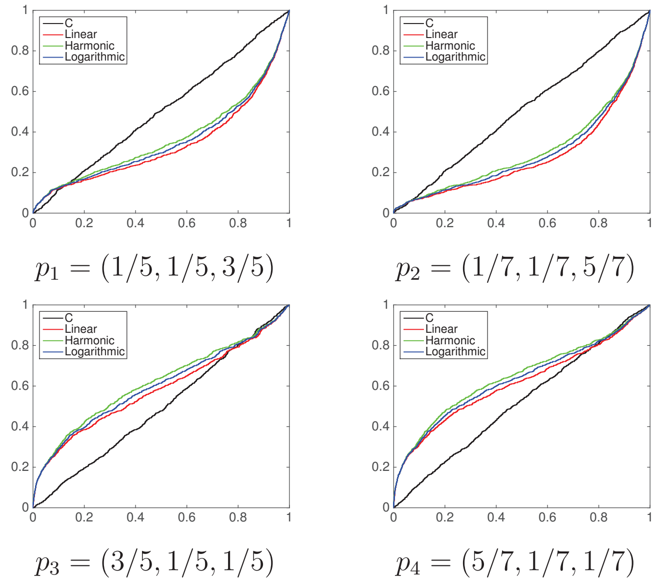
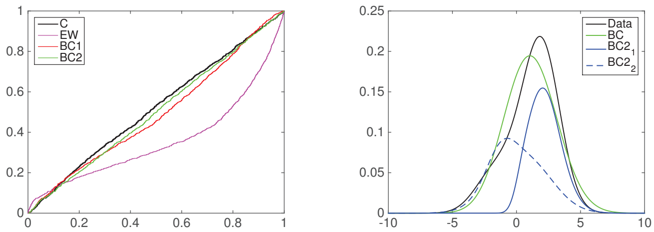


Figure 4.1: PITs cdf for EW model with linear, harmonic and logarithmic pools at different values of $\boldsymbol{p_i}$.

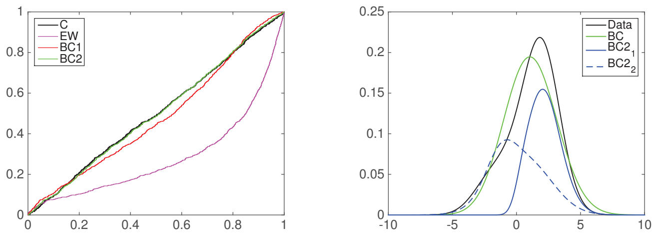
PITs of the equally weighted model (magenta line) lack to predict acceptably the standard uniform cdf of the data simulated by a mixture of normal distributions.

The beta-transformed models (BC1) (red line) predict better the uniformity than the EQ models, but they overestimate or underestimate the black line mainly in the central part of the support. In all the pool typology used, the beta-mixture models (BC2) provide the closest calibrated cdfs to the uniform one, being able to achieve a better flexibility among the others.

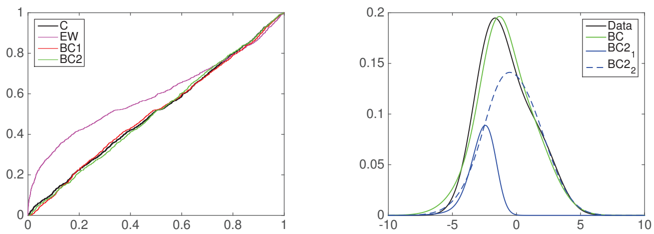
To highlight the behaviour of the two-component beta mixture (BC2), the right column of figures 4.2, 4.3 and 4.4 show the contribution in the calibration process of each element. As an example, consider the first panel of Fig. 4.2, the BC1 and BC2 models with linear pooling. The solid blue line represent the probability density function of the first component of the mixture, the dashed blue line the second component. Multimodality of the data is explained by two predictable functions: the first item $BC2_1$ calibrates mainly the predictive density over the positive part of the support; the second mixture component, denoted with $BC2_2$, calibrates the density over the negative part. Table 4.1 reports the following values for the weight ω: $\omega_1 = 0.191$ and $\omega_2 = 0.150$. This means that both components weights



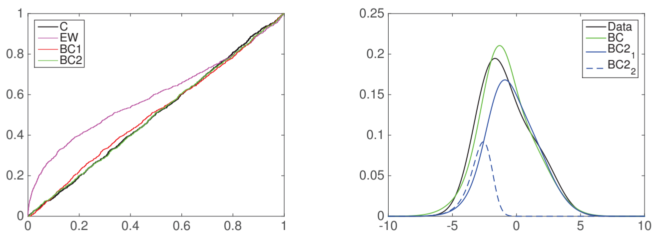
$p_1 = (1/5, 1/5, 3/5)$



$p_2 = (1/7, 1/7, 5/7)$

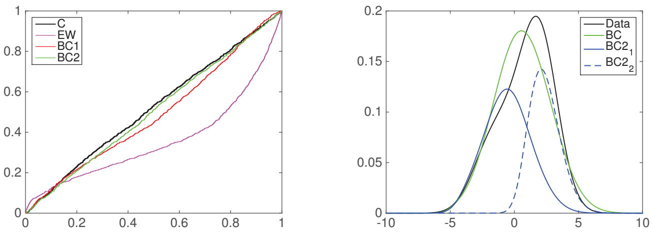


$p_3 = (3/5, 1/5, 1/5)$

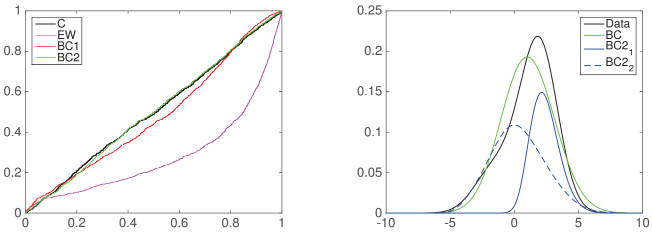


$p_4 = (5/7, 1/7, 1/7)$

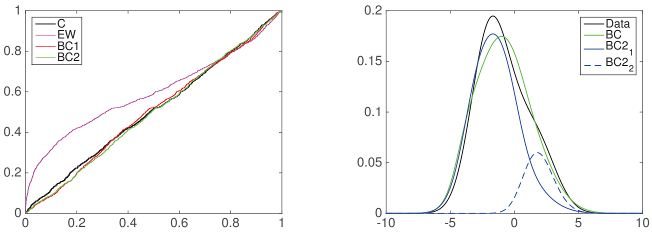
Figure 4.2: Left column: PITs cdf for linear pool at different values of $\boldsymbol{p_i}$. Right column: Contribution of the calibration components for BC1 (green line) and BC2 (blue line), where $BC2_1$ (solid) is the first component of the beta mixture in BC2, and $BC2_2$ (dashed) the second component.



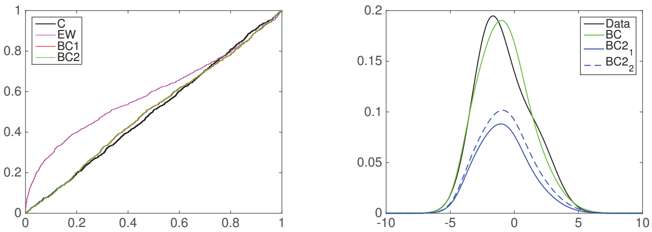
$p_1 = (1/5, 1/5, 3/5)$



$p_2 = (1/7, 1/7, 5/7)$

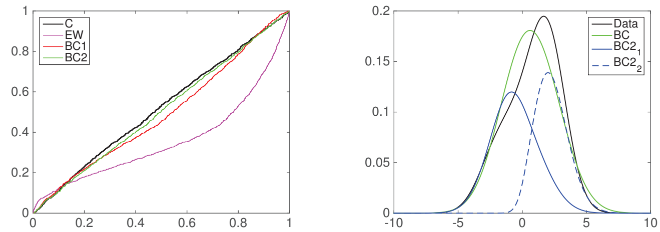


$p_3 = (3/5, 1/5, 1/5)$

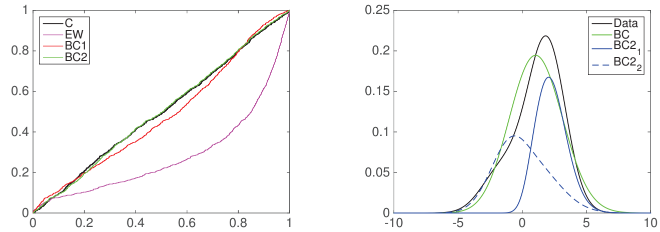


$p_4 = (5/7, 1/7, 1/7)$

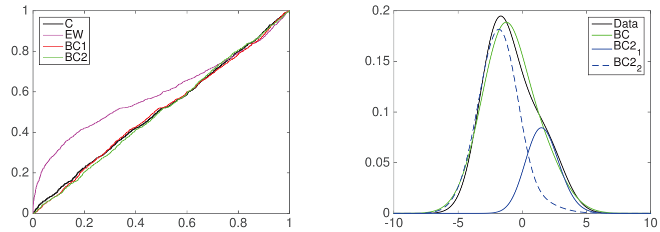
Figure 4.3: Left column: PITs cdf for Harmonic pool at different values of $\boldsymbol{p_i}$. Right column: Contribution of the calibration components for BC1 (green line) and BC2 (blue line), where $BC2_1$ (solid) is the first component of the beta mixture in BC2, and $BC2_2$ (dashed) the second component.



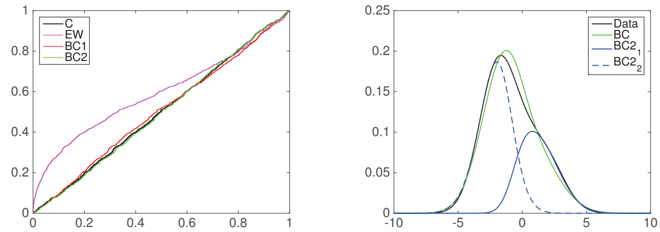
$p_1 = (1/5, 1/5, 3/5)$



$p_2 = (1/7, 1/7, 5/7)$



$p_3 = (3/5, 1/5, 1/5)$



$p_4 = (5/7, 1/7, 1/7)$

Figure 4.4: Left column: PITs cdf for Logarithmic pool at different values of $\boldsymbol{p_i}$. Right column: Contribution of the calibration components for BC1 (green line) and BC2 (blue line), where $BC2_1$ (solid) is the first component of the beta mixture in BC2, and $BC2_2$ (dashed) the second component.

more the second model in the pool, i.e. $\mathcal{N}(2,2)$.

In conclusion we had proved the validity of the result achieved in Bassetti et al. (2015) for calibration and linear combination of predictive densities are valid for and can be extended to other pooling schemes, including the harmonic pooling and the logarithmic pooling. Moreover, no clear preference for one combination scheme appears from our examples.

4.2 Financial Application: Standard&Poors500 Index

In this section we illustrate the effectiveness of the beta-mixture calibration with an application to the S&P500 index. Moreover this is an extension of the dynamic predictive density combinations in Casarin et al. (2015a) whose database for S&P500 is used to employ the following section.

We consider S&P500 daily percent log returns from January, 1^{st} 2007 to December, 31^{st} 2009; an updated version of the database used in Geweke and Amisano (2010), Geweke and Amisano (2011), and Fawcett et al. (2013). The price series $\{y_t\}$ were constructed assembling data from different sources: the WRDS database; Thompson/Data Stream; the total number of returns in the sample is $t = 784$. Many investors (also institutional) try to replicate the performance of S&P500 index with a set of stocks, not necessarily identical to those included in the index. Casarin et al. (2015a) individuate 3,712 stocks quoted in the NYSE and NASDAQ eligible for this purpose, whose 1,883 satisfy the control for liquidity (i.e. each stock has been traded a number of days corresponding to at least 40% of the sample size).
Then, a density forecast for each of the stock prices is produced by 4.2.

$$y_{it} = c_i + \kappa_{it}\zeta_{it}\kappa_{it}^2 = \theta_{i0} + \theta_{i1}\zeta_{i,t-1}^2 + \theta_{2\kappa_i,t-1}^2 \tag{4.2}$$

where y_{it} is the log return of stock $i = 1, \cdots, 1,883$ at day $t = 784$, $\zeta_{i,t-1} \sim \mathcal{N}(0,1)$ and $\zeta_{i,t-1} \sim \mathcal{T}(\nu_i)$ for the Normal and t-Student cases, respectively. Both models produce 784 one day ahead density forecast from January 1^{st} 2007 to December, 31^{st} 2009 by substituting the ML estimates for the unknown parameters $(c_i, \theta_{i0}, \theta_{i1}, \theta_{i2}, \nu_i)$ (to further details please refer to Casarin et al. (2015a)

The major contribution of this technic is the construction of a sequential cluster analysis to our forecasts. Casarin et al. (2015a) compute two clusters:

one for Normal GARCH(1,1) models and another for t-GARCH(1,1). Then, is our aim now to obtain a combined forecast of the S&P500 index combining and calibrating the two classes of predictive distribution functions, i.e. GARCH(1,1) and t-GARCH(1,1), through the equally weighted, the beta calibrated and the two-components beta mixture models with linear, harmonic and logarithmic pooling schemes.
The clustered weights here is assumed to be one and defined by:

$$\omega_i = \begin{cases} \frac{\omega}{3766}, & i \leq 1883 \\ \frac{1-\omega}{3766}, & i > 1883 \end{cases}$$

where 3766 are total the number of predictive distribution function: 1883 belonging to the class GARCH(1,1) and 1883 to the class t-GARCH(1,1). i.e. the combination model attains weight $\omega_i/3766$ to the class of GARCH(1,1) (first 1883 models) and $1 - \omega_i$ to the class of t-GARCH(1,1) (second 1883 models). The stage is open to further extensions, suggesting a less restricting weighting system.

The period taken into account is particularly interesting because it overpasses the financial crisis. The effect of this event is well evident in the figure 4.5 which shows the time variation and the clustering effect in the volatility of the daily daily percent log returns of S&P500 index over the time. Here the analysis is split in three subsample of 200 observations each representing three periods with different features: from January, 1^{st} 2007 to October, 5^{th} 2007, is consider a tranquil period, and the predictability of the index could be hypothesised better than the one from June, 20^{th} 2008 to March, 26^{th} 2009 during which the financial crisis has its development: here one can expect that the high volatility makes more difficult to predict the returns; finally, the third subsample considers data from March, 27^{th} 2009 to December, 31^{st} 2009, the post-crisis period and will be interesting to inquire if some difficulties in the forecastability is still present or not.

The two classes of predictive density functions GARCH(1,1), t-GARCH(1,1) are combined and calibrated through the models presented in section 4: the equally weighted (EW) model, the beta calibrated (BC1) model and the two-mixture beta calibrated (BC2) model for each combination scheme: linear, harmonic and logarithmic.
Figure 4.6 displays a comparison through PITs of linear, harmonic and logarithmic pools when those are combined with the equally weighted model and the 45 degree line, which represents the PITs for the unknown ideal model. Linear, harmonic and logarithmic pools have the same behaviour in

the mainly part of the support, the differences among them are in the tails, in particular in the left one. With respect to linear and logarithmic scheme indeed, the harmonic pool (blue line) underestimates more the frequency of the observations in the tails. The scheme closer to the 45 degree line is the harmonic one, thanks to its better performance in predicting tail events.

For a period of the first 200 days, from January, the 1^{st} 2007 to October, the 10^{th} 2007, where the volatility is roughly the same, the posterior means of BC1 and BC2 parameters (represented by the vector $\boldsymbol{\theta}$), are reported in Tables 4.4, 4.5, 4.6. Here, α and β stands for the parameters of the beta distribution in the estimated BC1 model and in the first component of the BC2 model, while the second component of BC2 is referred to α* and β*. In all the cases presented, the estimated BC1 models give zero weight to the class of t-GARCH(1,1) models, as well as the fist component of the beta mixture (BC2). While second component of the BC2, in the harmonic and logarithmic cases weights more class of t-GARCH(1,1) models than the class of GARCH(1,1) models. To better understand the effect of these parameters estimates, a graphical inspection of PITs is reported in figures 4.7, 4.8 4.9, for the pre-crisis, in-crisis and post crisis period respectively.

In all the pooling scheme applied the PITs of the equally weighted model (magenta line) lack to predict acceptably the ideal standard uniform cdf; as it is possible to see in figures 4.7 - 4.9, just in one case, the linear one, both BC1 and BC2 perform well, providing the closest calibration to the uniform one, being able to achieve a better flexibility for all the time periods analysed. In the harmonic and logarithmic cases, the BC1 model lack to calibrate class of GARCH(1,1) and the class of t-GARCH(1,1) models fitting even worsen than the EW model. However, a satisfactory calibration is obtain by the BC2 model, even if, not as good as that achieved in the linear pool. This is verified for all the periods of time analysed, even if, PITs calibration gets worse in the crisis and post crisis fases, highlighting some difficulties in being flexible. However, the linear pooling achieves good calibrated forecasts in both beta combination models; if the pool employed is chosen among the harmonic and the logarithmic schemes, satisfactory results are provided by the two-component beta mixture model.

In conclusion we had proved that the result achieved in Bassetti et al. (2015) for beta-mixture calibration and linear combination of predictive densities are still valid when harmonic and logarithmic combination schemes are used.

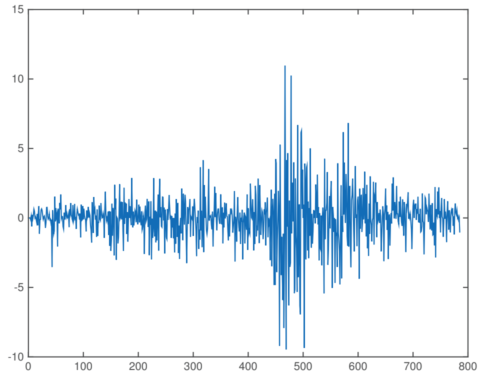


Figure 4.5: Different behaviour of the EW model for the three pool schemes applied to S&P500 daily percent log return .

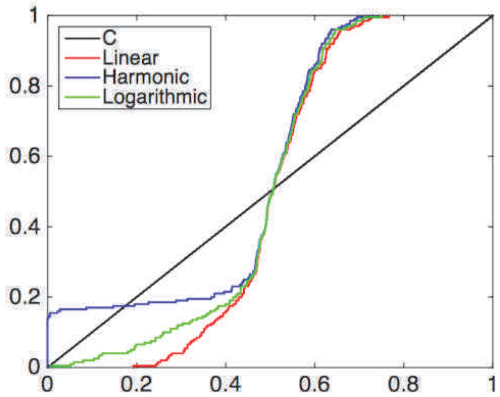


Figure 4.6: Different behaviour of the EW model for the three pool schemes applied to S&P500 daily percent log return .

Table 4.4: Parameters estimates in the different combination models for pre-crisis data subsample: January, 1^{st} 2007 - October, 5^{th} 2007.

P	Linear		Harmonic		Logarithmic	
θ	BC1	BC2	BC1	BC2	BC1	BC2
α	5.840	0.000	0.084	17.573	2.468	34.692
β	5.807	0.000	0.371	15.114	2.867	34.462
ω	1.000	0.000	1.000	0.863	1.000	0.706
$\alpha*$		5.812		0.020		1.781
$\beta*$		5.651		0.466		2.166
$\omega*$		1.000		0.199		0.93
ρ		0.000		0.7926		0.269

Table 4.5: Parameters estimates in the different combination models for in-crisis data subsample: June, 20^{th} 2008 - March, 26^{th} 2009.

P	Linear		Harmonic		Logarithmic	
θ	BC1	BC2	BC1	BC2	BC1	BC2
α	7.025	278.600	0.977	0.944	0.974	1.010
β	6.646	803.260	0.865	1.014	1.292	1.018
ω	1.000	1.000	0.740	0.263	0.821	0.031
$\alpha*$		6.760		0.975		1.131
$\beta*$		6.334		1.010		0.972
$\omega*$		1.000		0.247		0.298
ρ		0.000		0.000		0.000

Table 4.6: Parameters estimates in the different combination models for pre-crisis data subsample: March, 27^{th} 2009 - December, 31^{st} 2009.

P	Linear		Harmonic		Logarithmic	
θ	BC1	BC2	BC1	BC2	BC1	BC2
α	6.542	47110.000	1.031	0.972	1.127	1.007
β	6.071	0.000	0.419	0.942	2.275	1.066
ω	1.000	1.000	0.823	0.967	0.186	0.406
$\alpha*$		6.710		1.039		0.891
$\beta*$		6.307		0.938		1.015
$\omega*$		1.000		0.920		0.921
ρ		0.000		0.000		0.000

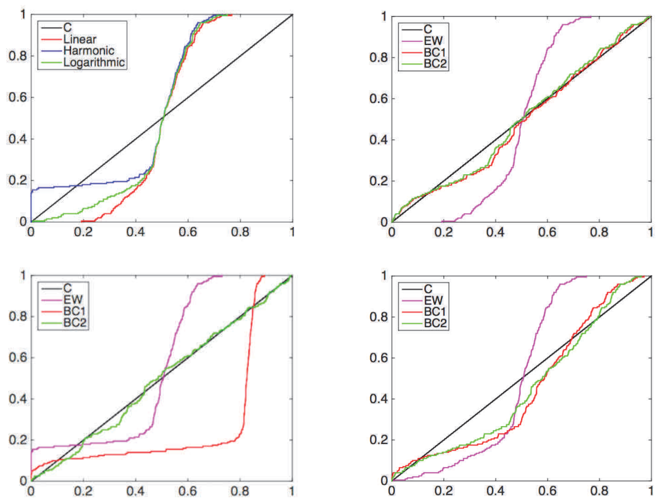


Figure 4.7: PITs cdf of the ideal model C (black line), EW (magenta), BC1(red) and BC2(green) for Linear (top right), Harmonic (bottom left) and Logarithmic (bottom right) pools and PiTs of EW models (top left) for linear (red), harmonic (blue) and logarithmic (green), in the first data subsample: pre-crisis period

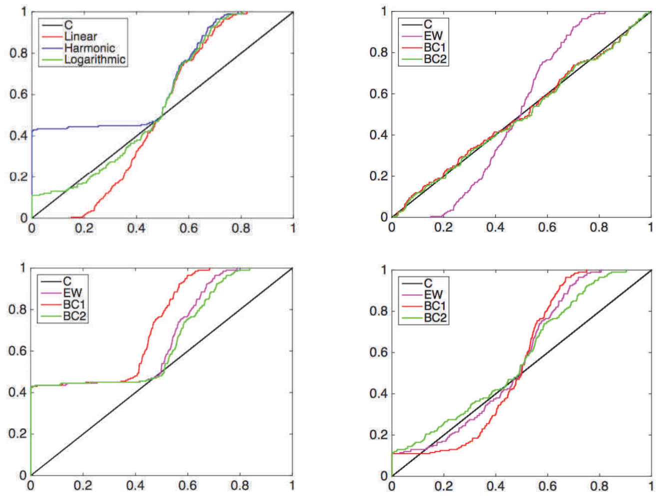


Figure 4.8: PITs cdf of the ideal model C (black line), EW (magenta), BC1(red) and BC2(green) for Linear (top right), Harmonic (bottom left) and Logarithmic (bottom right) pools and PiTs of EW models (top left) for linear (red), harmonic (blue) and logarithmic (green), in the second data subsample: in-crisis period

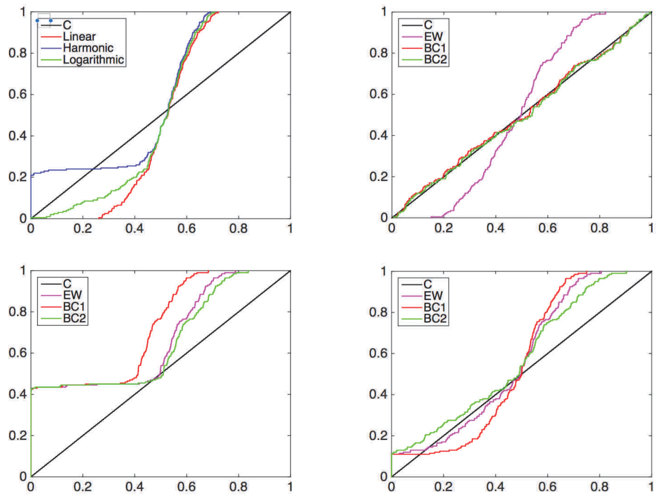


Figure 4.9: PITs cdf of the ideal model C (black line), EW (magenta), BC1(red) and BC2(green) for Linear (top right), Harmonic (bottom left) and Logarithmic (bottom right) pools and PiTs of EW models (top left) for linear (red), harmonic (blue) and logarithmic (green), in the third data subsample: post-crisis period

Chapter 5

Results

The aim of this paper was investigate combination and calibration models in some generalised pooling schemes. Starting from what explain in Bassetti et al. (2015), the combination and calibration models for the linear pool are generalised for other pooling schemes, among which the harmonic and the logarithmic pool. Since the three schemes behave differently, this paper would inquire if the results in Bassetti et al. (2015) are verified also with a generalise pools of predictive distributions.

The results of the simulation example supplied a substantially equivalence among the polling schemes, providing the same evidence for combination and calibration models than those in Bassetti et al. (2015). However, this equivalence was not verified in our application to daily log returns of S&P500 index for the period January 1^{st} 2007 - December, 31^{st} 2009: the application of combination and calibration schemes gives the results of the simulation example just for the linear pooling, while for the others schemes the results are verified just for what concerns the two-component beta mixture model.

Appendix A

Gibbs sampling details

Gibbs sampling details In this work, it was design a Metropolis-Hastings for posterior inference. In chapter 4, a beta and beta-mixture calibration models are presented and applied to simulated data.

A Metropolis-Hastings (MH) algorithm has been used to approximate the posterior distribution of the unknown parameters $\boldsymbol{\theta} = (\alpha, \beta, \omega)$ and $\boldsymbol{\theta} = (\alpha_1, \beta_1, \omega_1, \alpha_2, \beta_2, \omega_2, \rho)$ for the beta calibrated model and the beta-mixture calibrated model respectively. The joint posterior for $J = 2$ is reported in the chapter 3.

In order to account for the constrains on the parameters, the target distributions of the MH algorithm for μ, ν, ω and ρ is obtain by applying a change of variable: $\mu = 1/(1 + exp\{-\hat{\theta}_1\})$, $\nu = exp\{\hat{\theta}_2\}$ and $\omega = 1/(1+exp\{-0.1\hat{\theta}_3\})$ to the joint posterior for BC1 and the target distributions for $\mu_1, \mu_2, \nu_1, \nu_2, \omega_1, \omega_2$ and ρ the change of variable $\mu_1 = 1/(1 + exp\{-\hat{\theta}_1\})$, $\nu_1 = exp\{\hat{\theta}_2\}$, $\omega_1 = 1/(1 + exp\{-0.1\hat{\theta}_3\})$, $\mu_2 = 1/(1 + exp\{-0.1\hat{\theta}_4\})$, $\nu_2 = exp\{\hat{\theta}_5\}$, $\omega_2 = 1/(1 + exp\{-0.1\hat{\theta}_6\})$, $\rho = 1/(1 + exp\{-0.1\hat{\theta}_7\})$. The MH acceptance probability accounts for the Jacobian:

$$J(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = \left| \begin{pmatrix} -0.1\frac{exp\{-0.1\hat{\theta}_1\}}{(1+exp\{-0.1\hat{\theta}_1\})^2} & 0 & 0 \\ 0 & exp\{\hat{\theta}_2\} & 0 \\ 0 & 0 & -0.1\frac{exp\{-0.1\hat{\theta}_3\}}{(1+exp\{-0.1\hat{\theta}_3\})^2} \end{pmatrix} \right|$$

that is

$$\begin{aligned} J(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = 0.1^2 exp\{-0.1\hat{\theta}_1 + \hat{\theta}_2 - 0.1\hat{\theta}_3\}(1 + exp\{-0.1\hat{\theta}_1\})^{-2} \\ (1 + exp\{-0.1\hat{\theta}_3\})^{-2} \end{aligned} \tag{1}$$

for the BC1, and

$$J(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_6, \hat{\theta}_7) =$$

$$= \left| \begin{pmatrix} -0.1\frac{exp\{-0.1\hat{\theta}_1\}}{(1+exp\{-0.1\hat{\theta}_1\})^2} & 0 & 0 \\ 0 & exp\{\hat{\theta}_2\} & 0 \\ 0 & 0 & -0.1\frac{exp\{-0.1\hat{\theta}_3\}}{(1+exp\{-0.1\hat{\theta}_3\})^2} \end{pmatrix} \right| \cdot$$

$$\cdot \left| \begin{pmatrix} -0.1\frac{exp\{-0.1\hat{\theta}_4\}}{(1+exp\{-0.1\hat{\theta}_4\})^2} & 0 & 0 & 0 \\ 0 & exp\{\hat{\theta}_5\} & 0 & 0 \\ 0 & 0 & -0.1\frac{exp\{-0.1\hat{\theta}_6\}}{(1+exp\{-0.1\hat{\theta}_6\})^2} & 0 \\ 0 & 0 & 0 & -0.1\frac{exp\{-0.1\hat{\theta}_7\}}{(1+exp\{-0.1\hat{\theta}_7\})^2} \end{pmatrix} \right|$$

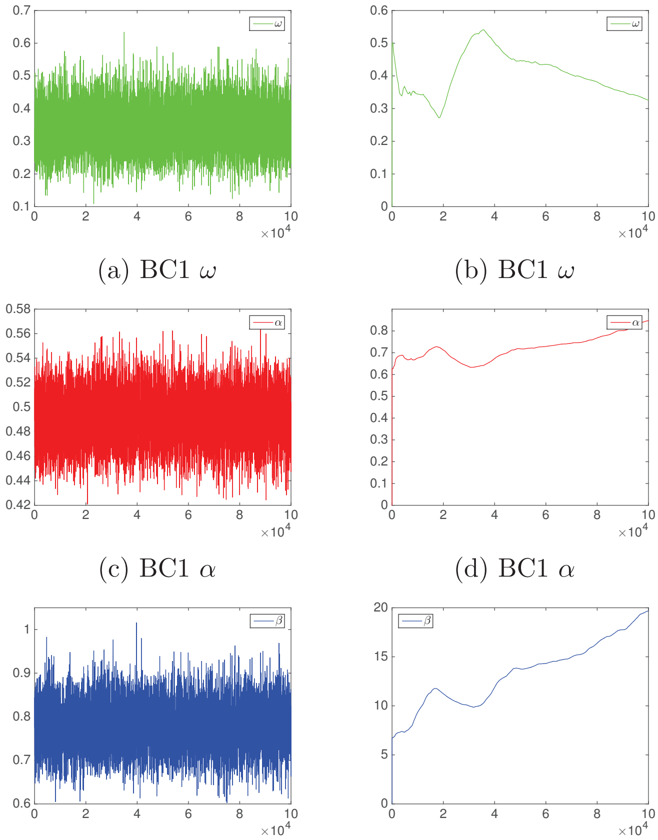
that is

$$\begin{aligned} J(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_6, \hat{\theta}_7) = 0.1^5 exp\{-0.1(\hat{\theta}_1 + \hat{\theta}_3 + \hat{\theta}_4 + \hat{\theta}_6 + \hat{\theta}_7) + \hat{\theta}_2 + \hat{\theta}_5\} \\ (1 + exp\{-0.1\hat{\theta}_1\})^{-2}(1 + exp\{-0.1\hat{\theta}_3\})^{-2}(1 + exp\{-0.1\hat{\theta}_4\})^{-2} \\ (1 + exp\{-0.1\hat{\theta}_6\})^{-2}(1 + exp\{-0.1\hat{\theta}_7\})^{-2} \quad (2) \end{aligned}$$

for the BC2 model.

Where equation 1 and 2 report the Jacobian used in the MH acceptance ratios for the BC1 and BC2 models, respectively. The variance/covariance matrix of the MH proposal distribution are $\Sigma = diag\{0.05^2, 0.1^2, 10^2\}$ and $\Sigma = diag\{0.05^2, 0.05^2, 10^2, 0.05^2, 0.05^2, 10^2, 10^2\}$ for the BC1 and BC2 models, respectively. We set the initial values of the MH sampler to the maximum likelihood estimate of the parameter $\boldsymbol{\theta}$.

The assumption of i.i.d. samples generated by the MH algorithm and the ergodic theorem guaranty the almost sure convergence of the MH empirical averages to the posterior mean. An example of the MCMC output and of ergodic means, for one of the experiments, is given in Fig. 1 for the parameters of the BC1 model and in Fig. 2-3 for the parameters in BC2 model.

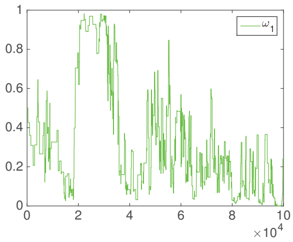


(a) BC1 ω (b) BC1 ω

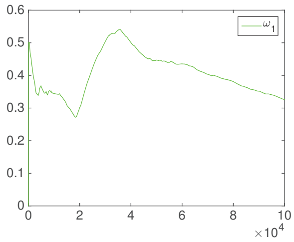
(c) BC1 α (d) BC1 α

(e) BC1 β (f) BC1 β

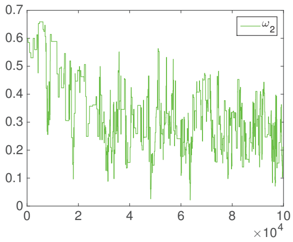
Figure 1: Samples of BC1 model: variable inspection and its empirical average for 1000 iteractions



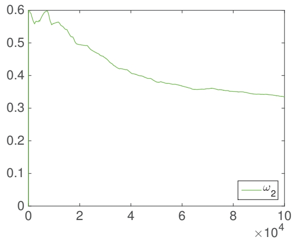
(a) BC2 $\omega 1$



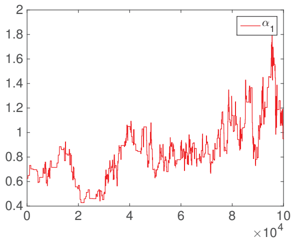
(b) BC2 $\omega 1$



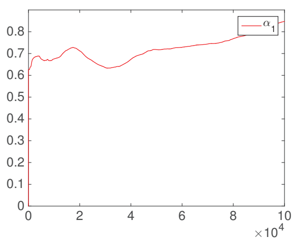
(c) BC2 $\omega 2$



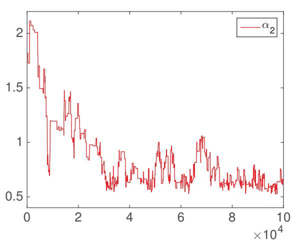
(d) BC2 $\omega 2$



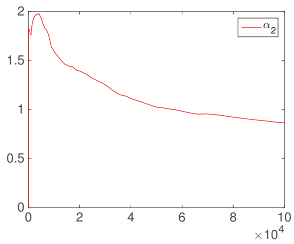
(e) BC2 $\alpha 1$



(f) BC2 $\alpha 1$

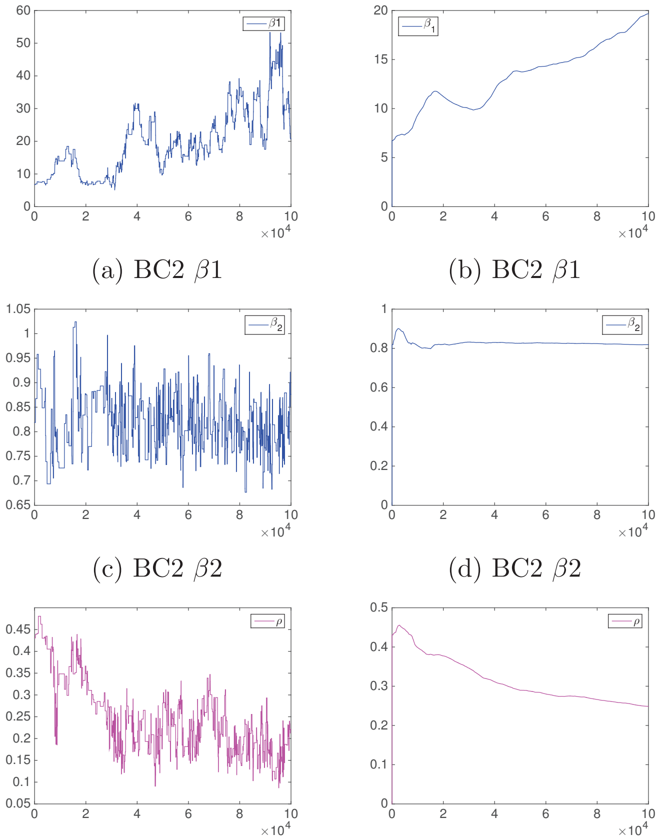


(g) BC2 $\alpha 2$



(h) BC2 $\alpha 2$

Figure 2: Samples of BC2 model: variable inspection and its empirical average for 1000 iteractions



(a) BC2 $\beta 1$ (b) BC2 $\beta 1$

(c) BC2 $\beta 2$ (d) BC2 $\beta 2$

(e) BC2 ρ (f) BC2 ρ

Figure 3: Samples of BC2 model: variable inspection and its empirical average for 1000 iteractions

Bibliography

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

Altomare, F. and Campiti, M. (1994). *Korovkin-type Approximation Theory and Its Application*. De Gruyter, Berlin.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2:1152–1174.

Bacharach, M. (1975). Group decisions in the face of differences of opinion. *Management Science*, 22:182–191.

Barnard, G. A. (1963). New methods of quality control. *Journal of the Royal Statistical Society, Series A*, 126:255–259.

Bassetti, F., Casarin, R., and Leisen, F. (2014). Beta-product dependent Pitman-Yor processes for Bayesian inference. *Journal of Econometrics*, 180:49–72.

Bassetti, F., Casarin, R., and Ravazzolo, F. (2015). Bayesian nonparametric calibration and combination of predictive distributions. *University Ca' Foscari of Venice, Dept. of Economics Research Paper Series No. 04/WP/2015.*

Bates, J. M. and Granger, C. W. J. (1969). Combination of forecasts. *Operational Research Quarterly*, 20:451–468.

Billio, M. and Casarin, R. (2010). Identifying Business Cycle Turning Points with Sequential Monte Carlo: An Online and Real-Time Application to the Euro Area. *Journal of Forecasting*, 29:145–167.

Billio, M. and Casarin, R. (2011). Beta autoregressive transition Markov-switching models for business cycle analysis. *Studies in Nonlinear Dynamics and Econometrics*, 15:1–32.

Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177:213–232.

Bouguila, N., Ziou, D., and Monga, E. (2006). Practical Bayesian estimation of a finite beta mixture through Gibbs sampling and its applications. *Statistics and Computing*, 16:215–225.

Burda, M., Harding, M., and Hausman, J. (2014). A Bayesian semiparametric competing risk model with unobserved heterogeneity. *Journal of Applied Econometrics*.

Casarin, R., Dalla Valle, L., and Leisen, F. (2012). Bayesian model selection for beta autoregressive processes. *Bayesian Analysis*, 7:1–26.

Casarin, R., Grassi, s., Ravazzolo, F., and van Dijk, H. K. (2015a). Dynamic predictive density combinations for large data sets in economics and finance. *Working Paper University Ca' Foscari of Venice, Dept. of Economics*.

Casarin, R., Leisen, F., Molina, G., and Ter Horst, E. (2015b). Beta markov random field calibration of the term structure of implied risk neutral densities. *Bayesian Analysis*, 10(4):791–819.

Chib, S. and Hamilton, B. H. (2002). Semiparametric bayes analysis of longitudinal data treatment models. *Journal of Econometrics*, 110:67–89.

Choudhuri, N., Ghosal, S., and Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059.

Corradi, V. and Swanson, N. R. (2006). Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics*, 135:187–228. 1-2.

Dawid, A. P. (1982). Intersubjective statistical models. *Exchangeability in Probability and Statistics*, pages 217–232.

Dawid, A. P. (1984). Statistical theory: The prequential approach. *Journal of the Royal Statistical Society Series A*, 147:278–290.

De Finetti, B. (1970). Logical foundation and measurement of subjective probability. *Acta Psychologica*, 34:129–145.

DeGroot, M. H., Dawid, A. P., and Mortera, J. (1995). Coherent combination of experts' opinions. *Test*, 4:263–313.

DeGroot, M. H. and Mortera, J. (1991). Optimal linear opinion pools. *Management Science*, 37(5):546–558.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39:863–883.

Epstein, E. S. (1966). Quality control for probability forecasts. *Monthly Weather Review*, 94:487–494.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89:268–277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588.

Fawcett, N., Kapetanios, G., Mitchell, J., and Price, S. (2013). Generalised density forecast combinations. *Journal of Econometrics, forthcoming*.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1:209–230.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer-Verlag, Berlin.

Genest, C. and Zidek, J. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–148.

Geweke, J. (2010). *Complete and Incomplete Econometric Models*. Princeton University Press, Princeton.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26:216–230.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164:130–141.

Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.*, 35(2):697–723.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian nonparametrics.* Springer Series in Statistics. Springer-Verlag, New York.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B*, 69:243–268.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151.

Gneiting, T., Raftery, A., Westveld, A. H., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133:1098–1118.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold and quantile weighted proper scoring rules. *Journal of Business and Economic Statistics*, 29:411–422.

Gneiting, T. and Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, 7:1747–1782.

Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3:197–204.

Griffin, J. E. (2011). Inference in infinite superpositions of non-Gaussian Ornstein-Uhlenbeck processes using Bayesian nonparametric methods. *Journal of Financial Econometrics*, 1:1–31.

Griffin, J. E. and Steel, M. F. J. (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, 101:179–194.

Griffin, J. E. and Steel, M. F. J. (2011). Stick-breaking autoregressive processes. *Journal of Econometrics*, 162:383–396.

Guidolin, M. and Timmermann, A. (2009). Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach. *Journal of Econometrics*, 150:297–311.

Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23:1–13.

Hatjispyros, S. J., Nicoleris, T. N., and Walker, S. G. (2011). Dependent mixtures of Dirichlet processes. *Computational Statistics & Data Analysis*, 55:2011–2025.

Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica*, 70:781–799.

Hjort, N. L., Homes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–417.

Hora, S. C. (2010). An analytic method for evaluating the performance of aggregation rules for probability densities. *Operations Research*, 58:1440–1449.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96:161–173.

Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87:371–390.

Ishwaran, H. and Zarepour, M. (2009). Series representations for multivariate generalized gamma processes via a scale invariance principle. *Statistica Sinica*, 19:1665–1682.

Jensen, J. M. and Maheu, M. J. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157:306–316.

Jochmann, M. (2015). Modeling U.S. inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews*, 34(5):537–558.

Jore, A. S., Mitchell, J., and Vahey, S. P. (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25:621–634.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, 21:93–105.

Kling, J. L. and Bessler, D. A. (1989). Calibration-based predictive distributions: An application of prequential analysis to interest rates, money, prices, and output. *Journal of Business*, 62:477–499.

Laddaga, R. (1977). Lehrer and the consensus proposal. *Synthese*, 36:473–477.

Lerch, S. and Thorarinsdottir, T. L. (2013). Comparison of nonhomogeneous regression models for probabilistic wind speed forecasting. *Tellus Series A*, 65:21206.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics*, 12:351–357.

MacEachern, S. N. and Müller, P. (1998). Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 7:223–238.

McConway, K. J. (1981). Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76:410–415.

Mitchell, J. and Hall, S. G. (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESER "fan" charts of inflation. *Oxford Bulletin of Economics and Statistics*, 67:995–1033.

Mitchell, J. and Wallis, K. F. (2010). Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration and Sharpness. *Journal of Applied Econometrics*, forthcoming.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122:73–119.

Müller, P., Quintana, F., and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society B*, 66:735–749.

Norets, A. and Pelenis, J. (2012). Bayesian modeling of joint and conditional distributions. *Journal of Econometrics*, 168(332-346).

Norets, A. and Pelenis, J. (2015). Posterior consistency in conditional density estimation by covariate dependent mixtures. *EconometricTheory, forthcoming.*

Papaspiliopoulos, O. (2008). A note on posterior sampling from Dirichlet mixture models. Technical Report 8-20, CRiSM, University of Warwick.

Papaspiliopoulos, O. and Roberts, G. (2008). Retrospective Markov chain Monte Carlo for Dirichlet process hierarchical models. *Biometrika*, 95:169–186.

Pati, D., Dunson, D., and Tokdar, S. (2013). Posterior consistency in conditional distribution estimation. *Journal of Multivariate Analysis*, 116:456–472.

Pelenis, J. (2014). Bayesian regression with heteroscedastic error density and parametric mean function. *Journal of Econometrics*, 178:624–638.

Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics*, 27:105–126.

Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26:373–393.

Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society Series B*, 64:79–100.

Pinson, P. and Hagedorn, R. (2012). Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Meteorological Applications*, 19:484–500.

Ranjan, R. and Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):71–91.

Robert, C. P. and Rousseau, J. (2002). A mixture approach to Bayesian goodness of fit. Technical Report 02009, CEREMADE, Université Paris-Dauphine.

Roberts, H. V. (1965). Probabilistic prediction. *Journal of American Statistical Association*, 60:50–62.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2010). Latent stick-breaking processes. *Journal of the American Statistical Association*, 105:647–659.

Rodriguez, A. and ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis*, 3:339–366.

Rosenblatt, M. (1952). Remarks on multivariate transformation. *Annals of Mathematical Statistics*, 23:1052–1057.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester.

Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics*, 32:1339–1342.

Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking the intensity of violent crime. *Journal of the American Statistical Association*, 105:1403–1417.

Taddy, M. A. and Kottas, A. (2009). Markov switching Dirichlet process mixture regression. *Bayesian Analysis*, 4:793–816.

Tang, Y. and Ghosal, S. (2007). Posterior consistency of Dirichlet mixtures for estimating a transition density. *J. Statist. Plann. Inference*, 137(6):1711–1726.

Tang, Y., Ghosal, S., and Roy, A. (2007). Nonparametric Bayesian estimation of positive false discovery rates. *Biometrics*, 63(4):1126–1134, 1312.

Tay, A. S. and Wallis, K. F. (2000). Density forecasting: A survey. *Journal of Forecasting*, 19:235–254.

Terui, N. and van Dijk, H. K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, 18:421–438.

Thorarinsdottir, T. L. and Gneiting, T. (2010). Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society Series A*, 173:371–388.

Thorarinsdottir, T. L. and Johnson, M. S. (2012). Probabilistic wind gust forecasting using non-homogeneous Gaussian regression. *Monthly Weather Review*, 140:889–897.

Timmermann, A. (2006). Forecast combinations. In Elliot, G., Granger, C., and Timmermann, A., editors, *North-Holland*, volume 1 of *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier.

Tsyplakov, A. (2011). Evaluating density forecasts: A comment. *MPRA paper*, 31233.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics – Simulation and Computation*, 36:45–54.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, 67:983–994.

Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *Journal of the Royal Statistical Society Series B*, 62:159–180.

Wiesenfarth, M., C.M., H., Kneib, T., and Cadarso-Suarez, C. (2014). Bayesian nonparametric instrumental variables regression based on penalized splines and dirichlet process mixtures. *Journal of Business and Economic Statistics*, 32(3):468–482.

Wu, Y. and Ghosal, S. (2009a). Correction to: “Kullback Leibler property of kernel mixture priors in Bayesian density estimation” [mr2399197]. *Electron. J. Stat.*, 3:316–317.

Wu, Y. and Ghosal, S. (2009b). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electron. J. Stat.*, 2:298–331.