University of Innsbruck

A primer on bayesian distributional regression

Thomas Kneib, Nikolaus Umlauf

Working Papers in Economics and Statistics

2017-13

Research Platform
eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
Research platform “Empirical and Experimental Economics”
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

A Primer on Bayesian Distributional Regression

Thomas Kneib [1] **and Nikolaus Umlauf** [2]

[1] Chairs of Statistics and Econometrics, Georg-August-Universität Göttingen, Germany
[2] Department of Statistics, Faculty of Economics and Statistics, Universität Innsbruck, Austria

Address for correspondence: Thomas Kneib, Chairs of Statistics and Econometrics, Georg-August-Universität Göttingen, Humboldtallee 3, 37073 Göttingen, Germany.
E-mail: tkneib@uni-goettingen.de.
Phone: (+490) 551 39 25678.
Fax: (+490) 551 39 27279.

Abstract: Bayesian methods have become increasingly popular in the past two decades. With the constant rise of computational power even very complex models can be estimated on virtually any modern computer. Moreover, interest has shifted from conditional mean models to probabilistic distributional models capturing location, scale, shape and other aspects of a response distribution, where covariate effects can have flexible forms, e.g., linear, nonlinear, spatial or random effects. This tutorial paper discusses how to select models in the Bayesian distributional regression setting, how to monitor convergence of the Markov chains, evaluate relevance of effects using simultaneous credible intervals and how to use simulation-based inference also for quantities derived from the original model parameterisation. We exemplify the work flow using daily weather data on (i) temperatures on Germany's highest mountain and (ii) extreme values of precipitation all over Germany.

Key words: Distributional regression; generalized additive models for location, scale and shape; Markov chain Monte Carlo simulations; semiparametric regression; tutorial;

1 Introduction

While it is well known that the distribution of any quantity of interest is not well characterised by the mean alone, this is still frequently ignored in regression analyses since most common model specifications (linear models, generalized linear models, generalized additive models, etc.) focus on relating the mean of the response variable to a set of explanatory variables. While such models have the advantage of being easy to estimate and to interpret, they also have the risk of model misspecification which will often imply that conclusions drawn from the model estimates are misleading. In addition, in many applications the analysist may not only be interested in effects on the mean of the response, but rather on additional features of the response distribution. This is for example the case if one is

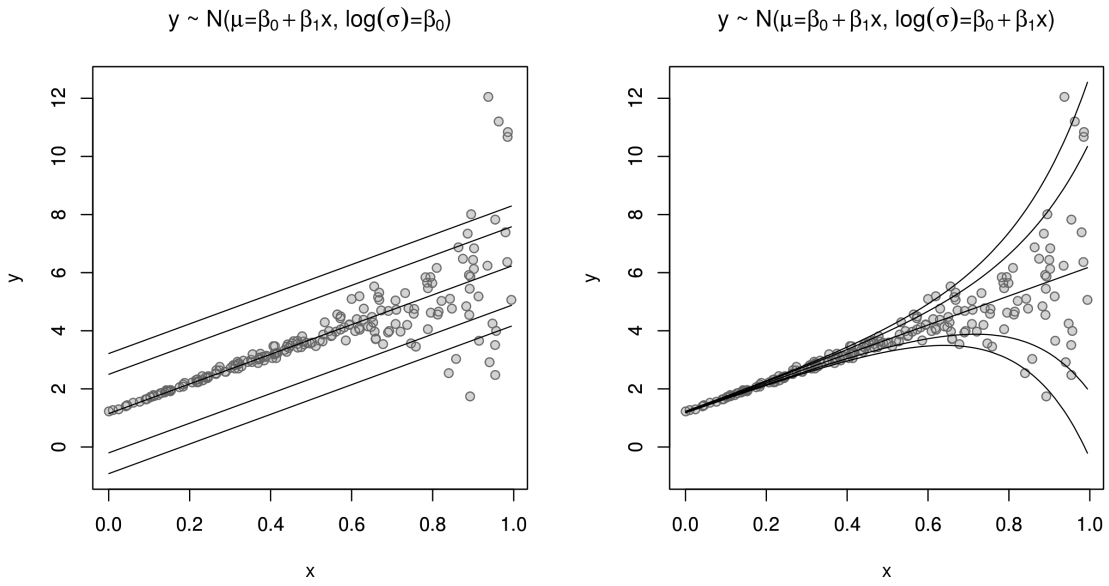

Figure 1: Simulated data example, shown are the estimated 2.5%, 10%, 50%, 90% and 97.5% quantiles of a homoscedastic normal model, left panel, and a heteroscedastic loaction-scale normal model.

interested in quantifying risks or extreme observations rather than the average behaviour of the response.

One common approach to overcome the obsession with means in regression analyses are quantile regression methods (see Koenker, 2005, for a detailled introduction) which are also discussed in another contribution to this special issue. While conditional quantiles provide a characterisation of local characteristics of the response distribution (indexed by the quantile level), we will focus on distributional regression models that provide a complete probabilistic characterisation of the response distribution in one joint model. These models will rely on the assumption of a specific distributional type for the response but allow potentially all parameters of this distribution to depend on covariates. This is illustrated in Figure 1 for simulated data from a normal distribution where both the mean the variance are related to a continuous covariate x. In particular, not only the mean but also the variance increases with larger values of the covariate such that the data are subject to heterogeneity. The left plot shows an estimated linear model for this data along with a 95% and an 80% prediction interval. While the tendency in the mean is still approximated reasonably well despite the misspecification of the variance, the prediction intervals derived from the model will be highly misleading. In contrast, the estimates in the right plot represent regression effects not only for the mean but also for the variance and nicely reflect the increasing variability associated with larger values of the covariate.

While in the particular case of a normal location-scale model with only one single covariate as in Figure 1 it is still fairly easy to understand the structure and implications of a dis-

tributional regression specification, this typically becomes more difficult in more complex specifications with either a more flexible specification of the regression predictor or the response distribution. This tutorial therefore aims at providing more insight for practitioners into the work flow of building distributional regression models illustrated alongside meteorological examples of daily temperature observations of the Zugspitz mountain in Germany and on extreme values of precipitation all over Germany. For inference, we will follow a Bayesian approach based on Markov chain Monte Carlo (MCMC) which are particularly attractive in distributional regression, since the inferential framework provides valid credible intervals for estimators in situations where confidence intervals for corresponding maximum likelihood estimators based on asymptotic properties fail and also enables immediate inference for quantities from the original model parameterisation. Since MCMC approaches imply a number of additional challenges for applied researchers (in particular monitoring mixing and convergence of Markov chains), we will also provide step by step guidance on those aspects of the analysis. Model fitting is carried out using the R (R Core Team, 2017) package **bamlss** (Umlauf et al., 2017). The R code to reproduce all examples is provided in the online materials of this paper. Since the Bayesian distributional modeling framework is very flexible and a lot of models are contained special cases, e.g., the generalized additive model (GAM, Hastie and Tibshirani, 1990), this tutorial paper also provides useful information on how to implement Bayesian inference in such simplified model types.

The remainder of this paper is structured as follows: The data used to illustrate the modeling work flow are presented in Section 2. In Section 3, the models and model terms supported by the distributional regression framework are briefly introduced. Section 4 gives an overview on monitoring convergence of the Markov chains while tools for model selection are discussed in Section 5. Bayesian inference based on the posterior samples is the illustrated in Section 6. The final section 7 comments on additional sources of more detailled information on distributional regression models.

2 German daily weather data

As illustrative examples we use two data sets of German daily weather observations provided by the *German Meteorological Service* (Deutscher Wetterdienst, DWD), which are freely available (`ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/daily/kl/historical/`). The data are a collection of over 1000 meteorological stations all across Germany and comprise several physical parameters. In the first analysis, we focus on 2 meters above the ground air temperature (`T`, in degree Celsius) of the highest mountain in Germany, the Zugspitze (47°N, 11°E; 2964m a.s.l.). The second example focuses on modeling extreme values of precipitation (`P`, in millimeters) all over Germany, i.e., values higher than 30 millimeters of daily precipitation sums.

2.1 Zugspitz daily temperature

The main interest in the first example is to evaluate trends in temperatures not only in terms of the average temperature but also in terms of changes of the general distributional shape. For estimating time trends using meteorological data, especially temperature, it is important to ensure that the data produced by the weather station is homogenized. For example, if a station is moved to another location with higher altitude, this can result in a dramatic drop of measured temperatures, similarly, exchanging physical sensors can shift meteorological observations which needs to be accounted for prior to analyzing the data. Since the Zugspitz daily temperature time series is one of the longest homogenized series in Germany, it is especially suited for estimating long term time trends.

In the top row of Figure 2, a histogram of the daily temperature observations together with a plot vs. the day of year is shown. The histogram shows that the distribution is not symmetric but slightly left-skewed towards lower temperatures, which had a minimum of $-33.1\,^{\circ}\mathrm{C}$ in February 1940. The maximum temperature of $+13.3\,^{\circ}\mathrm{C}$ was measured more recently in August of 2012, which may be seen as some indication of an increase in temperature over the years. The top right plot shows the seasonal pattern over all years. The peak temperatures are clearly during the summer times around August. Moreover, the 5% and 95.5% daily quantiles suggest that in the winter season from November to March temperatures vary more compared to the summer.

In the bottom row of Figure 2, the full temperature time series is shown. Compared to the global mean temperature of $-4.6\,^{\circ}\mathrm{C}$ it seems that after 1980 a temperature increase takes place, as indicated by the 20 years running mean. Before, the running mean fluctuates around its global mean in a more unstructured way. In addition, there is a very small increase in temperature variation over the last 30 years visible when comparing the 5% and 95% 20 years running quantiles.

In practice, (temperature) climatology models are are oftentimes used for improving probabilistic weather forecasts, e.g., by statistical downscaling or as a baseline model for verification. Therefore, it is important to capture the full information of the distribution such that correct inference can be made for all quantities of interest, e.g., for 5% and 95% quantiles. Therefore, as already suggested from the descriptive data analysis in Figure 2, next to autocorrelated observations, subsequent models need to account for complex seasonal patterns, possibly varying over time, as well as varying uncertainty over season and observation time. Due to the skewness of the response data, a simple Gaussian model is most likely not capable to model all features of the data. Therefore, we exemplify a complex distributional modeling task using three different distributions for daily temperature observations.

2.2 German extreme precipitation

This example focuses on extreme values of precipitation all over Germany, i.e., observations of the very right hand side of the distribution that exceed a threshold of 30 millimeters of precipitation (corresponding to the empirical 99.5% quantile of all precipitation measure-

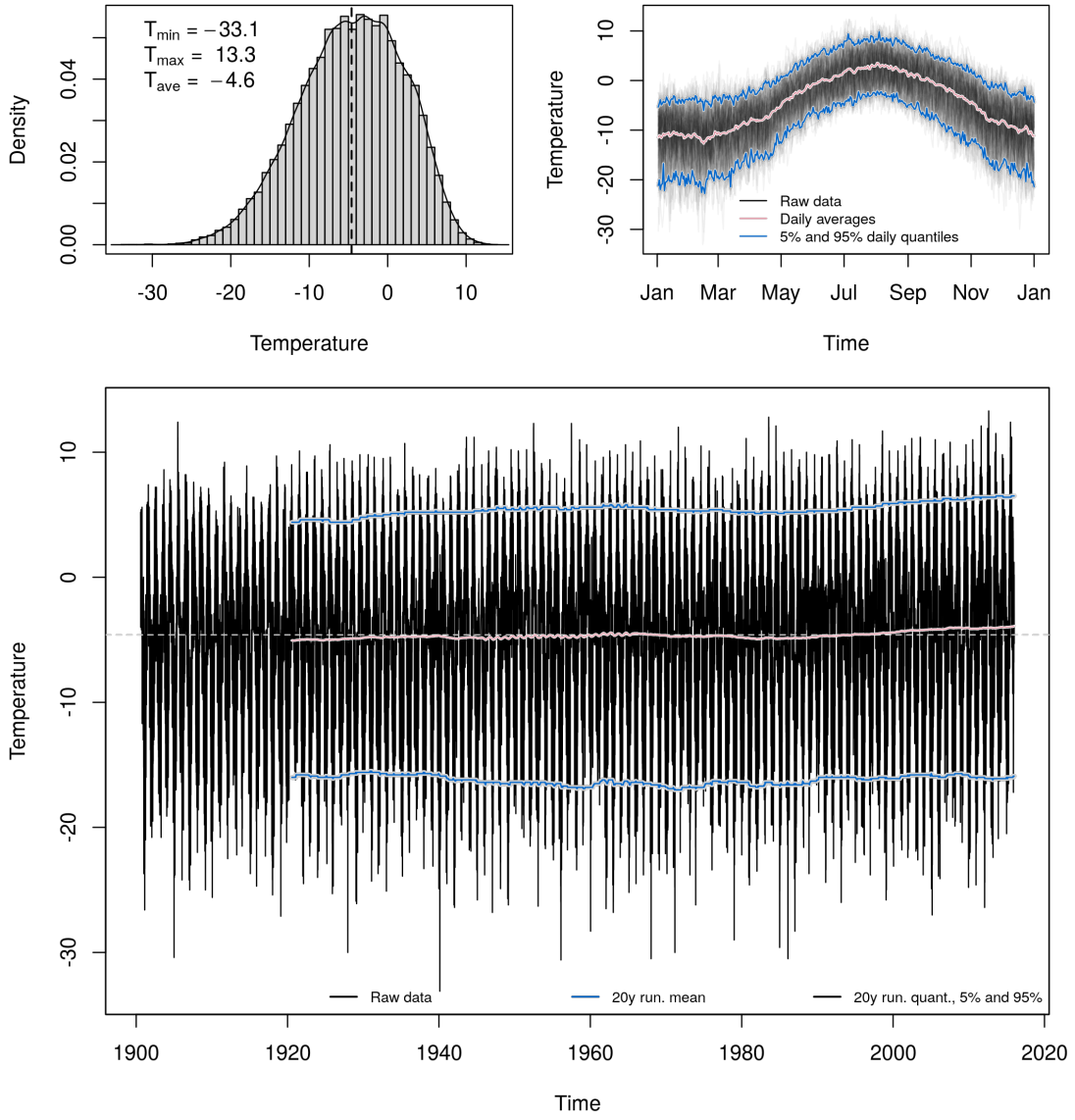

Figure 2: The Zugspitz daily temperature data set.

ments). Extreme precipitation events can lead to strong environmental impacts, therefore, evaluating the amount of maximum precipitation over a certain time period, e.g., the level of daily precipitation sums that can be observed at least once in 100 years (100 year return level), is an important information for environmental risk management. This analysis can be seen as an enhancement of Fischer et al. (2013) which use station-wise models all over Germany applying a generalized extreme value (GEV) distribution on monthly block maxima to compute return levels. Instead of station-wise modeling, the main focus of this analysis is to develop a model for all of Germany and provide a smooth map of 100 year return levels.

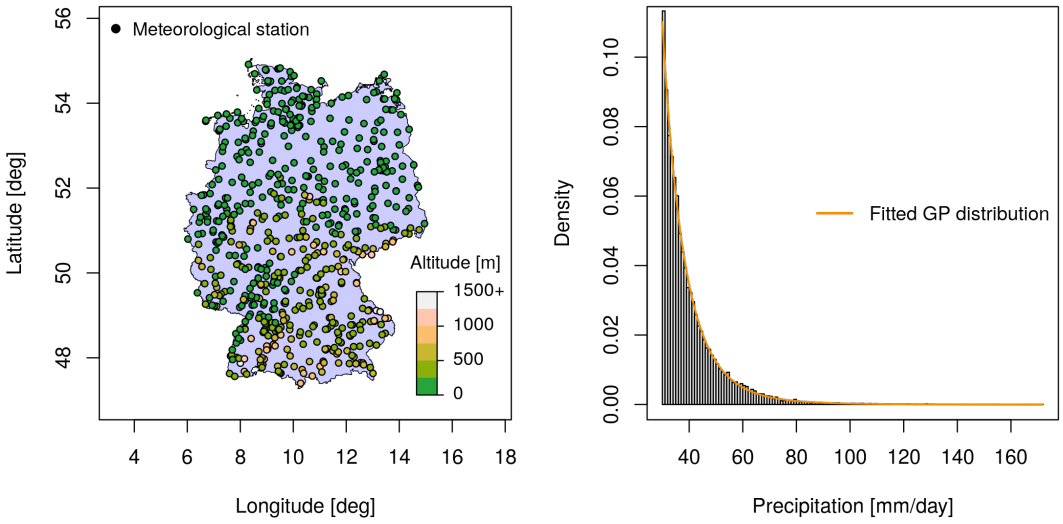

Figure 3: Distribution of available meteorological stations (left panel) fitted generalized Pareto distribution to the daily precipitation observations (right panel).

Similarly to the first example, to account for possible time trends in extreme precipitation it is necessary to use homogenized long term time series. Therefore, we only used observations of meteorological stations with the highest quality, as reported by the quality flags from the DWD. Additionally, we analyzed data from stations that did not change in altitude more than 10 meters, only. The final data set consists of 28634 observations of 640 meteorological stations.

The spatial distribution of the stations is shown in the left panel of Figure 3. A histogram of observations over 30 millimeters of daily precipitation sums is shown in the right panel of Figure 3. A natural distribution for such extreme value data, or peak over threshold (POT) data, is the generalized Pareto (GP) family (see, e.g., Coles, 2001), which has the probability density function

$$f(y;\xi,\sigma,\mu) = \frac{1}{\sigma}\left(1+\frac{\xi(y-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)}.$$

Here, parameter μ is the threshold that is usually specified in advance by, e.g., visual inspection of so called mean residual life plots (Coles, 2001). Parameters ξ and σ are the shape and scale parameters of the distribution. In the right panel of Figure 3, the orange line represents the corresponding estimated density of a simple GP model illustrating the goodness-of-fit to the precipitation data. Note that the mean of the GP distribution is given by

$$\mu + \frac{\sigma}{1-\xi}, \quad \text{for} \quad \xi < 1.$$

Another special interest of this analysis is treating the question if the average amount of extreme precipitation has changed over time. However, as mentioned above the main focus

is to provide a smooth map of Germany with 100 years return levels. The 100 year return levels of daily observations are calculated with

$$\mu + \frac{\sigma}{\xi}(365 \cdot 100 \cdot \lambda_\mu)^\xi - 1,$$

where $\lambda_\mu = \text{Prob}(Y > \mu)$ which is estimated by the empirical threshold exceedance rate. Therefore, this analysis is an example of distributional regression where the interest lies not only on determining the impact of covariates on the parameters of the response distribution but also on quantities derived from transformations of the distributional parameters. This is particularly easy in Bayesian inference based on Markov chain Monte Carlo simulations.

3 Model structure

Most common regression specifications relate the conditional expectation of the response variable of interest y_i to a set of explanatory variables (or covariates) $\mathbf{x}_i$ via

$$E(y_i|\mathbf{x}_i) = h(\eta(\mathbf{x}_i)), \quad i = 1, \ldots, n,$$

where $h(\cdot)$ is a response function that ensures that only admissible values for the expectation can be observed (e.g. ensuring nonnegativity in case of responses with only positive support) while $\eta(\mathbf{x}_i)$ is a regression predictor formed from the covariates (e.g. $\eta(\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$ in case of generalized linear models). Of course, specifying a regression model for the expectation alone provides only an incomplete picture for the distribution of the response. While an additional distributional assumption has to be made for y_i to facilitate maximum likelihood or Bayesian inference that both require the specification of a likelihood, all parameters potentially involved in the likelihood except for the mean are treated as nuisance parameters (for example the error variance in case of a linear model with normal errors) and are, in particular, not related to regression effects.

Distributional regression models overcome this limitation and allow for regression effects on potentially all parameters of the response distribution. More specifically, distributional regression assumes that the conditional distribution is of a specific type (e.g. normal, gamma, etc.) with K parameters and that all parameters are determined in a regression, i.e.

$$y_i|\mathbf{x}_i \sim \mathcal{D}\left(\vartheta_1(\mathbf{x}_i), \ldots, \vartheta_K(\mathbf{x}_i)\right),$$

where $\mathcal{D}$ denotes the parametric distribution for the response variable y_i. In our analysis of the Zugspitz temperature data presented in Section 2, the normal distribution $\texttt{T}_i \sim \text{N}(\mu(\mathbf{x}_i), \sigma(\mathbf{x}_i)^2)$ is a natural starting point since temperatures data do not involve natural limits on the domain but have continuous support over the whole real line. Since we are also allowing for regression effects on the variance parameter $\sigma(\mathbf{x}_i)^2$, the distributional normal model already allows us to account for regression effects beyond the mean in the form of covariate-dependent, heteroscedastic variances. However, as illustrated in the previous paragraph, there is also some empirical indication that the temperature data

exhibit skewness (see Figure 2) which may also change over time such that it makes sense to consider some competing distributional specifications.

We therefore work with two alternative, more flexible distributions with distinguished parameters for skewness and kurtosis. More precisely, we consider the Box-Cox t distribution $\mathtt{T}_i \sim \text{BCT}(\mu(\mathbf{x}_i), \sigma(\mathbf{x}_i), \nu(\mathbf{x}_i), \tau(\mathbf{x}_i))$ and the Box-Cox Power Exponential distribution $\mathtt{T}_i \sim \text{BCPE}(\mu(\mathbf{x}_i), \sigma(\mathbf{x}_i), \nu(\mathbf{x}_i), \tau(\mathbf{x}_i))$ which involve not only a location parameter $\mu(\mathbf{x}_i)$ and a scale parameter $\sigma(\mathbf{x}_i)$ that determine the location and the variability of the distribution but also additional shape parameters $\nu(\mathbf{x}_i)$ and $\tau(\mathbf{x}_i)$ controlling the skewness and kurtosis of the respective distribution. We will later discuss possibilities to compare the fit of these candidate distributions in more detail.

For modeling extreme values of precipitation, we rely on the generalized Pareto distribution $\mathtt{P}_i \sim \text{GP}(\xi(\mathbf{x}_\mathrm{i}), \sigma(\mathbf{x}_\mathrm{i}))$ which naturally arises from extreme value theory based on the peak over threshold approach.

For relating the different distributional parameters $\vartheta_1(\mathbf{x}_i), \ldots, \vartheta_K(\mathbf{x}_i)$ to the covariates, we rely on additive predictors of the form

$$\eta_i^{\vartheta_k} = f_1^{\vartheta_k}(\mathbf{x}_i; \boldsymbol{\beta}_1^{\vartheta_k}) + \ldots + f_{J_k}^{\vartheta_k}(\mathbf{x}_i; \boldsymbol{\beta}_{J_k}^{\vartheta_k}),$$

where the unspecified (possibly nonlinear) functions $f_j^{\vartheta_k}(\cdot)$, $j = 1, \ldots, J_k$ relate to different types of regression effects characterised by regression parameters $\boldsymbol{\beta}_j^{\vartheta_k}$. In analogy to generalized linear models, the predictors are then linked to the distributional parameters via known monotonic and twice differentiable response functions $h_k(\cdot)$ such that

$$\vartheta_k(\mathbf{x}_i) = h_k\left(\eta_i^{\vartheta_k}\right).$$

In the case of our analysis of the daily Zugspitz temperatures, the location and skewness parameters ($\mu(\mathbf{x}_i)$ and $\nu(\mathbf{x}_i)$) are real-valued such that no response function is required. In contrast, the scale parameter $\sigma(\mathbf{x}_i)$ as well as the kurtosis parameter $\tau(\mathbf{x}_i)$ have strictly positive support and we therefore use the exponential response function to ensure this restriction. For simplicity, we treat the kurtosis parameter as a constant ($\tau(\mathbf{x}_i) \equiv \tau$ independently of covariates) while for all other parameters, we consider a predictor of the form

$$\eta_i^\circ = \beta_0^\circ + f_1^\circ(\mathtt{T}_i - \mathtt{T}_{i-1}) + f_2^\circ(\mathtt{yday}_i) + f_3^\circ(\mathtt{year}_i) + f_4^\circ(\mathtt{yday}_i, \mathtt{year}_i), \quad i = 1, \ldots, 42414,$$

where $\circ$ denotes any of the parameters $\mu(\mathbf{x}_i)$, $\sigma(\mathbf{x}_i)$, $\nu(\mathbf{x}_i)$ and $i = 1, \ldots, 42414$ identifies the consecutive days of the observation period. The effects $f_1(\cdot)$ to $f_4(\cdot)$ relate to different types of regression effects chosen to fit with certain stylized features of the data:

- Temperature data are temporally correlated and we therefore include daily temperature differences $\mathtt{T}_i - \mathtt{T}_{i-1}$ in the nonlinear effect $f_1(\cdot)$ to account for this temporal correlation assuming that rather the change in temperature than the absolute value is relevant.

- Due to seasonal variation over the year, we include a cyclic trend component based on the day of the year in $f_2(\cdot)$ which is allowed to vary smoothly over the year but repeats every year while ensuring that the nonlinear effect on December 31 and January 1 are fused together smoothly.
- To account for a potential long term trend in temperatures, we include an overall time trend based on the years from 1954 to 2015 in $f_3(\cdot)$.
- To allow for time-variation in the seasonal component, we furthermore include an interaction between day of the year and year in $f_4(\cdot)$.

In the second example, for modeling extreme values of precipitation, we use exponential response functions for both parameters, $\xi(\mathbf{x}_i)$ and $\sigma(\mathbf{x}_i)$, as well as the predictor

$$\begin{aligned}\eta_i^\circ \quad = \quad & \beta_0^\circ + f_1^\circ(\texttt{alt}) + f_2^\circ(\texttt{year}_i) + f_3^\circ(\texttt{yday}_i) + \\ & f_4^\circ(\texttt{lon}_i, \texttt{lat}_i) + f_5^\circ(\texttt{yday}_i, \texttt{lon}_i, \texttt{lat}_i), \quad i = 1, \ldots, 28634,\end{aligned}$$

where we include the following effects:

- Function $f_1(\cdot)$ reflects the natural, potentially nonlinear effect of altitude on precipitation.
- Similarly to the temperature model, effects $f_2(\cdot)$ and $f_3(\cdot)$ account for a possible long term time trend as well as seasonal trends of the day of the year.
- In addition, to account for spatial variation we include a spatially correlated effect $f_4(\cdot)$ of longitude and latitude coordinates modelled as a two-dimensional smooth surface.
- Moreover, spatial variation might be pronounced for the seasonal trend, therefore, a spatially-varying seasonal effect $f_5(\cdot)$ is included as an interaction.

For all model components, we make use of penalized splines (Eilers and Marx, 2003; Wood, 2003) and their tensor product interactions since they provide a flexible and versatile tool for estimating nonlinear effects also with cyclic constraints. Note, however, that the framework of additive predictors applicable in distributional regression is much broader as indicated by the overview in Table 1.

Estimation of the different model specifications will be based on Markov chain Monte Carlo (MCMC) simulations using iteratively weighted least squares (IWLS) proposals. Since distributional regression models are rather complex including a quite high number of parameters, finding good starting values is important. Therefore, a backfitting algorithm for finding posterior mode estimates is commonly applied before running the MCMC simulation. Details on the implementation of Bayesian inference in distributional regression can be found in Klein et al. (2015d).

Obviously, Bayesian distributional regression involves a number of distinct choices and challenges an analyst is faced with. These include

Covariates	Effect types
Scalar covariates (categorical or continuous)	Linear effect $x \cdot \beta$
	Linear interaction $x_1 \cdot x_2 \cdot \beta$
	Smooth nonlinear effect $f(x)$
	Varying coefficient $f(x_2) \cdot x_1$
	Smooth interaction surface $f(x_1, \ldots, x_L)$
Grouping variable s	Random intercept β_s
	Spatial effect $f(s)$
Grouping and scalar, time variable t	Random slope $x \cdot \beta_s$
	Space-time effect $f(s,t)$
	Functional random intercept $f_s(t)$

Table 1: Commonly used effect types in distributional regression models.

- the choice of a suitable class of potential response distributions,
- the determination of candidate regression predictors for the different distributional parameters, and
- the monitoring and evaluation of Bayesian, simulation-based inference.

In the following, we will provide guidance on these various aspects of performing an analysis based on Bayesian distributional regression.

4 Monitoring and convergence of Markov chains.

Let us start with the situation that a fixed model specification is given and we are estimating the model based on MCMC simulations. In this case, it is important to monitor the mixing and convergence behaviour of the underlying Markov chain to avoid that spurious results are obtained. All methods discussed in the following rely on inspecting the posterior samples either based on graphical visualisation or certain summary statistics. In the following, we assume some familiarity with the principles of Bayesian inference based on MCMC as provided, for example in Green (2001).

The necessity to monitor the convergence of the Markov chain arises from the fact that MCMC simulations are not optimising a certain estimation criterion (such as least squares or the log-likelihood) where convergence implies that the estimate of interest converges to the optimal value but rather we are interested in achieving convergence to the posterior distribution. This unfortunately implies that convergence is way harder to determine in practice although it is guaranteed from a theoretical perspective under mild regularity conditions. Furthermore, since we are generating a Markov chain, the samples will necessarily exhibit some degree of autocorrelation that has to be taken into account. To illustrate different aspects of monitoring Markov chains in MCMC, Figure 4 contains sampling paths as well as autocorrelation functions (ACFs) for three different scenarios.

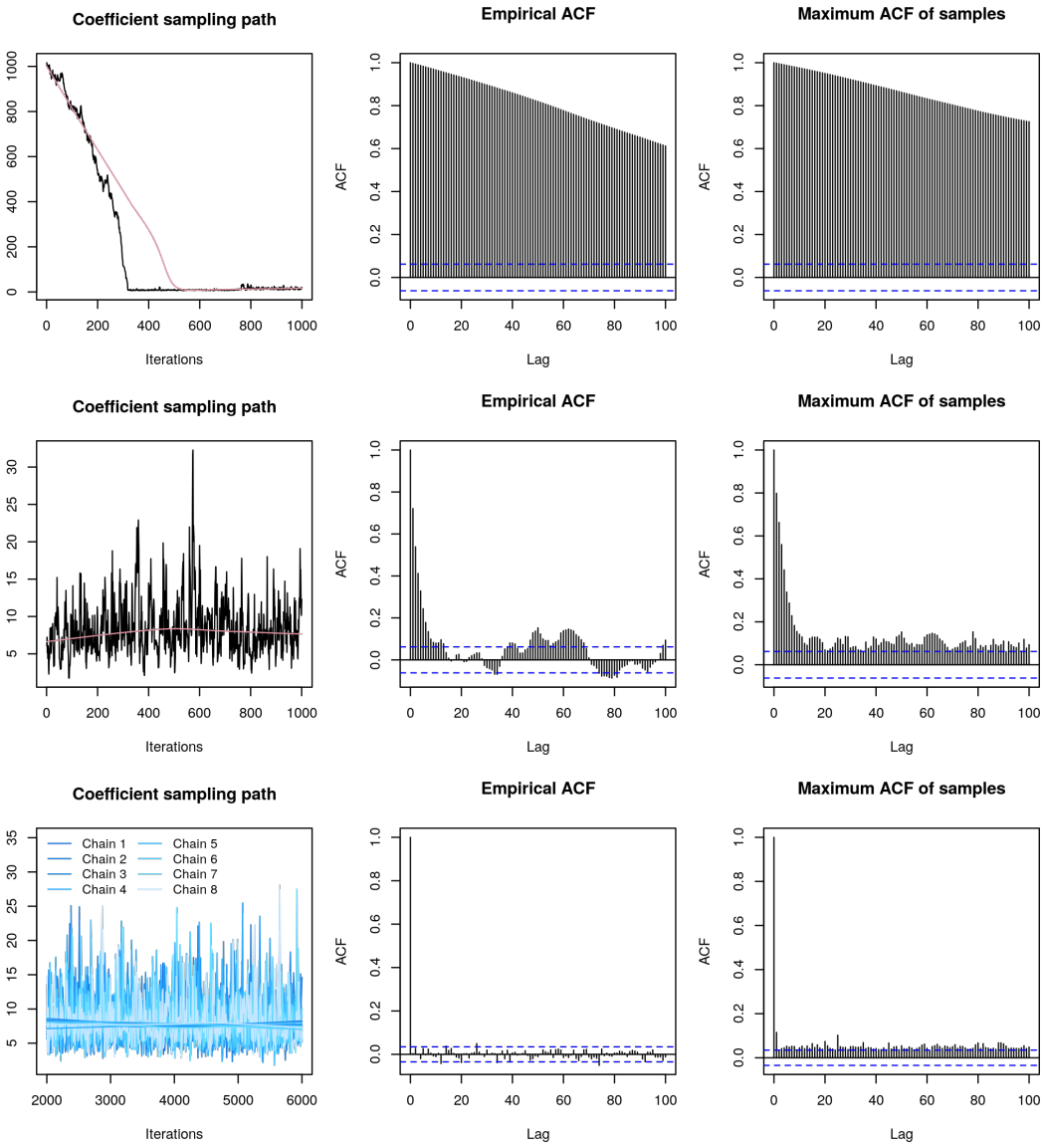

Figure 4: Trace- and ACF plots generated from the Gaussian daily temperature model. The samples of one smoothing variance of model term f_4 of the μ parameter are shown in the first column together with a smooth trend line. The second column shows the corresponding empirical ACF, while the third column visualizes the maximum ACF values over all model parameters. The upper row illustrates the behavior of the Markov chain using poor starting values. The second row shows the effect of using good starting values, e.g., generated by a backfitting algorithm. The bottom row shows samples generated from 8 parallel Markov chains with 6000 samples, a burnin phase of 2000 and thinning of 10 for each chain.

The first row illustrates the behaviour of the Markov chain when starting from a badly chosen starting value. In this case, it takes roughly 300 iterations until the Markov chain has reached the part of the parameter space that is supported by high values of the posterior distribution. This phase is known as the burn in period of MCMC simulations and should be discarded so that results are not impaired by the convergence towards the posterior. Unfortunately, even after the burn in period, the mixing of the Markov chain shown in the first row is far from satisfactory and shows insufficient mixing as demonstrated by the extremely high autocorrelations.

The Markov chain in the second row starts from better initial guesses and therefore only shows a smaller trend towards the posterior and already considerably reduced autocorrelation. Still, the samples can not yet be considered as being independent. The third row contains samples from eight independent chains after removing a burn in period and after thinning the Markov chain to make the samples approximately independent. The Markov chains do not show any indication of a trend, autocorrelations are very low, and repeated runs of the Markov chain produce a similar behaviour. In summary, graphical inspection of the sampling paths and the autocorrelation function provide good guidance on convergence and mixing of the Markov chains with respect to the following aspects: determination of the burn in period, thinning of the Markov chain to make the samples independent and mixing of the Markov chain by comparing different chains.

The major drawback of graphical approaches is that for complex models with hundreds or even thousands of parameters it is hard to monitor all parameters simultaneously. In this case, it can be useful to rely on convergence diagnostics that can be computed from multiple Markov chains produced for the same model (as implemented in the R package **coda**, Plummer et al., 2006). The most common approaches are

- Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992; Brooks and Gelman, 1998),
- Geweke's convergence diagnostic (Geweke, 1992),
- Heidelberger and Welch's convergence diagnostic (Heidelberger and Welch, 1981, 1983).

The Gelman and Rubin's convergence diagnostic is a general approach for monitoring convergence of parallel MCMC chains. Using different starting values, convergence is indicated if all chains approach stationarity within the first half of the samples which is tested by calculating within-chain and between-chain variances. For each variable, if MCMC chains have converged the value of the diagnostic is approximately 1, e.g., the diagnostic for the final BCPE model for daily temperatures indicates convergence with a total range of the diagnostic of $[0.99, 1.01]$ calculated over all single chains.

Similarly, Geweke's diagnostic is testing for convergence based on differences of the mean of the first and last part of the MCMC chains. The test statistic is asymptotically normal distributed, i.e, if MCMC chains have converged differences should be small when quantifying them using a standard Z-score. If the Z-score shows a lack of convergence, it is useful

to successively discard more and more samples until the chain is halfed in order identify the iteration from which convergence is achieved. Again, convergence of the MCMC chains of the final BCPE model is achieved according to a 1% significance level since the range of almost 95% of all Z-values is $[-2.27, 1.94]$ when dividing the chains in half.

Heidelberger and Welch's run length control test is computed by successively discarding samples until either the test is passed or half of the chain has been discarded. In practice, the diagnostic can be used to automatically define an accurate burn-in phase for the MCMC chains. When computing the test for the MCMC chains of the final BCPE model all chains passed the stationarity test.

5 Model selection

We now turn our attention to the actual determination of suitable models for an empirical analysis. A common way can be to proceed along the following steps:

1. Compile a list of potential response distributions: Based on some stylized features of the data and in particular the domain of the observed response values, one can often come up with a sensible list of candidate response distributions. For example, in case of the Zugspitz temperature data, we are interested in continuous data without any domain restrictions and potentially skewed distributions while for the precipitation data the peak over threshold approach implied the generalized Pareto distribution as a suitable model class. Other types of common response structures relate to continuous nonnegative responses (e.g. log-normal or gamma distribution), count data (Poisson, negative binomial, zero inflated Poisson, zero-inflated negative binomial, hurdle models), zero-adjusted responses (log-normal or gamma with an additional spike in zero), fractional responses (beta), etc. A good reference for available distributions is `http://www.gamlss.org/`.

2. Define a candidate predictor: Based on subject-matter considerations about covariates to include and exploratory analyses, define a candidate predictor to be considered for potentially all parameters of the response distribution. This will often be a quite complex model version that defines an upper limit of what the analyst would potentially be willing to consider. Note that very complex models are likely to be a numerical challenge and may lead to difficulties in estimation especially in case of small data sets. The candidate predictor should therefore rather be considered the basis for building sparser models than the ultimate goal for the analysis.

3. Eliminate covariates / candidate distributions: To determine a suitable subset of covariate effects and to evaluate the ability of the candidate distributions to fit the data, one can rely on different criteria such as quantile residuals, information criteria and predictive scoring rules (see below for details). In particular, information criteria can be used to implement different types of stepwise model selection although the

complexity of the model space increases considerably when working with models with multiple predictors.

4. Interpret the results: Once having reduced the list of potential model specifications to a subset of specifications that are compatible with the data, one can interpret the results by identifying both common results and differences between rival specifications. We will return to this issue later based on the temperature data. Note that data-driven model decisions will lead to overly optimistic uncertainty assessments in the final model estimates such that results in such models should be interpreted with care. For confirmatory analyses, model choice steps should therefore be avoided to allow for a meaningful interpretation of significances and the like.

Randomized quantile residuals (Dunn and Smyth, 1996) are based on the probability integral transform that indicates that for a continuous random variable Y with cumulative distribution function F, we have

$$F(Y) \sim U(0,1)$$

i.e. evaluating the cumulative distribution function on the random variable yields a uniform distribution. For discrete random variables, there is a corresponding approximate version. For distributional regression models, one can now replace F with the estimated cumulative distribution function $\hat{F} = F_{\hat{\boldsymbol{\vartheta}}}$ and evaluate it on the observed response values. If the model provides a good approximation of the true data generating mechanism, this should yield a sample of approximately i.i.d. uniform distributed random variables that can be visualized in a probability integral transform histogram. Alternatively, a quantile-quantile-plot can be used to determine deviations of $\Phi^{-1}(F(Y))$ from the standard normal distribution, as demonstrated in Figure 5 for the Zugspitz temperature data. Clearly, the mean of the quantile residuals using the normal distribution slightly deviates from zero and the residuals exhibit some skewness. This is also indicated by the corresponding Q-Q plot where for high and low temperatures the quantiles deviate from the diagonal line. Modeling the skewness and kurtosis parameter seems to improve the model fit, the histogram of the residuals of the Box-Cox t model already seems to be more or less symmetric around zero, however, concerning the Q-Q plots the Box-Cox Power Exponential model is more normal such that the confidence bands computed using all parameter samples almost cover the diagonal over the full range. Therefore, the Box-Cox Power Exponential model is considered as the best model for the daily temperature data

Quantile residuals and the probability integral transform mostly provide a basic check whether a candidate model provides a reasonably flexible description of the observed data but they are not well suited to discriminate between rival model specifications in terms of covariate effects. The latter can, for example, be achieved by Bayesian analogues to frequentist information criteria such as Akaike's information criterion (AIC). The most popular approaches are the deviance information criterion (DIC **?**) and the widely applicable information criterion (WAIC Watanabe, 2010). Both can be easily determined based on the samples for the paramaters of a Bayesian model specification and can be interpreted in complete analogy to the AIC (i.e. smaller values indicate a better fit). However, since

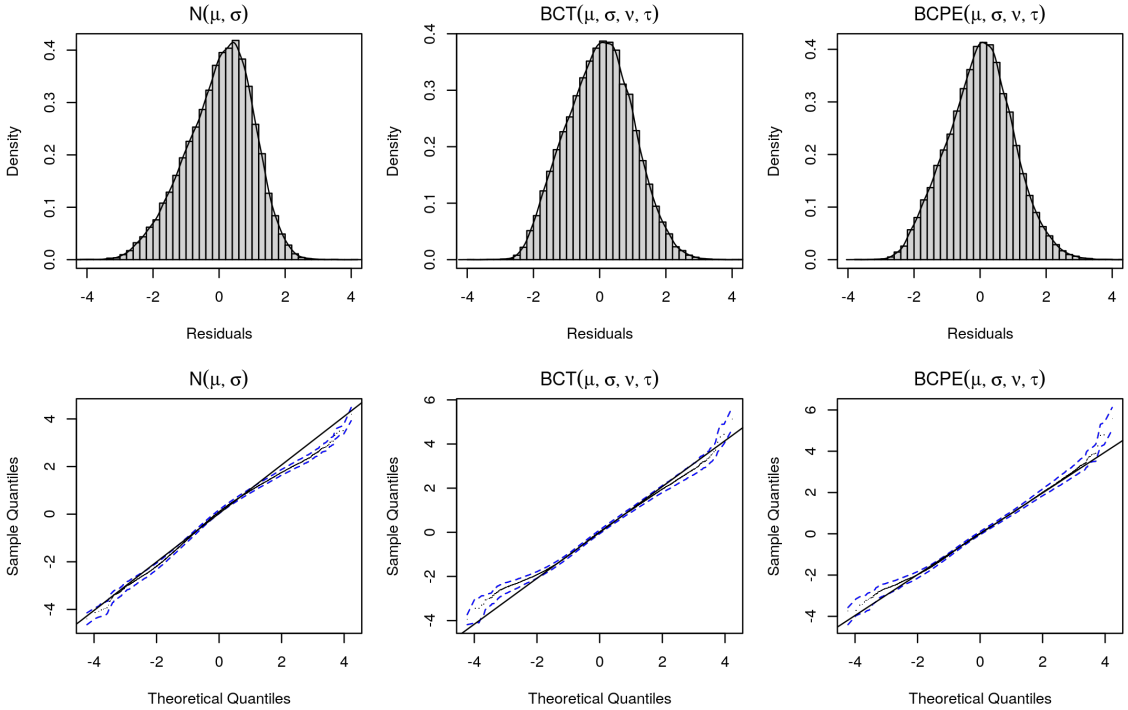

Figure 5: Randomized quantile residuals of the daily temperature models The top row shows histograms of posterior mean quantile residuals together with a kernel density estimate, the bottom row the resulting quantiles vs. the theoretical quantiles of the standard normal distribution. The small black dots represent the posterior mean quantiles of over all posterior samples, the blue dashed lines the corresponding 95% credible bands.

they are estimated from the data, there is inherent sampling variability such that small differences in the information criteria should not be overinterpreted. Table 2 reports the DIC for the three different models for the Zugspitz daily temperature data along with the estimated effective number of parameters (that provide an estimate for the effective dimension of the model). The DIC also clearly favors the Box-Cox-Power-Exponential model over the competing approaches although it is also the model with the largest number of parameters.

When being interested in the predictive ability of distributional regression model, comparing results based on predictive proper scoring rules (Gneiting et al., 2007) can be a valuable alternative. Klein et al. (2015d) provide details on different scores for various types of responses. Basically the goal of all specifications is to evaluate the predictive ability based on rules that favor honest predictions while simultaneously taking sharpness (i.e. precision) of the prediction into account. To obtain a truly predictive assessment, the scores will usually be evaluated based on a cross validation approach which makes their computation numerically demanding compared to information criteria and quantile residuals which can be obtained from one single estimation run.

Model	DIC	pd
$\mathrm{N}(\mu, \sigma^2)$	232426.6	138.7
$\mathrm{BCT}(\mu, \sigma, \nu, \tau)$	231168.1	181.7
$\mathrm{BCPE}(\mu, \sigma, \nu, \tau)$	**230938.6**	189.7

Table 2: Deviance information criterion (DIC) together with estimated effective number of parameters (pd) of the daily temperature models.

6 Empirical Analyses

6.1 Zugspitz daily temperature

From our previous considerations, we found that the Box-Cox-Power-Exponential model provided the best fit for the Zugspitz daily temperature data. Estimated effects of this model are shown in Figure 6. Note that all effects are shown on the scale of the predictor and all estimates are based on all samples of the MCMC chains. Concerning possible shifts of seasonal variability in the first row of Figure 6, the plots indicate that the two-dimensional effect of the day of the year and time for parameter μ and σ is not very pronounced and could be dropped in further analysis. Only for the skewness parameter ν the time-variation of the seasonality seems to be more distinct. For the second and third row, additional to the mean effect the 95% credible intervals are shown by the gray shaded areas. For all distributional parameters, the effect of temperature differences $\mathtt{T}_t - \mathtt{T}_{t-1}$ is the largest. The effect for parameter μ is increasing for negative differences, i.e., if it was warmer at day $t-1$, and is zero for positive differences. The effect for the scale parameter σ is decreasing, i.e., uncertainty is increasing if the weather is getting colder. The effect on the skewness parameter ν has a small positive peak for small differences and is about zero for larger differences. The time-trend effect in the last row of Figure 6 is significantly increasing for parameter μ, especially after the year of 1980. The increase is also significant for parameter σ indicating higher variation of temperatures. Besides, the effect on the skewness parameter ν is not significant, since the 95% credible intervals cover the zero line over the full time period.

A more detailed picture of the temperature time trend is provided in Figure 7. The left panel shows estimated 10%, 50%, 90% and 97.5% temperature quantiles of the Box-Cox Power Exponential model. The quantiles are computed using the quantile function of the Box-Cox Power Exponential distribution by calculating each distributional parameter for each MCMC iteration and take the average over all iterations. Compared to the 20 years empirical running quantiles, the model seems to nicely adapt to the distribution of the data. The middle panel shows the estimated mean temperature trend together with 95% credible intervals. The time trend is significantly increasing with a difference of about 1.6 °C. The right panel of Figure 7 additionally shows the estimated slopes of the 10%, 50%, 90% and 97.5% temperature quantiles. The plot indicates that only for the lowest 10% temperatures the time trend was decreasing from the beginning of the observational period. After 1980 all quantile curves show a positive slope, i.e., regardless of looking at cold or warm days,

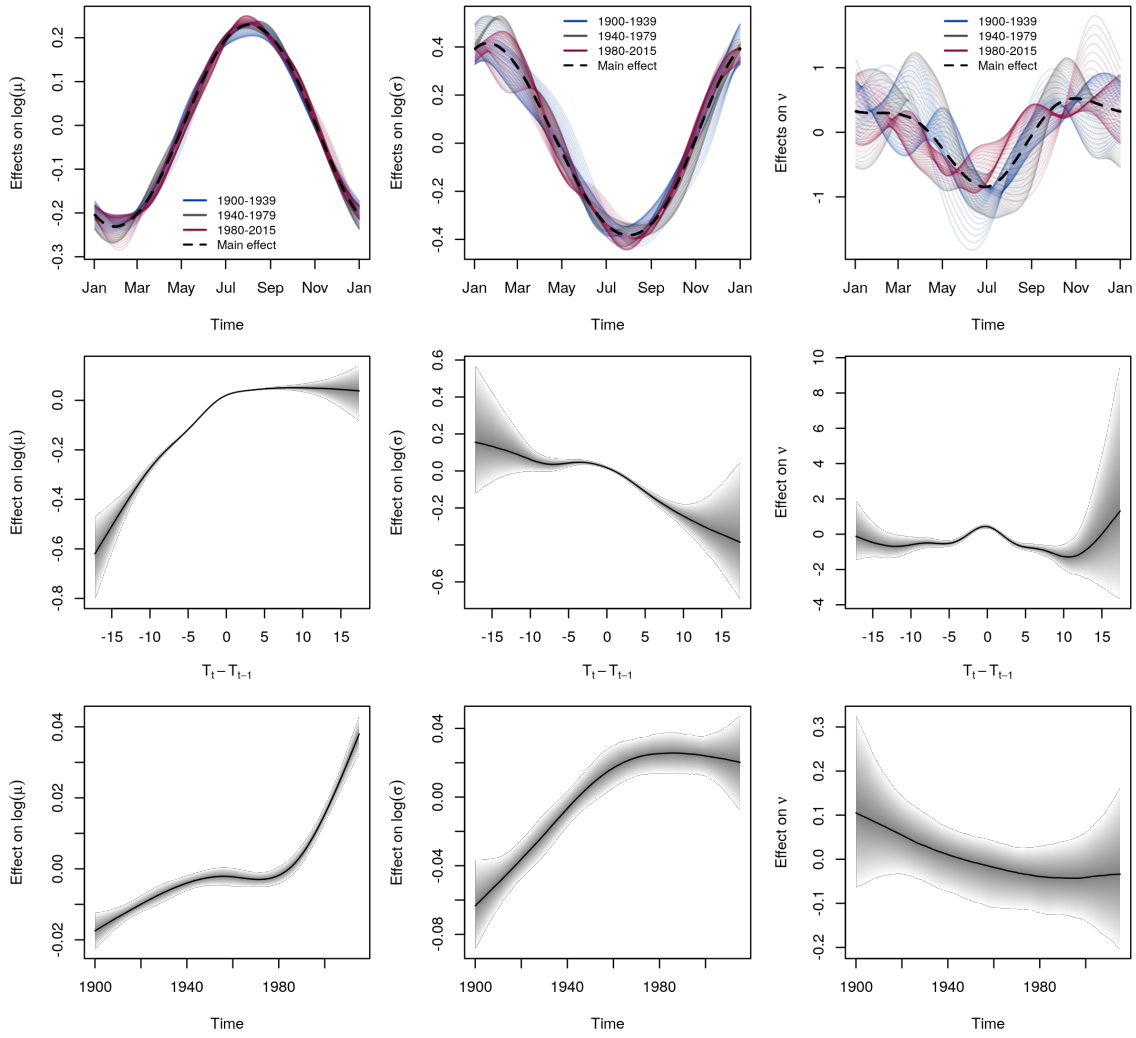

Figure 6: Estimated effects of the daily BCPE temperature model. 95% credible intervals are shown in the middle and bottom row as gray shaded areas.

temperature seems to increase after 1980.

6.2 German extreme precipitation

In this example the focus is on the computation of the mean of the generalized Pareto distribution and the corresponding 100 year return levels. As already mentioned in Section 2.2, these quantities are transformations of the parameters of the distribution. Using the distributional regression approach in combination with MCMC simulation, it is relatively straightforward to provide correct inference for any desired quantity, e.g., the 100 year return levels. This results from the fact that all inferences are sampling based. Once samples for all model parameters are available, we can determine samples also for all derived

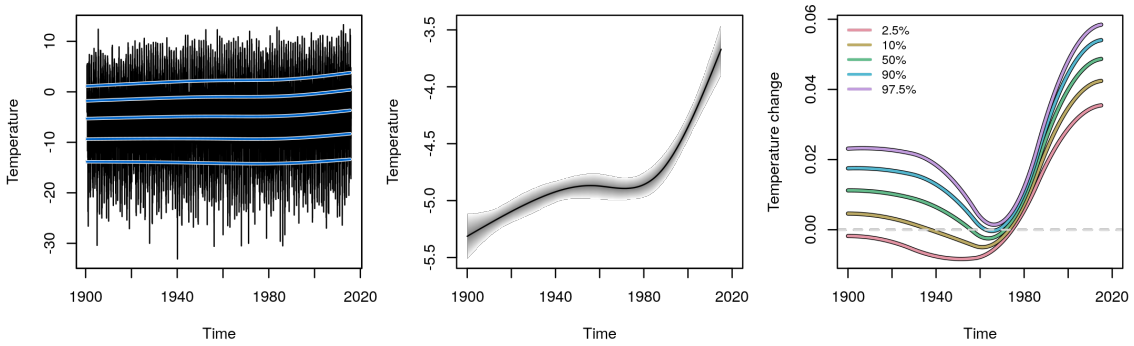

Figure 7: Estimated mean temperature trends of the daily BCPE temperature model. The left panel shows the estimated 2.5%, 10%, 50%, 90% and 97.5% temperature quantiles of the time trend, blue lines, together with the raw data, black lines. The middle panel shows the mean time trend together with 95% credible intervals, gray shaded area. Estimated slopes of corresponding quantiles are shown in the right panel.

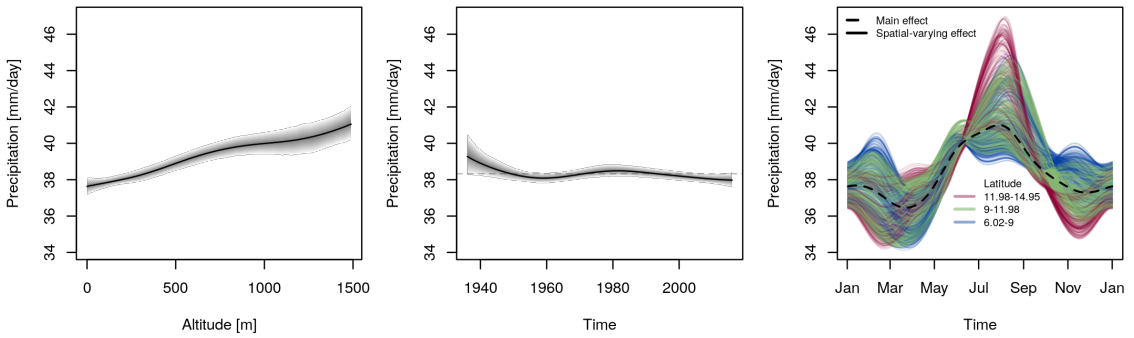

Figure 8: Estimated effects on the mean of the generalized Pareto distribution of the precipitation model.

quantities and base inferences on those.

The estimated effects on the mean of the GP distribution are shown in Figure 8. Note that all effects are visualized using the same scale on the y-axis, such that the comparison of the importance of individual effects is straightforward. The left panel shows the effect of altitude on precipitation. The effect is increasing as expected and basically linear, i.e., a linear instead of a spline representation of the function seems to be sufficient. The estimated time trend in the middle panel shows two maxima around 1940 and 1980, however, no consistent trend can be identified as the estimated function basically follows the estimated average mean precipitation of 40mm. The spatially varying seasonal effect in the right panel has a clear peak in late summer concerning the main effect represented by the black dashed line. Moreover, considerable variation can be identified between north and south of Germany. Some regions in the north obviously have the highest precipitation from Juli to September which are most probably driven by extreme local events. The seasonal effect is also the most important effect on extreme precipitation with the highest variation of about 14mm

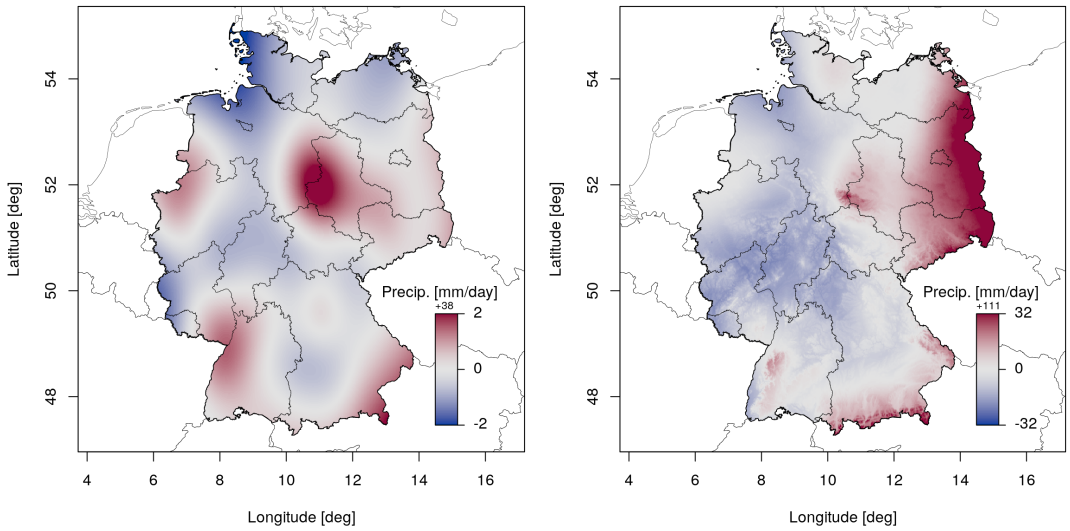

Figure 9: Estimated spatial effect, left panel, estimated maximum 100 year return levels, right panel. The spatial effect is centered around the overall mean of 40mm, the 100 year return levels are centered around the overall mean of 105mm of precipitation.

compared to the altitude effect with 4mm and the time trend with 3mm. The estimated spatial effect is shown in Figure 9. Note that the effect is centered around the estimated mean of 40mm of precipiation, i.e., blue colors represent areas with lower estimated mean and red color areas with higher values. The importance of the effect is similar to the altitude effect with a range of 4mm. The map shows that regions in the south of Germany that are close to the mountains have a positive effect on mean precipitation and, similarly, regions around the highest mountain of central Germany, the Brocken (51°N, 10°E; 1141.2m a.s.l.), and in the east of the Brocken have positive effects. This reflects that regions in the north-east have extreme local events driven by the season as shown in Figure 8, i.e., compared to the north-west where the weather is more continuous. For example, in an extreme event the area of Lindenberg (Brandenburg) 172mm of precipitation where measured in 1978, the highest value in the data set. Finally the map of estimated 100 return levels is shown in the right panel of Figure 8. Initially, the return levels are calculated for each day of the year, afterwards, the maximum over the whole year yields the final return level map. Again note that the return levels are centered around its mean of 111mm of precipitation and that the range of the colorlegend is based on the 97% quantile of estimated return levels. The full range of the maximum 100 year return levels is $[90, 181]$. Mostly because of the altitude effect, the highest return levels can be identified in the mountainous regions in the south of Germany as well as regions around the Brocken in the center and in the north-east.

7 Summary and Conclusions

Distributional regression models as described in this tutorial are a flexible extension of generalized additive models (Wood, 2006; Ruppert et al., 2003; Fahrmeir et al., 2013) that allows to apply flexible, additive predictor structures not only on the mean but on potentially all parameters of even complex response distributions. This idea has been introduced in a penalized likelihood context by Rigby and Stasinopoulos (2005) as generalized additive models for location, scale and shape (GAMLSS) and transferred to a Bayesian context in Klein et al. (2015b,c). We prefer the notion of distributional regression over GAMLSS since in many distributions the parameters are not or only very indirectly related to location or scale of the distribution.

The main competitor to distributional regression in applied regression analyses is quantile regression as also discussed in one of the contributions to this special issue. While the main advantage of quantile regression is that one can avoid an explicit distributional assumption for the response distribution which also alleviates the danger of model misspecification. On the other hand, quantile regression requires strictly continuous responses while distributional regression provides a convenient framework for continuous, discrete and mixed discrete-continuous distributions. Furthermore, quantile regression typically faces the risk of quantile crossing while distributional regression provides a coherent estimate for the complete distribution of the response. A final advantage of distributional regression is that it also facilitates the consideration of multivariate response structures (Klein et al., 2015a; Klein and Kneib, 2016).

Considering Bayesian approaches to distributional regression based on MCMC simulations is particularly attractive not only due to its flexibility, but due to the virtues of sampling based inference. MCMC provides samples for all parameters of the distributional regression model which can be transferred to samples for any quantity of interest that depends on these parameters. In our application on extreme precipitation values, one such quantity were the 100 year return levels. Based on samples from these return levels, one can determine point estimates as well as credible intervals without requiring asymptotic considerations.

A particular challenge of distributional regression is the interpretation of results, which often can no longer be achieved in a ceteris paribus type fashion. As a consequence, designing scenarios of interest for which covariate effects can be evaluated in combination with carefully chosen visualisation tools is extremely important. In the future, we will work on making such graphical tools more broably available and easily accessible.

References

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**(4), 434–455. doi: 10.1080/10618600.1998.10474787.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, volume Springer series in statistics. Springer.

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236–245. doi: 10.2307/1390802.

Eilers, P. H. C. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, **66**(2), 159–174. ISSN 0169-7439. doi: 10.1016/s0169-7439(03)00029-7.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression - Models, Methods and Applications*. Springer, Berlin.

Fischer, M., Rust, H. W., and Ulbrich, U. (2013). Seasonal extreme value statistics for precipitation in germany. In *Proceedings of the 1st International Short Conference on Advances in Extreme Value Analysis and Application to Natural Hazards (EVAN 2013)*, pages 144–152. URL https://www.uni-siegen.de/evan2013/home/proceedings/evan2013_proceedings.pdf.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.*, **7**(4), 457–472. doi: 10.1214/ss/1177011136.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, J. F. M., editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, Oxford.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, **69**(2), 243–268. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2007.00587.x.

Green, P. J. (2001). A primer in Markov Chain Monte Carlo. In Barndorff-Nielsen, O. E., Cox, D. R., and Klüppelberg, C., editors, *Complex Stochastic Systems*, pages 1–62. Chapman and Hall.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall/CRC.

Heidelberger, P. and Welch, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Commun. ACM*, **24**(4), 233–245. ISSN 0001-0782. doi: 10.1145/358598.358630.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, **31**(6), 1109–1144. ISSN 0030364X, 15265463. URL http://www.jstor.org/stable/170841.

Klein, N. and Kneib, T. (2016). Simultaneous inference in structured additive conditional copula regression models: A unifying Bayesian approach. *Statistics and Computing*, **26**, 841–860.

Klein, N., Kneib, T., Klasen, S., and Lang, S. (2015a). Bayesian structured additive distributional regression for multivariate responses. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **64**, 569–591. doi/10.1111/rssc.12090.

Klein, N., Kneib, T., and Lang, S. (2015b). Bayesian generalized additive models for location, scale and shape for zero-inflated and overdispersed count data. *Journal of the American Statistical Association*, **110**, 405–419. doi:10.1080/01621459.2014.912955.

Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015c). Bayesian structured additive distributional regression with with an application to regional income inequality in Germany. *Annals of Applied Statistics*, **9**, 1024–1052.

Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015d). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *Annals of Applied Statistics*, **9**, 1024–1052. doi: 10.1214/15-aoas823.

Koenker, R. (2005). *Quantile Regression*. Cambrigde University Press, New York. Economic Society Monographs.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). **coda**: Convergence diagnosis and output analysis for MCMC. *R News*, **6**(1), 7–11. doi: 10.18637/jss.v021.i11.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. ISBN 3-900051-07-0.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **54**, 507–554.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Umlauf, N., Klein, N., Zeileis, A., and Köhler, M. (2017). ***bamlss**: Bayesian Additive Models for Location Scale and Shape (and Beyond)*. URL http://CRAN.R-project.org/package=bamlss. R package version 0.1-2.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, **11**, 3571–3594.

Wood, S. N. (2006). *Generalized Additive Models : An Introduction with R*. Chapman & Hall/CRC, New York/Boca Raton.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(1), 95–114. ISSN 1467-9868. doi: 10.1111/1467-9868.00374.

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2017-13 **Thomas Kneib, Nikolaus Umlauf:** A Primer on Bayesian Distributional Regression

2017-12 **Susanne Berger, Nathaniel Graham, Achim Zeileis:** Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R

2017-11 **Natalia Danzer, Martin Halla, Nicole Schneeweis, Martina Zweimüller:** Parental leave, (in)formal childcare and long-term child outcomes

2017-10 **Daniel Muller, Sander Renes:** Fairness views and political preferences - Evidence from a large online experiment

2017-09 **Andreas Exenberger:** The Logic of Inequality Extraction: An Application to Gini and Top Incomes Data

2017-08 **Sibylle Puntscher, Duc Tran Huy, Janette Walde, Ulrike Tappeiner, Gottfried Tappeiner:** The acceptance of a protected area and the benefits of sustainable tourism: In search of the weak link in their relationship

2017-07 **Helena Fornwagner:** Incentives to lose revisited: The NHL and its tournament incentives

2017-06 **Loukas Balafoutas, Simon Czermak, Marc Eulerich, Helena Fornwagner:** Incentives for dishonesty: An experimental study with internal auditors

2017-05 **Nikolaus Umlauf, Nadja Klein, Achim Zeileis:** BAMLSS: Bayesian additive models for location, scale and shape (and beyond)

2017-04 **Martin Halla, Susanne Pech, Martina Zweimüller:** The effect of statutory sick-pay on workers' labor supply and subsequent health

2017-03 **Franz Buscha, Daniel Müller, Lionel Page:** Can a common currency foster a shared social identity across different nations? The case of the Euro.

2017-02 **Daniel Müller:** The anatomy of distributional preferences with group identity

2017-01 **Wolfgang Frimmel, Martin Halla, Jörg Paetzold:** The intergenerational causal effect of tax evasion: Evidence from the commuter tax allowance in Austria

2016-33 **Alexander Razen, Stefan Lang, Judith Santer:** Estimation of spatially correlated random scaling factors based on Markov random field priors

2016-32 **Meike Köhler, Nikolaus Umlauf, Andreas Beyerlein, Christiane Winkler, Anette-Gabriele Ziegler, Sonja Greven:** Flexible Bayesian additive joint models with an application to type 1 diabetes research

2016-31 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Simultaneous ensemble post-processing for multiple lead times with standardized anomalies

2016-30 **Alexander Razen, Stefan Lang:** Random scaling factors in Bayesian distributional regression models with an application to real estate data

2016-29 **Glenn Dutcher, Daniela Glätzle-Rützler, Dmitry Ryvkin:** Don't hate the player, hate the game: Uncovering the foundations of cheating in contests

2016-28 **Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

2016-27 **Michael Razen, Matthias Stefan:** Greed: Taking a deadly sin to the lab

2016-26 **Florian Wickelmaier, Achim Zeileis:** Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

2016-25 **Michel Philipp, Carolin Strobl, Jimmy de la Torre, Achim Zeileis:** On the estimation of standard errors in cognitive diagnosis models

2016-24 **Florian Lindner, Julia Rose:** No need for more time: Intertemporal allocation decisions under time pressure

2016-23 **Christoph Eder, Martin Halla:** The long-lasting shadow of the allied occupation of Austria on its spatial equilibrium

2016-22 **Christoph Eder:** Missing men: World War II casualties and structural change

2016-21 **Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies *published in Monthly Weather Review*

2016-20 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Incentive schemes, private information and the double-edged role of competition for agents

2016-19 **Martin Geiger, Richard Hule:** Correlation and coordination risk

2016-18 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

2016-17 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Guilt-averse or reciprocal? Looking at behavioural motivations in the trust game

2016-16 **Esther Blanco, Tobias Haller, James M. Walker:** Provision of public goods: Unconditional and conditional donations from outsiders

2016-15 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Predictive bookmaker consensus model for the UEFA Euro 2016

2016-14 **Martin Halla, Harald Mayr, Gerald J. Pruckner, Pilar García-Gómez:** Cutting fertility? The effect of Cesarean deliveries on subsequent fertility and maternal labor supply

2016-13 **Wolfgang Frimmel, Martin Halla, Rudolf Winter-Ebmer:** How does parental divorce affect children's long-term outcomes?

2016-12 **Michael Kirchler, Stefan Palan:** Immaterial and monetary gifts in economic transactions. Evidence from the field

2016-11 **Michel Philipp, Achim Zeileis, Carolin Strobl:** A toolkit for stability assessment of tree-based learners

2016-10 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** Affirmative action or just discrimination? A study on the endogenous emergence of quotas *published in Journal of Economic Behavior and Organization*

2016-09 **Loukas Balafoutas, Helena Fornwagner:** The limits of guilt

2016-08 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Spatial ensemble post-processing with standardized anomalies

2016-07 **Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model

2016-06 **Michael Razen, Jürgen Huber, Michael Kirchler:** Cash inflow and trading horizon in asset markets

2016-05 **Ting Wang, Carolin Strobl, Achim Zeileis, Edgar C. Merkle:** Score-based tests of differential item functioning in the two-parameter model

2016-04 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Non-homogeneous boosting for predictor selection in ensemble post-processing

2016-03 **Dietmar Fehr, Matthias Sutter:** Gossip and the efficiency of interactions

2016-02 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Rankings and risk-taking in the finance industry

2016-01 **Sibylle Puntscher, Janette Walde, Gottfried Tappeiner:** Do methodical traps lead to wrong development strategies for welfare? A multilevel approach considering heterogeneity across industrialized and developing countries

University of Innsbruck

Working Papers in Economics and Statistics

2017-13

Thomas Kneib, Nikolaus Umlauf

A primer on bayesian distributional regression

Abstract
Bayesian methods have become increasingly popular in the past two decades. With the constant rise of computational power even very complex models can be estimated on virtually any modern computer. Moreover, interest has shifted from conditional mean models to probabilistic distributional models capturing location, scale, shape and other aspects of a response distribution, where covariate effects can have flexible forms, e.g., linear, nonlinear, spatial or random effects. This tutorial paper discusses how to select models in the Bayesian distributional regression setting, how to monitor convergence of the Markov chains, evaluate relevance of effects using simultaneous credible intervals and how to use simulation-based inference also for quantities derived from the original model parameterisation. We exemplify the work flow using daily weather data on (i) temperatures on Germany's highest mountain and (ii) extreme values of precipitation all over Germany.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)