Extending extended logistic regression for ensemble post-processing: Extended vs. separate vs. ordered vs. censored

Jakob W. Messner, Georg J. Mayr,
Daniel S. Wilks, Achim Zeileis

Working Papers in Economics and Statistics

2013-32

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 Economics
- Department of Public Finance
- Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
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.

Extending Extended Logistic Regression for Ensemble Post-Processing: Extended vs. Separate vs. Ordered vs. Censored

Jakob W. Messner
Universität Innsbruck

Georg J. Mayr
Universität Innsbruck

Daniel S. Wilks
Cornell University

Achim Zeileis
Universität Innsbruck

Abstract

Extended logistic regression is a recent ensemble calibration method that extends logistic regression to provide full continuous probability distribution forecasts. It assumes conditional logistic distributions for the (transformed) predictand and fits these using selected predictand category probabilities. In this study we compare extended logistic regression to the closely related ordered and censored logistic regression models. Ordered logistic regression avoids the logistic distribution assumption but does not yield full probability distribution forecasts, whereas censored regression directly fits the full conditional predictive distributions.

To compare the performance of these and other ensemble post-processing methods we used wind speed and precipitation data from two European locations and ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). Ordered logistic regression performed similarly to extended logistic regression for probability forecasts of discrete categories whereas full predictive distributions were better predicted by censored regression.

Keywords: probabilistic forecasting, extended logistic regression, ordered logistic regression, heteroscedasticity.

1. Introduction

Important applications such as severe weather warnings or decision making in agriculture, industry, and finance strongly demand accurate weather forecasts. Usually numerical weather prediction (NWP) models are used to provide these weather forecasts. Unfortunately, because of the only roughly known current state of the atmosphere and unknown or unresolved physical processes, NWP models are always subject to error. To estimate these errors many forecasting centers nowadays provide ensemble forecasts. These are several NWP forecasts with perturbed initial conditions and/or different model formulations. However, the perturbed initial conditions do not necessarily represent initial condition uncertainty (Hamill, Snyder, and Whitaker 2003; Wang and Bishop 2003) and some structural deficiencies in the models are also not accounted for. Thus, the ensemble forecasts usually do not represent the full uncertainty of NWP models. Ensemble forecasts therefore typically need to be statistically post-processed to achieve well-calibrated probabilistic forecasts.

In the past decade a variety of different ensemble post-processing methods have been proposed. Examples are ensemble dressing (Roulston and Smith 2003), Bayesian model averaging (Raftery, Gneiting, Balabdaoui, and Polakowski 2005), heteroscedastic linear regression (Gneiting, Raftery, Westveld, and Goldman 2005), or logistic regression (Hamill, Whitaker, and Wei 2004). Comparisons of these and other post-processing methods (Wilks 2006a; Wilks and Hamill 2007) showed that logistic regression performs relatively well. Recently, Wilks (2009) extended logistic regression by including the (transformed) predictand thresholds as an additional predictor variable. In addition to requiring fewer coefficients and providing coherent probabilistic forecasts this extended logistic regression allows derivation of full continuous predictive distributions. Extended logistic regression has been used frequently (Schmeits and Kok 2010; Ruiz and Saulo 2012; Roulin and Vannitsem 2012; Hamill 2012; Ben Bouallègue 2013; Scheuerer 2013; Messner, Zeileis, Mayr, and Wilks 2013) and has been further extended to additionally account for conditional heteroscedasticy (Messner *et al.* 2013). Recently, several studies noticed that extended logistic regression assumes a conditional logistic distribution for the transformed predictand (Scheuerer 2013; Schefzik, Thorarinsdottir, and Gneiting 2013; Messner *et al.* 2013) where this logistic distribution is fitted to selected predictand category probabilities.

In this study we compare (heteroscedastic) extended logistic regression with two closely related regression models from statistics that are particularly popular in econometrics (and more broadly the social sciences):

1. (Heteroscedastic) ordered logistic regression also provides coherent forecasts of category probabilities. However it differs from extended logistic regression in that no continuous distribution is assumed or specified by the model.

2. (Heteroscedastic) censored regression also fits conditional logistic distributions to a transformed predictand but employs the full set of training-data points (as opposed to a set of thresholds) for fitting the model.

The performance of these statistical models is tested on wind speed and precipitation data from two European locations and ensemble forecasts from the European Centre for Medium Range Weather Forecasts (ECMWF). In addition to heteroscedastic ordered logistic regression, heteroscedastic extended logistic regression, and heteroscedastic censored logistic regression, also separate logistic regressions (Hamill *et al.* 2004) and for wind speed forecasts heteroscedastic truncated Gaussian regression (Thorarinsdottir and Gneiting 2010) are tested.

The following Section 2 describes the different statistical models in detail. A brief description of the data can be found in Section 3. Finally, Section 4 presents the results that are summarized and discussed in Section 5.

2. Statistical models

This section describes different statistical models to predict conditional probabilities $P(y \leq q_j|\mathbf{x})$ of a continuous predictand y falling below a threshold q_j, given a vector of predictor variables $\mathbf{x} = (1, x_1, x_2, \dots)^\top$ (i.e., NWP forecasts). Conditional category probabilities of y

to fall between two thresholds q_a and q_b can then easily be derived with $P(q_a < y \leq q_b) = P(y \leq q_b|\mathbf{x}) - P(y \leq q_a|\mathbf{x})$.

2.1. Separate logistic regressions (SLR)

Logistic regression was one of the first statistical methods that were proposed to post-process ensemble forecasts (Hamill *et al.* 2004). Originally it is a regression model from the generalized linear model framework (Nelder and Wedderburn 1972) to model the probability of binary responses:

$$P(y \leq q_j|\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \beta)}{1 + \exp(\mathbf{x}^\top \beta)} = \Lambda(\mathbf{x}^\top \beta) \tag{1}$$

where $\beta = (\beta_0, \beta_1, \beta_2, \dots)^\top$ is a coefficient vector and $\Lambda(\cdot) = \exp(\cdot)/(1+\exp(\cdot))$ is notationally equivalent to the cumulative distribution function of the standard logistic distribution. The coefficient vector β is estimated by maximizing the the log-likelihood

$$\ell = \sum_{i=1}^{N} \log(\pi_i) \tag{2}$$

as a function of β as defined in Equation 1, where N is the number of events in the data set and π_i is the predicted probability of the i-th observed outcome:

$$\pi_i = \begin{cases} P(y_i \leq q_j|\mathbf{x}_i) & y_i \leq q_j \\ 1 - P(y_i \leq q_j|\mathbf{x}_i) & y_i > q_j \end{cases} \tag{3}$$

Often separate logistic regressions (i.e., with separate coefficient vectors β) are fitted for several thresholds q_j of interest (e.g., Hamill *et al.* 2004; Wilks 2006a; Wilks and Hamill 2007). This implies that the regression lines for different thresholds can cross, so that for some values of the predictor variables $\mathbf{x}$, $P(y \leq q_a|\mathbf{x}) > P(y \leq q_b|\mathbf{x})$ although $q_a < q_b$ which leads to nonsense negative probability for y to fall between q_a and q_b.

2.2. Heteroscedastic extended logistic regression (HXLR)

To avoid these negative probabilities and to reduce the number of regression coefficients Wilks (2009) proposed to include a transformation of the predictand thresholds as an additional predictor variable in logistic regression.

$$P(y \leq q_j|\mathbf{x}) = \Lambda(\alpha g(q_j) - \mathbf{x}^\top \beta) \tag{4}$$

where α is an additional coefficient that has to be estimated and the transformation $g()$ is a non-decreasing function. Equation 4 also differs from standard logistic regression, where β is estimated separately for each threshold, in that here β is the same for all thresholds. Thus, one interpretation of Equation 4 is that it defines parallel regression lines (in log-odds space) with equal slope but different intercepts ($\theta_j = \alpha g(q_j) - \beta_0$). Figure 1 shows examples of these regression lines schematically.

Extended logistic regression not only avoids the problem of crossing regression lines but also allows for computing probabilities for any threshold value q_j (and not only the the thresholds employed for estimating the model). In other words, Equation 4 can also be interpreted as a cumulative distribution function that describes a full continuous predictive distribution. After some reformulation (see Messner *et al.* 2013), Equation 4 can also be written as

$$P(g(y) \leq g(q_j)|\mathbf{x}) = P(y \leq q_j|\mathbf{x}) = \Lambda\left(\frac{g(q_j) - \mathbf{x}^\top\beta/\alpha}{1/\alpha}\right) \tag{5}$$

which shows that the predictive distribution of the transformed predictand $g(y)$ is a logistic distribution with location parameter $\mathbf{x}^\top\beta/\alpha$ and scale parameter $1/\alpha$. Thus, the transformation $g()$ must be chosen such that the transformed predictand can be assumed to follow a conditional (on the predictors $\mathbf{x}$) logistic distribution.

To effectively utilize uncertainty information contained in the ensemble spread, Messner *et al.* (2013) proposed to use additional predictor variables ($\mathbf{z} = (1, z_1, z_2, \dots)^\top$; e.g., the ensemble spread) to directly control the dispersion (variance) of the logistic predictive distribution:

$$P(y \leq q_j|\mathbf{x}) = \Lambda\left(\frac{g(q_j) - \mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right) \tag{6}$$

where $\gamma = (1, \gamma_1, \gamma_2, \dots)^\top$ and $\delta = (1, \delta_1, \delta_2, \dots)^\top$ are the coefficient vectors that have to be estimated.

These coefficient vectors γ and δ are also estimated by maximizing the log-likelihood function given by Equation 2. However, the probability of the observed outcome for the multi-categorical predictand is

$$\pi_i = \begin{cases} P(y_i \leq q_1|\mathbf{x}_i) & y_i \leq q_1 \\ P(y_i \leq q_j|\mathbf{x}_i) - P(y_i \leq q_{j-1}|\mathbf{x}_i) & q_{j-1} < y_i \leq q_j \\ 1 - P(y_i \leq q_J|\mathbf{x}_i) & y_i > q_J \end{cases} \tag{7}$$

(Messner *et al.* 2013) where J is the number of thresholds q_j that have been selected for the fitting calculation.

2.3. Heteroscedastic ordered logistic regression (HOLR)

Ordered logistic regression – also known as ordered logit, proportional odds logistic regression, or cumulative link model – is a popular regression model from statistics and econometrics for ordinal data, which has not received much attention in meteorology so far. Like extended logistic regression it is an extension of standard logistic regression for multi-categorical and ordered predictands. Different to extended logistic regression, separate intercepts θ_j are fitted for each selected threshold instead of modeling them as a linear function of the (transformed) thresholds.

$$P(y \leq q_j|\mathbf{x}) = \Lambda(\theta_j - \mathbf{x}^\top\beta) \tag{8}$$

where the estimated separate intercepts θ_j are only constrained to be ordered ($\theta_1 \leq \theta_2 \leq \dots \leq \theta_J$). Because the intercepts of the regression lines are fully determined by θ_j further intercepts are not needed anymore so that $\mathbf{x} = (x_1, x_2, \dots)^\top$ must not contain any constant.

The separate intercepts for each threshold imply the estimation of more coefficients than for extended logistic regression. Furthermore only the probabilities for the thresholds q_j employed in the estimation can be derived, so that Equation 8 does not specify full continuous predictive distributions. In return, ordered logistic regression does not assume a continuous distribution

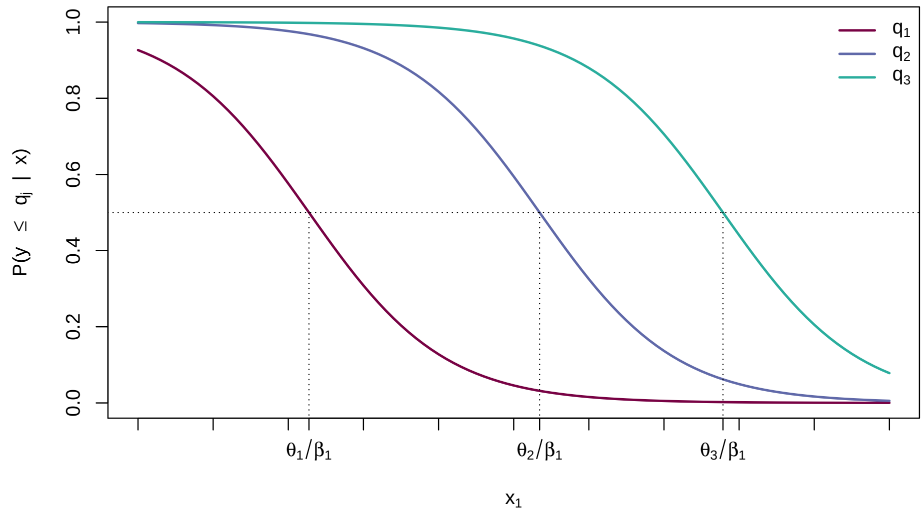

Figure 1: Schematic figure of regression lines fitted with extended, ordered, or censored logistic regression with one predictor variable x_1 and $J = 3$ thresholds (q_1, q_2, q_3).

for the transformed predictand. Thus no (possibly non-existent) transformation has to be determined to fulfill this assumption.

Similar to heteroscedastic extended logistic regression, a heteroscedastic version of ordered logistic regression also allows control of the scale (variance) of an underlying latent distribution with additional predictor variables (Agresti 2002).

$$P(y \leq q_j|\mathbf{x}) = \Lambda\left(\frac{\theta_j - \mathbf{x}^\top\beta}{\exp(\mathbf{z}^\top\delta)}\right) \tag{9}$$

Note that here also no constant is needed in $\mathbf{z} = (z_1, z_2, \dots)^\top$.

Maximum likelihood estimation with the same log-likelihood function as for extended logistic regression (Equations 2 and 7) is used to estimate the coefficients θ_j, β, and δ.

2.4. Heteroscedastic censored logistic regression (HCLR)

Above we have shown that extended logistic regression assumes a conditional logistic distribution for the transformed predictand. The maximum likelihood estimation with the log-likelihood function given by Equations 2 and 7 fits the selected category probabilities. However, if the predictand is given in continuous form, the model described by Equation 6 can also be estimated with the log-likelihood function from Equation 2 with

$$\pi_i = \lambda\left(\frac{g(y_i) - \mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right) = \frac{\exp\left(-\frac{g(y_i)-\mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right)}{\exp(\mathbf{z}^\top\delta)\left[1 + \exp\left(-\frac{g(y_i)-\mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right)\right]^2} \tag{10}$$

where $\lambda(\cdot)$ denotes the likelihood function of the standard logistic distribution. The likelihood is notationally identical to the probability density function (i.e., the derivative of Equation 6 with respect to $g(q_j)$), but differs because it is a function of the parameter vectors γ and δ

Model	*SLR*	*(H)OLR*	*(H)XLR*	*(H)CLR*
Type	Separate	Ordered	Extended	Censored
Intercepts	Unconstrained	Ordered	Lin. fun. of $g(q)$	Lin. fun. of $g(q)$
Slopes	Separate	Joint	Joint	Joint
Number of parameters	KJ	$K+J$	$K+2$	$K+2$
Implies cont. distribution	No	No	Yes	Yes
Estimation based on	Thresholds	Thresholds	Thresholds	Cont. distribution
Heteroscedasticity	No	Yes (optional)	Yes (optional)	Yes (optional)

Table 1: Overview over the different logistic regression models with respect to their parametrization and the likelihood. K is the number of predictor variables $(x_1, x_2, \dots, z_1, z_2, \dots)$ and J is the number of thresholds q_j.

for a fixed predictand value y_i, rather than being a function of y_i given fixed values for γ and δ. In this way, the π_i employed for fitting the model are not the likelihoods for predictands falling into discrete intervals, but rather the likelihoods that they take on their exact observed values. This model can also be interpreted as a linear regression model with a (heteroscedastic) logistic error distribution.

Non-negative variables, e.g., wind speeds or precipitation amounts, are only continuous for positive values and have a natural threshold at 0. This non-negativity can easily accommodated using censored regression (first discussed by Tobin 1958, for the Gaussian case) where the π_i are replaced by

$$\pi_i = \begin{cases} \Lambda\left(\frac{g(0)-\mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right) & y_i = 0 \\ \lambda\left(\frac{g(y_i)-\mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right) & y_i > 0 \end{cases} \tag{11}$$

in Equation 2.

This heteroscedastic censored logistic regression fits a logistic error distribution with point mass at zero to the transformed predictand. While such an error distribution seems reasonable for square root transformed precipitation amounts (Scheuerer 2013; Schefzik *et al.* 2013), usually other error distributions are assumed for wind speed. For example Thorarinsdottir and Gneiting (2010) proposed to fit a truncated normal distribution to the *untransformed* wind speed. In this case, in Equations 6 and 10 the logistic distribution is replaced with a truncated normal distribution and $g(y)$ is set to $g(y) = y$. Note that Thorarinsdottir and Gneiting (2010) called this model also heteroscedastic censored regression although actually the data is considered to be truncated and not censored. In the following we therefore denote this model as heteroscedastic truncated Gaussian regression (*HTGR*) which we also employ as benchmark model for wind speed.

2.5. Comparison

Table 1 summarizes the major differences between the 4 different logistic regression models that were presented above. Extended logistic regression (*XLR*) and censored logistic regression (*CLR*) (and their heteroscedastic versions HXLR and HCLR, respectively) are essentially the same models and only differ in their parameter estimation. They have the fewest parameters of the compared models but imply continuous distribution assumptions. Ordered logistic regression (*OLR*) and its heteroscedastic version (HOLR) avoid this continuous distribution assumption but require estimation of more coefficients than *(H)XLR* and *(H)CLR*. With its

Model		$g(y)$	$\mathbf{x}$	$\mathbf{z}$
SLR	Separate logistic regressions	–	$(1, M, S * M)^\top$	–
HOLR	Het. ordered logistic regression	–	M	S
HXLR	Het. extended logistic regression	$\sqrt{y}$	$(1, M)^\top$	$(1, S)^\top$
HCLR	Het. censored logistic regression	$\sqrt{y}$	$(1, M)^\top$	$(1, S)^\top$
HTGR	Het. truncated Gaussian regression	y	$(1, M_r)^\top$	$(1, S_r)^\top$

Table 2: List of different statistical models. $g(y)$ is the transformation, $\mathbf{x}$ are vectors of predictor variables for the location (mean) and $\mathbf{z}$ are predictor variables for the scale (variance). M and S are the mean and standard deviation of square root transformed ensemble forecasts respectively and M_r and S_r are the mean and standard deviation of the untransformed ensemble forecasts respectively. For wind speed forecasts M, S, M_r, and S_r are derived from 10m wind speed ensemble forecasts and for precipitation forecasts M and S are derived from total precipitation ensemble forecasts.

unconstrained slope estimates, separate logistic regressions *SLR* is more flexible than *OLR* but requires estimation of even more coefficients. Figure 1 shows schematic parallel regression lines for *XLR*, *CLR*, or *OLR*. In contrast to these models, regression lines from *SLR* would not be constrained to be parallel and so could potentially cross, which would lead to nonsense negative probabilities.

3. Data

To compare the presented ensemble post-processing methods, we used 10 meter wind speed observations (10 minute average) and 24-h accumulated precipitation amount from the two European weather stations *Paris–Orly* (48.717 N, 2.383 E) and *Wien–Hohe-Warte* (48.249 N, 16.356 E). As input for the statistical models, 10m wind speed and total precipitation ensemble forecasts from the European Centre for Medium Range Weather Forecasts (ECMWF) were linearly interpolated from neighboring grid points to the station locations. The data were available from April 2010 to December 2012 and we mainly considered the lead times 24, 48, and 96 hours for this study.

Since the predictands were square root transformed for most regression models (see Section 4) we mainly used the mean and standard deviation of *square root transformed* ensemble forecasts as predictor variables. For *HTGR* the untransformed predictand is used, following Thorarinsdottir and Gneiting (2010). Consequently we employed the mean and standard deviation of the *untransformed* ensemble forecasts as input for this model.

As thresholds q_j we defined $M = 9$ climatological deciles that are estimated for each location and predictand variable separately. Note that for precipitation several deciles are 0 and are merged to one threshold.

We found the ensemble spread to improve the forecasts of all statistical models, indicating useful spread-skill relationships. Therefore we only show results for the heteroscedastic models in the following. For separate logistic regressions the product of ensemble mean and spread is included as additional predictor variable (Wilks and Hamill 2007). Table 2 lists the different models that are compared in the following in detail.

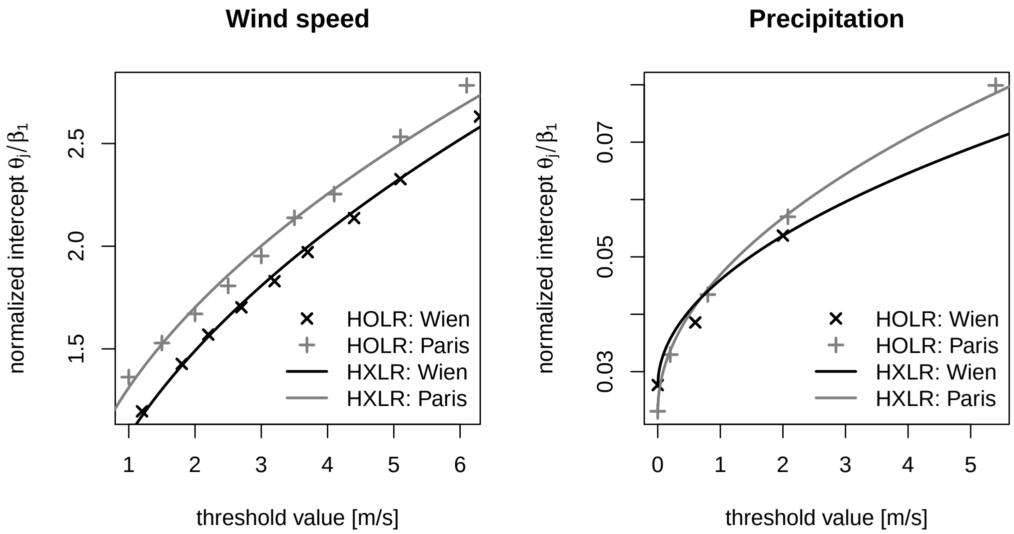

Figure 2: Intercepts θ_j from heteroscedastic ordered ($HOLR$) and extended logistic regression ($HXLR$) relative to threshold values, for the locations *Wien–Hohe-Warte* and *Paris–Orly*, lead time 48 hours, and the predictands wind speed (left) and 24 hours accumulated precipitation amount (right). For better comparability intercepts are normalized with β_1 respectively. The square root is used as transformation for $HXLR$ ($g(q_j) = \sqrt{q_j}$).

4. Results

Before comparing the performance of the different ensemble post-processing methods we show how ordered logistic regression can be used to determine appropriate transformations $g()$ for extended logistic regression. The crosses and plus-signs in Figure 2 show the fitted intercepts from ordered logistic regression ($HOLR$) for the two locations and two predictands. For both locations and variables these plots suggest that the intercepts can be parameterized as being proportional to the square roots of the thresholds. Thus we fitted $HXLR$ models with $g(q_j) = \sqrt{q_j}$ and added the corresponding $HXLR$ intercept functions $\theta_j = \beta_0 + \alpha\sqrt{q_j}$ as lines in Figure 2. For both predictand variables and locations the $HXLR$ intercept functions fit the $HOLR$ intercepts reasonably well. Note that similar figures can also be used to compare the intercepts of extended logistic regression with those of separate logistic regression (e.g., Ruiz and Saulo 2012). However, the varying slope coefficients then complicate the comparison.

Figure 2 already suggests that $HXLR$ and $HOLR$ predict similarly well. In the following we compare these and the other statistical models more thoroughly. Because all models provide probabilistic forecasts for discrete intervals we mainly employ the ranked probability score (RPS; Epstein 1969; Wilks 2006b) to characterize forecast accuracy:

$$RPS = \sum_{j=1}^{J} (P(y \leq q_j|\mathbf{x}) - I(y \leq q_j))^2 \qquad (12)$$

where J is the number of thresholds and $I(\cdot)$ is the indicator function. For each model, forecast location, and lead time we applied 10-fold cross validation to get independent training and

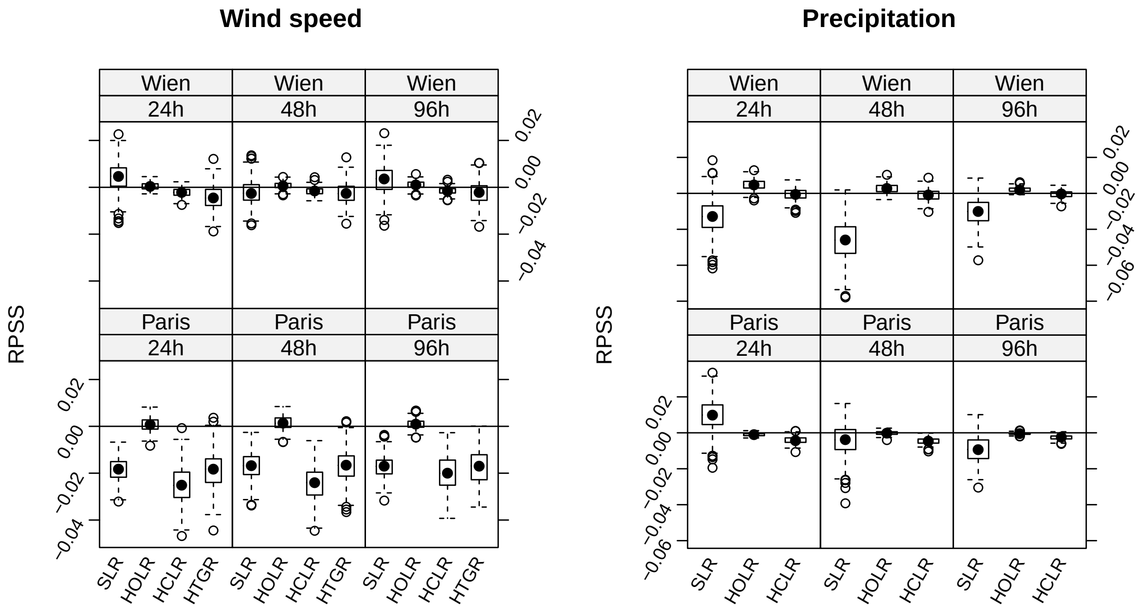

Figure 3: Ranked probability skill score (*RPSS*) relative to heteroscedastic extended logistic regression (*HXLR*) of wind speed (left) and 24 hours accumulated precipitation amount (right) for different models (see Table 2 for details) and locations. 9 climatological deciles that were computed separately for each forecast location are used as thresholds. Positive values indicate improvements over *HXLR*. The solid circles mark the median and the boxes the interquartile ranges of the 250 values from the bootstrapping approach, the whiskers show the most extreme values that are less than 1.5 times the length of the box away from the box, and empty circles are plotted for values that are outside the whiskers.

test data sets. To estimate the sampling distribution for the average $\overline{RPS}$ we computed means of 250 bootstrap samples. To compare the models with a reference model we finally computed ranked probability *skill* scores (*RPSS*):

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS}_{ref}} \tag{13}$$

where $\overline{RPS}_{ref}$ is the $\overline{RPS}$ of appropriate reference forecasts.

Figure 3 shows the *RPSS* relative to *HXLR* for different models, lead times, locations, and predictand variables. *HOLR* performs equally well or slightly better than *HXLR* for all locations, lead times, and predictand variables. For precipitation in *Paris* forecasts of *HXLR* and *HOLR* are nearly identical which is consistent with Figure 2 where the *HXLR* intercept function almost perfectly interpolates the *HOLR* intercepts. Separate logistic regressions (*SLR*) mostly performs worse than *HXLR*. Exceptions are wind speed forecasts in *Wien* for 24 and 96 hours lead time and precipitation forecasts in *Paris* for 24 hours lead time. However, note that the *RPS* (Equation 12) does not penalize the partly inconsistent forecasts from *SLR*. *HCLR* and *HTGR* also tend to perform worse than *HXLR*. While for *Paris HTGR* is slightly better than *HCLR* there is no clear preference for one of these models in *Wien*.

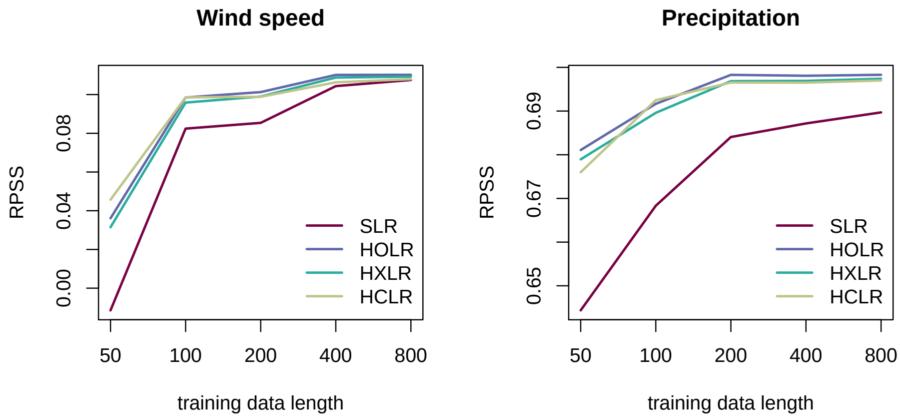

Figure 4: Ranked probability skill score (*RPSS*) relative to the raw ensemble (ensemble relative frequencies within each interval) in *Wien–Hohe-Warte* for different training data lengths and models (see Table 2 for details) and lead time 48 hours. 9 climatological deciles are used as thresholds for wind speed and 3 for precipitation.

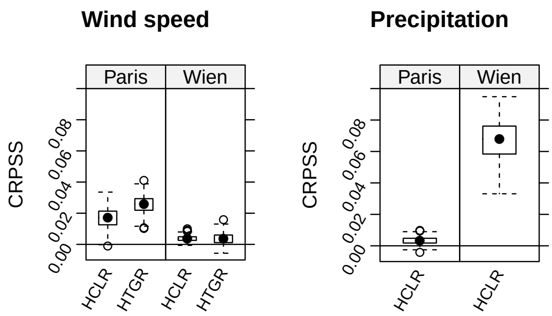

Figure 5: Continuous ranked probability skill score (*CRPSS*) relative to heteroscedastic extended logistic regression (*HXLR*) and their bootstrap sampling distributions in *Wien–Hohe-Warte* for different predictands, models (see Table 2 for details), and locations. Positive values indicate improvements over *HXLR*.

Because the different statistical models differ considerably in their number of estimated coefficients (*SLR*: $3J$, *HOLR*: $2 + J$, *HXLR*, *HCLR*, *HTGR*: 4) it is also interesting to compare their performance for different training data lengths. Figure 4 shows RPSS for windspeed and precipitation forecasts for 48 hours lead time at Wien, relative to the raw ensemble interval relative frequencies. It can be seen that all models lose skill with a reduced training data set. With the largest parameter count *SLR* clearly loses most and for wind speed even performs worse than the raw ensemble ($RPSS < 0$) when the training data contains only 50 days. The other models exhibit comparable skill reductions in response to decreasing training data.

HCLR basically fits the same model as *HXLR*, with the only difference being that the estimated model parameters optimize either the selected category probabilities (*HXLR*) or the continuous predictive distribution (*HCLR*). Since the *RPS* only measures the quality of the

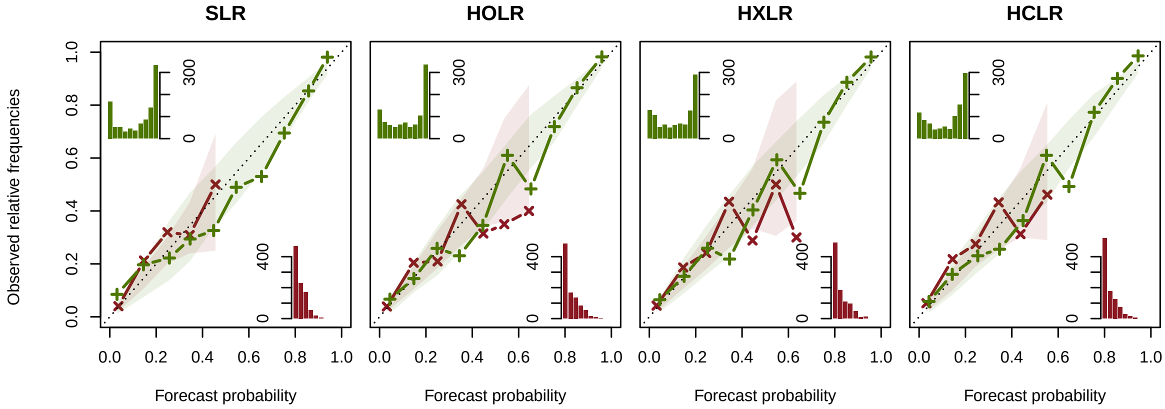

Figure 6: Reliability diagrams for predicted probabilities to fall below the first climatological decile $P(y \leq q_1|\mathbf{x})$ for *Wien–Hohe-Warte*, lead time 48 hours, and different models. Forecasts are aggregated in 0.1 probability intervals. Calibration functions for *wind speed* are plotted as red 'x' and for *precipitation amount* as green '+' and are only shown for intervals with more than 10 forecasts. Refinement distributions for wind speed are plotted in the bottom right corner in red and for precipitation in the top left corner in green. 95% consistency intervals derived from consistency re-sampling (Bröcker and Smith 2007) are shown as red and green shaded areas respectively. Note that due to the frequent zero observations $q_1 = q_2 = \cdots = q_6$ for precipitation so that $P(y \leq q_1|\mathbf{x}) = P(y \leq q_6|\mathbf{x}) = P(y = 0|\mathbf{x})$.

selected category probabilities the better *RPS* of *HXLR* in Figure 3 is not surprising. To compare also the quality of the full predictive distributions we therefore employ the continuous ranked probability score (*CRPS*; Matheson and Winkler 1976; Hersbach 2000; Wilks 2006b) that generalizes the *RPS* to full predictive distributions.

$$CRPS = \int_{-\infty}^{\infty} (P(y_i \leq t|\mathbf{x}) - I(y_i \leq t))^2 dt \tag{14}$$

Analogously to Figure 3 the continuous ranked probability skill score (*CRPSS*) relative to *HXLR* is shown in Figure 5. In contrast to the *RPSS* (Figure 3) the *CRPSS* clearly favors *HCLR* for both locations and predictand variables. Note that the large improvement of *HCLR* over *HXLR* for precipitation in *Wien* mainly stems from *HXLR*'s bad forecast performance for very high precipitation amounts. The inclusion of additional thresholds in the parameter fitting process (e.g., climatological 0.95-quantile) substantially improved the *CRPS* of *HXLR* and consequently diminished the *CRPSS* of *HCLR* (not shown).

For wind speed, Figure 5 also shows the *CRPSS* for *HTGR*. As in Figure 3 *HCLR* and *HTGR* show similar *CRPSS* for *Wien* while *HTGR* is slightly preferred for *Paris*, which suggests that there the real error distribution is better estimated by a truncated normal than by a censored transformed logistic distribution.

Finally Figures 6 and 7 show reliability diagrams (e.g., Wilks 2006b) for the lower and upper climatological deciles, respectively, for 48 hours lead time at *Wien*. With few exceptions the observed conditional relative frequencies of both predictand variables lie within the 95% consistency intervals (Bröcker and Smith 2007) with only minor differences between the different statistical models. Similarly, the refinement distributions in Figures 6 and 7 show only little

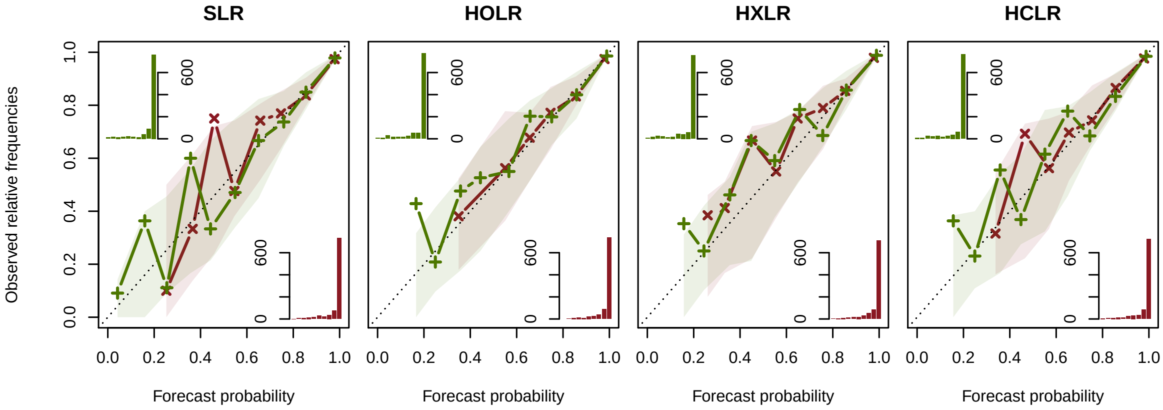

Figure 7: Same as Figure 6 but for predicted probabilities to fall below the upper climatological decile $P(y \leq q_9|\mathbf{x})$.

differences between the different models. Only for zero precipitation *SLR* and *HOLR* have slightly sharper forecasts than *HXLR* and *HCLR* (forecasts more frequently close to 0 and 1).

5. Summary and conclusion

Extended logistic regression fits predictand category probabilities by assuming a conditional logistic distribution for the transformed predictand (Scheuerer 2013; Schefzik *et al.* 2013; Messner *et al.* 2013). However, for some applications the transformed predictand cannot be assumed to follow a logistic distribution. Moreover, fitting selected category probabilities implies disregarding available information when the predictand is actually given in continuous form.

In this study we compared extended logistic regression with two closely related regression models from statistics and econometrics. Ordered logistic regression is very similar to extended logistic regression but avoids a continuous distribution assumption. On the other hand, censored logistic regression fits the same model as extended logistic regression but uses each individual predictand value in the training data set instead of the selected category probabilities. As further benchmark models we also employed separate logistic regressions and a truncated Gaussian regression model (Thorarinsdottir and Gneiting 2010). The performance of the different statistical models was tested with wind speed and precipitation data from two European locations and ensemble forecasts from the ECMWF. Overall, the logistic distribution assumption seemed to be quite appropriate for the square-root-transformed predictands, at both locations and for both predictand variables. Thus, the performance differences between ordered and extended logistic regression were only minor. However, because no continuous distribution has to be assumed, ordered logistic regression should generally be preferred if solely threshold probabilities are required.

Since extended logistic regression fits selected category probabilities, it is actually not surprising that *RPS* skills are higher for this model than for censored logistic regression, which fits the full continuous predictive distribution. For the same reason it is unsurprising that censored logistic regression performed better than extended logistic regression according to

CRPS skill, which evaluates accuracy of the full predictive distributions.

Extended and censored logistic regression assume censored conditional logistic distributions for the transformed predictand. In contrast, wind speed was assumed to follow a truncated normal distribution in Thorarinsdottir and Gneiting (2010). A comparison between censored and truncated regression models showed that the assumption of a truncated normal distribution resulted in slightly better wind speed forecasts than the assumption of a censored transformed logistic distribution.

As input for the statistical methods we only employed NWP model forecasts for the observation location and forecast variable. However, all models might potentially be improved with additional inputs such as NWP model forecasts of other variables and/or locations.

Nevertheless, our results show that the optimal statistical model strongly depends on the intended application. Ordered logistic regression was best suited for category probability predictions for the forecasts considered here, given sufficiently long training series. When the transformed predictand can be assumed to follow a conditional logistic distribution then extended logistic regression provides equally good category probability forecasts while requiring fewer coefficients and additionally specifying full predictive distributions. However, if the primary interest is in predicting full continuous probability distributions, censored or truncated regression models should be preferred because they use the information contained in the training data more fully.

Computational details

Our results were obtained on Ubuntu Linux using the statistical software R 3.0.1 (R Core Team 2013). Heteroscedastic extended logistic regression and heteroscedastic censored logistic regression were fitted using the package crch 0.1-0 (Messner and Zeileis 2013). For ordered logistic regression models we used the package ordinal 2013.9-30 (Christensen 2013).

Acknowledgments

This study was supported by the Austrian Science Fund (FWF): L615-N10. The first author was also supported by a PhD scholarship from the University of Innsbruck, *Vizerektorat für Forschung*. Data from the ECMWF forecasting system were obtained from the ECMWF Data Server.

References

Agresti A (2002). *Categorical Data Analysis*. 2nd edition. John Wiley & Sons.

Ben Bouallègue Z (2013). "Calibrated Short-Range Ensemble Precipitation Forecasts Using Extended Logistic Regression with Interaction Terms." *Weather and Forecasting*, **28**(2), 515–524.

Bröcker J, Smith LA (2007). "Increasing the Reliability of Reliability Diagrams." *Weather and Forecasting*, **22**(3), 651–661.

Christensen RHB (2013). *ordinal: Regression Models for Ordinal Data.* R package version 2013.09-30, URL http://CRAN.R-project.org/package=ordinal.

Epstein ES (1969). "A Scoring System for Probability Forecasts of Ranked Categories." *Journal of Applied Meteorology*, **8**(6), 985–987.

Gneiting T, Raftery AE, Westveld AH, Goldman T (2005). "Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation." *Monthly Weather Review*, **133**(5), 1098–1118.

Hamill TM (2012). "Verification of TIGGE Multimodel and ECMWF Reforecast-Calibrated Probabilistic Precipitation Forecasts over the Contiguous United States." *Monthly Weather Review*, **140**(7), 2232–2252.

Hamill TM, Snyder C, Whitaker JS (2003). "Ensemble Forecasts and the Properties of Flow-Dependent Analysis-Error Covariance Singular Vectors." *Monthly Weather Review*, **131**(8), 1741–1758.

Hamill TM, Whitaker JS, Wei X (2004). "Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts." *Monthly Weather Review*, **132**(6), 1434–1447.

Hersbach H (2000). "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems." *Weather and Forecasting*, **15**(5), 559–570.

Matheson JE, Winkler RL (1976). "Scoring Rules for Continuous Probability Distributions." *Management Science*, **22**(10), 1087–1096.

Messner JW, Zeileis A (2013). *crch: Censored Regression with Conditional Heteroscedasticity.* R package version 0.1-0, URL http://CRAN.R-project.org/package=crch.

Messner JW, Zeileis A, Mayr GJ, Wilks DS (2013). "Heteroscedastic Extended Logistic Regression for Post-Processing of Ensemble Guidance." *Monthly Weather Review*, p. in press.

Nelder JA, Wedderburn RWM (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society A*, **135**(3), 370–384.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005). "Using Bayesian Model Averaging to Calibrate Forecast Ensembles." *Monthly Weather Review*, **133**(5), 1155–1174.

Roulin E, Vannitsem S (2012). "Postprocessing of Ensemble Precipitation Predictions with Extended Logistic Regression Based on Hindcasts." *Monthly Weather Review*, **140**(3), 874–888.

Roulston MS, Smith LA (2003). "Combining Dynamical and Statistical Ensembles." *Tellus A*, **55**(1), 16–30.

Ruiz JJ, Saulo C (2012). "How Sensitive are Probabilistic Precipitation Forecasts to the Choice of Calibration Algorithms and the Ensemble Generation Method? Part I: Sensitivity to Calibration Methods." *Meteorological Applications*, **19**(3), 302–313.

Schefzik R, Thorarinsdottir TL, Gneiting T (2013). "Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling." *Statistical Science*, p. in press.

Scheuerer M (2013). "Probabilistic Quantitative Precipitation Forecasting Using Ensemble Model Output Statistics." *Quarterly Journal of the Royal Meteorological Society*. In press.

Schmeits MJ, Kok KJ (2010). "A Comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging, and Extended Logistic Regression Using ECMWF Ensemble Precipitation Reforecasts." *Monthly Weather Review*, **138**(11), 4199–4211.

Thorarinsdottir TL, Gneiting T (2010). "Probabilistic Forecasts of Wind Speed: Ensemble Model Output Statistics by Using Heteroscedastic Censored Regression." *Journal of the Royal Statistical Society A*, **173**(2), 371–388.

Tobin J (1958). "Estimation of Relationships for Limited Dependent Variables." *Econometrica*, **26**(1), 24–36.

Wang X, Bishop CH (2003). "A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes." *Journal of the Atmospheric Sciences*, **60**(9), 1140–1158.

Wilks DS (2006a). "Comparison of Ensemble-MOS Methods in the Lorenz '96 Setting." *Meteorological Applications*, **13**(3), 243–256.

Wilks DS (2006b). *Statistical Methods in the Atmospheric Sciences*, volume 14. 2nd edition. Academic Press.

Wilks DS (2009). "Extending Logistic Regression to Provide Full-Probability-Distribution MOS Forecasts." *Meteorological Applications*, **368**(March), 361–368.

Wilks DS, Hamill TM (2007). "Comparison of Ensemble-MOS Methods Using GFS Reforecasts." *Monthly Weather Review*, **135**(6), 2379–2390.

Affiliation:

Jakob W. Messner, Georg J. Mayr
Institute of Meteorology and Geophysics
Universität Innsbruck
Innrain 52
6020 Innsbruck, Austria
E-mail: Jakob.Messner@uibk.ac.at, Georg.Mayr@uibk.ac.at

Daniel S. Wilks
Department of Earth and Atmospheric Sciences
Cornell University
Ithaca, New York, United States of America
E-mail: dsw5@cornell.edu

Achim Zeileis
Department of Statistics

Faculty of Economics and Statistics
Universität Innsbruck
Universitätsstraße 15
6020 Innsbruck, Austria
E-mail: Achim.Zeileis@R-project.org
URL: http://eeecon.uibk.ac.at/~zeileis/

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

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

2013-32 **Jakob W. Messner, Georg J. Mayr, Daniel S. Wilks, Achim Zeileis:** Extending extended logistic regression for ensemble post-processing: Extended vs. separate vs. ordered vs. censored

2013-31 **Anita Gantner, Kristian Horn, Rudolf Kerschbamer:** Fair division in unanimity bargaining with subjective claims

2013-30 **Anita Gantner, Rudolf Kerschbamer:** Fairness and efficiency in a subjective claims problem

2013-29 **Tanja Hörtnagl, Rudolf Kerschbamer, Rudi Stracke, Uwe Sunde:** Heterogeneity in rent-seeking contests with multiple stages: Theory and experimental evidence

2013-28 **Dominik Erharter:** Promoting coordination in summary-statistic games

2013-27 **Dominik Erharter:** Screening experts' distributional preferences

2013-26 **Loukas Balafoutas, Rudolf Kerschbamer, Matthias Sutter:** Second-degree moral hazard in a real-world credence goods market

2013-25 **Rudolf Kerschbamer:** The geometry of distributional preferences and a non-parametric identification approach

2013-24 **Nadja Klein, Michel Denuit, Stefan Lang, Thomas Kneib:** Nonlife ratemaking and risk management with bayesian additive models for location, scale and shape

2013-23 **Nadja Klein, Thomas Kneib, Stefan Lang:** Bayesian structured additive distributional regression

2013-22 **David Plavcan, Georg J. Mayr, Achim Zeileis:** Automatic and probabilistic foehn diagnosis with a statistical mixture model

2013-21 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis, Daniel S. Wilks:** Extending extended logistic regression to effectively utilize the ensemble spread

2013-20 **Michael Greinecker, Konrad Podczeck:** Liapounoff's vector measure theorem in Banach spaces *forthcoming in Economic Theory Bulletin*

2013-19 **Florian Lindner:** Decision time and steps of reasoning in a competitive market entry game

2013-18 **Michael Greinecker, Konrad Podczeck:** Purification and independence

2013-17 **Loukas Balafoutas, Rudolf Kerschbamer, Martin Kocher, Matthias Sutter:** Revealed distributional preferences: Individuals vs. teams

2013-16 **Simone Gobien, Björn Vollan:** Playing with the social network: Social cohesion in resettled and non-resettled communities in Cambodia

2013-15 **Björn Vollan, Sebastian Prediger, Markus Frölich:** Co-managing common pool resources: Do formal rules have to be adapted to traditional ecological norms?

2013-14 **Björn Vollan, Yexin Zhou, Andreas Landmann, Biliang Hu, Carsten Herrmann-Pillath:** Cooperation under democracy and authoritarian norms

2013-13 **Florian Lindner, Matthias Sutter:** Level-k reasoning and time pressure in the 11-20 money request game *forthcoming in* *Economics Letters*

2013-12 **Nadja Klein, Thomas Kneib, Stefan Lang:** Bayesian generalized additive models for location, scale and shape for zero-inflated and overdispersed count data

2013-11 **Thomas Stöckl:** Price efficiency and trading behavior in limit order markets with competing insiders *forthcoming in* *Experimental Economics*

2013-10 **Sebastian Prediger, Björn Vollan, Benedikt Herrmann:** Resource scarcity, spite and cooperation

2013-09 **Andreas Exenberger, Simon Hartmann:** How does institutional change coincide with changes in the quality of life? An exemplary case study

2013-08 **E. Glenn Dutcher, Loukas Balafoutas, Florian Lindner, Dmitry Ryvkin, Matthias Sutter:** Strive to be first or avoid being last: An experiment on relative performance incentives.

2013-07 **Daniela Glätzle-Rützler, Matthias Sutter, Achim Zeileis:** No myopic loss aversion in adolescents? An experimental note

2013-06 **Conrad Kobel, Engelbert Theurl:** Hospital specialisation within a DRG-Framework: The Austrian case

2013-05 **Martin Halla, Mario Lackner, Johann Scharler:** Does the welfare state destroy the family? Evidence from OECD member countries

2013-04 **Thomas Stöckl, Jürgen Huber, Michael Kirchler, Florian Lindner:** Hot hand belief and gambler's fallacy in teams: Evidence from investment experiments

2013-03 **Wolfgang Luhan, Johann Scharler:** Monetary policy, inflation illusion and the Taylor principle: An experimental study

2013-02 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Tensions between the resource damage and the private benefits of appropriation in the commons

2013-01 **Jakob W. Messner, Achim Zeileis, Jochen Broecker, Georg J. Mayr:** Improved probabilistic wind power forecasts with an inverse power curve transformation and censored regression

2012-27 **Achim Zeileis, Nikolaus Umlauf, Friedrich Leisch:** Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond

2012-26 **Francisco Campos-Ortiz, Louis Putterman, T.K. Ahn, Loukas Balafoutas, Mongoljin Batsaikhan, Matthias Sutter:** Security of property as a public good: Institutions, socio-political environment and experimental behavior in five countries

2012-25 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Appropriation in the commons: variations in the opportunity costs of conservation

2012-24 **Edgar C. Merkle, Jinyan Fan, Achim Zeileis:** Testing for measurement invariance with respect to an ordinal variable *forthcoming in Psychometrika*

2012-23 **Lukas Schrott, Martin Gächter, Engelbert Theurl:** Regional development in advanced countries: A within-country application of the Human Development Index for Austria

2012-22 **Glenn Dutcher, Krista Jabs Saral:** Does team telecommuting affect productivity? An experiment

2012-21 **Thomas Windberger, Jesus Crespo Cuaresma, Janette Walde:** Dirty floating and monetary independence in Central and Eastern Europe - The role of structural breaks

2012-20 **Martin Wagner, Achim Zeileis:** Heterogeneity of regional growth in the European Union

2012-19 **Natalia Montinari, Antonio Nicolo, Regine Oexl:** Mediocrity and induced reciprocity

2012-18 **Esther Blanco, Javier Lozano:** Evolutionary success and failure of wildlife conservancy programs

2012-17 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Beliefs and truth-telling: A laboratory experiment

2012-16 **Alexander Sebald, Markus Walzl:** Optimal contracts based on subjective evaluations and reciprocity

2012-15 **Alexander Sebald, Markus Walzl:** Subjective performance evaluations and reciprocity in principal-agent relations

2012-14 **Elisabeth Christen:** Time zones matter: The impact of distance and time zones on services trade

2012-13 **Elisabeth Christen, Joseph Francois, Bernard Hoekman:** CGE modeling of market access in services

2012-12 **Loukas Balafoutas, Nikos Nikiforakis:** Norm enforcement in the city: A natural field experiment *forthcoming in European Economic Review*

2012-11 **Dominik Erharter:** Credence goods markets, distributional preferences and the role of institutions

2012-10 **Nikolaus Umlauf, Daniel Adler, Thomas Kneib, Stefan Lang, Achim Zeileis:** Structured additive regression models: An R interface to BayesX

2012-09 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** History repeating: Spain beats Germany in the EURO 2012 Final

2012-08 **Loukas Balafoutas, Glenn Dutcher, Florian Lindner, Dmitry Ryvkin:** The optimal allocation of prizes in tournaments of heterogeneous agents

2012-07 **Stefan Lang, Nikolaus Umlauf, Peter Wechselberger, Kenneth Harttgen, Thomas Kneib:** Multilevel structured additive regression

2012-06 **Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang:** Bayesian semiparametric additive quantile regression

2012-05 **Eric Mayer, Sebastian Rueth, Johann Scharler:** Government debt, inflation dynamics and the transmission of fiscal policy shocks *forthcoming in Economic Modelling*

2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints? *forthcoming in Economica*

2012-03 **Uwe Dulleck, David Johnston, Rudolf Kerschbamer, Matthias Sutter:** The good, the bad and the naive: Do fair prices signal good types or do they induce good behaviour?

2012-02 **Martin G. Kocher, Wolfgang J. Luhan, Matthias Sutter:** Testing a forgotten aspect of Akerlof's gift exchange hypothesis: Relational contracts with individual and uniform wages

2012-01 **Loukas Balafoutas, Florian Lindner, Matthias Sutter:** Sabotage in tournaments: Evidence from a natural experiment *published in Kyklos*

University of Innsbruck

Working Papers in Economics and Statistics

2013-32

Jakob W. Messner, Georg J. Mayr, Daniel S. Wilks, Achim Zeileis

Extending extended logistic regression for ensemble post-processing: Extended vs. separate vs. ordered vs. censored

Abstract
Extended logistic regression is a recent ensemble calibration method that extends logistic regression to provide full continuous probability distribution forecasts. It assumes conditional logistic distributions for the (transformed) predictand and fits these using selected predictand category probabilities. In this study we compare extended logistic regression to the closely related ordered and censored logistic regression models. Ordered logistic regression avoids the logistic distribution assumption but does not yield full probability distribution forecasts, whereas censored regression directly fits the full conditional predictive distributions. To compare the performance of these and other ensemble post-processing methods we used wind speed and precipitation data from two European locations and ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). Ordered logistic regression performed similarly to extended logistic regression for probability forecasts of discrete categories whereas full predictive distributions were better predicted by censored regression.

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