University of Innsbruck

Improved probabilistic wind power forecasts with an inverse power curve transformation and censored regression

Jakob W. Messner, Achim Zeileis,
Jochen Broecker, Georg J. Mayr

Working Papers in Economics and Statistics

2013-01

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.

Improved Probabilistic Wind Power Forecasts with an Inverse Power Curve Transformation and Censored Regression

Jakob W. Messner
Universität Innsbruck

Achim Zeileis
Universität Innsbruck

Jochen Broecker
University of Reading

Georg J. Mayr
Universität Innsbruck

Abstract

Forecasting wind power is an important part of a successful integration of wind power into the power grid. Forecasts with lead times longer than 6 hours are generally made by using statistical methods to postprocess forecasts from numerical weather prediction systems. Two major problems that complicate this approach are the nonlinear relationship between wind speed and power production and the limited range of power production between zero and nominal power of the turbine. In practice, the nonlinearity is often tackled by using nonlinear nonparametric regression methods while the limited range is typically not addressed explicitly. However, such an approach ignores valuable and readily available information: the power curve of the turbine's manufacturer. Much of the nonlinearity can be directly accounted for by transforming the observed power production into wind speed via the inverse power curve so that simpler linear regression models can be used. Furthermore, the limited range of the transformed power production can be easily exploited by adopting censored regression models.

In this study, we evaluate quantile forecasts from a range of methods: (a) using parametric and nonparametric models, (b) with and without the proposed inverse power curve transformation, and (c) with and without censoring. The results show that with our inverse (power-to-wind) transformation, simpler linear regression models with censoring perform equally or better than nonlinear models with or without the frequently used wind-to-power transformation.

Keywords: wind power, probabilistic forecasting, power curve transformation, censored regression, quantile regression.

1. Introduction

The importance of wind energy has increased significantly in the past decades. In 2011 approximately 21% of installed power capacity in Europe was from wind power (Wilkes, Moccia, and Dragan 2012). One problem of integrating wind power into the electricity grid is the volatility of wind speed and consequently of power production. Prediction of power production is therefore crucial for energy trading and management. In this context, probabilistic forecast methods have been receiving increased attention recently because of their higher value in decision making when compared to single value (point) forecasts (Pinson, Kariniotakis, Nielsen, and Madsen 2006; Roulston and Smith 2003; Bremnes 2004). Probabilistic forecasts can be for example quantile or interval forecasts, full predictive distributions, or risk indices in addition to point forecasts.

The general approach to make probabilistic power production forecasts with lead times ≥ 6 hours is to statistically postprocess forecasts (mainly wind speed forecasts) from numerical weather prediction (NWP) models (Giebel, Brownsword, Kariniotakis, Denhard, and Draxl 2011). In the atmospheric sciences, this approach is termed model output statistics (MOS, Glahn and Lowry 1972). However, standard linear regression analysis, as typically used for MOS, cannot be used due to two major problems:

1. The relationship between wind speed and power production is clearly nonlinear (see Figures 1 and 2) .
2. The range of power production is limited (censored) between zero and nominal power so that typical parametric distribution assumptions (e.g., Gaussian) are inappropriate.

To overcome these problems, nonlinear and often also nonparametric regression methods are used frequently in the literature. For example, Bremnes (2004, 2006) uses locally weighted quantile regression methods. Nielsen, Madsen, and Nielsen (2006) employs quantile regression with spline basis functions. Pinson *et al.* (2006) accounts for the nonlinearity with a fuzzy inference model and Juban, Fugon, and Kariniotakis (2007) suggests to use quantile regression forests or a kernel density estimator. However, the disadvantages of such nonparametric nonlinear models are that generally a large number of parameters have to be estimated and therefore these estimations can be unstable, especially in cases where few data are available. Furthermore, the resulting models are sometimes hard to interpret and, more importantly, neglect the available information about the form of the power curve and the censoring.

Therefore we propose a new (set of) approach(es):

1. Transform the observed power observations into wind speed observations prior to MOS regression modeling by using the inverse of the power curve function. Note that this transforms the limited range from zero to nominal power into the limited range from cut-in wind speed to nominal wind speed.
2. Exploit the information about this limited range by using censored models in “wind space” where typically much simpler (more) linear regressions can be used and parametric distributions work well.

Figure 4 shows the relationship between power observations, transformed with the inverse power curve on the y-axis and NWP wind speed forecasts on the x-axis. Clearly, this seems to be almost linear and just the censoring of the transformed power observations at cut-in and nominal wind speed has to be accounted for in a regression model. While such censored regression techniques are not very frequently used for MOS, they are among the standard regression models in statistics and econometrics and easily available in many software statistics packages. Thus we can obtain probabilistic forecasts in “wind space” with a relatively simple model and then employ the power curve again to transform these to probabilistic power production forecasts.

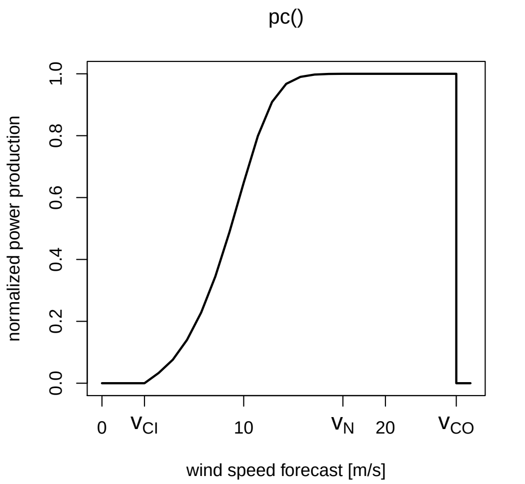

Figure 1: Power curve function $pc()$ of the turbine manufacturer: Power production by wind speed.

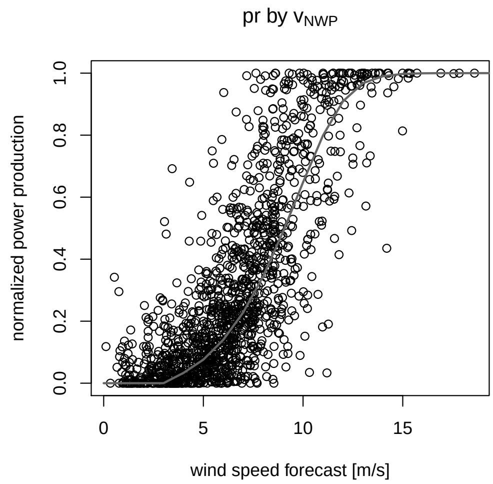

Figure 2: Normalized power production (black points) by ECMWF wind speed forecasts with power curve (gray line).

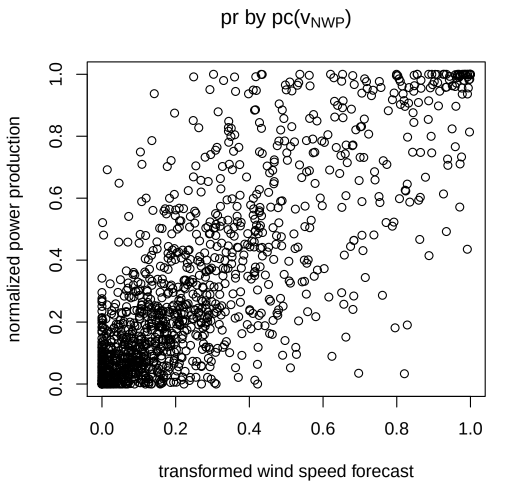

Figure 3: Power curve transformation (wind to power): Observed power production by transformed ECMWF wind speed forecasts.

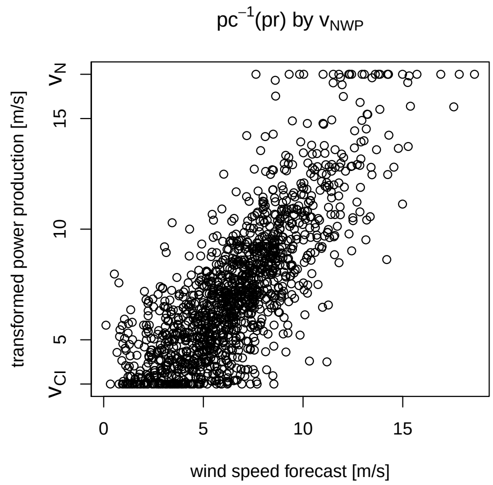

Figure 4: Inverse power curve transformation (power to wind): Transformed power production by ECMWF wind speed forecasts.

We are not the first to suggest usage of the known power curve to address the nonlinearity issue. However, previous approaches employed the power curve itself rather than its inverse to transform the NWP wind speed forecasts into power forecasts prior to regression modeling (Lange 2005; Nielsen, Madsen, Nielsen, Badger, Giebel, Landberg, Sattler, and Feddersen 2004; Roulston 2003). While this is also very easy to carry out (see Figure 3 for an example), it has a crucial disadvantage: In the steep parts of the power curve, errors in the NWP wind speed forecasts are strongly amplified while errors of low and high NWP wind speed forecasts are suppressed. Hence, the resulting relationship between the (wind-to-power) transformed NWP wind speed forecasts and observed power production exhibits strong heteroskedasticity which leads to less reliable estimates in regression models. Note the higher variance in the center of Figure 3 as compared with the lower variance on the left and right side. In contrast, the inverse power-to-wind transformed relationship in Figure 4 has a rather low and stable variance (only limited by censoring at cut-in and nominal speed).

In this study, we demonstrate how both parametric and nonparametric censored (linear) regression model can be employed for inverse power curve transformed data (i.e., in wind space). The resulting models are assessed and compared with previously suggested approaches for untransformed data as well as wind-to-power transformed data (i.e., in power space), showing that in many situations we can get similar or even better performance from models that are easier to compute and interpret. As observation data, we use 3 years of wind turbine data from a turbine located in Austria. As NWP forecasts, high resolution and ensemble forecasts of wind in different heights from the European Centre for Medium-Range Weather Forecasts (ECMWF) are employed.

The remainder of the paper is organized as follows: In Section 2, the data used for testing the transformations and models are described briefly. The regression models are introduced in Section 3. The verification measures are specified in Section 4 and the corresponding results are shown in Section 5. Finally, a conclusion of the paper is provided in Section 6.

2. Data

As observation data, we utilize power production data from a wind turbine in eastern Austria with a nominal power of 2000kW. Measurements with 10 minute temporal resolution are available from 2006 to 2009. Data values when the turbine was off because of maintenance are removed.

As input for the statistical models, we use NWP forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). In particular, we use wind speed forecasts, linearly interpolated from neighbouring model levels to turbine hub height, as this has been shown to be the best predictor from ECMWF for wind speed on wind turbines (Drechsel, Mayr, Messner, and Stauffer 2012). No further variables (such as wind direction or air density) are added because they do not improve forecasts significantly for the data considered. To capture heteroskedasticity (i.e., the standard deviation of the observations) some of our models additionally employ the 10 meter wind speed ensemble standard deviation form the ECMWF ensemble prediction system (EPS). To combine the observation data (with temporal resolution of 10 minutes) with the NWP data (with resolution of 3 hours), means of the observation data are computed for 1 hour around the times for which forecasts are available.

Thus for each lead time, 1340 forecast-observation pairs are available. For lead time of 24 hours, the data is plotted in Figure 2. Note that all used ECMWF forecasts are initialized at 00UTC. 12 and 36 hour forecasts are therfore always for midday while 24 and 48 hour forecasts are always for midnight.

In the next sections the following notations are used:

n**:** Number of forecast-observation pairs.

pr_i**:** Power production; $i = 1, ..., n$.

v_i^*: Wind speed; $i = 1, ..., n$.

v_{CI}: Cut-in wind speed (wind speed where turbine starts to rotate).

v_N: Nominal wind speed (wind speed where turbine reaches maximum power).

$pc()$: Power curve function given by the turbine manufacturer; $pc(v_i^*) = pr_i$ (see Figure 1).

$v_i = pc^{-1}(pr_i)$: Inverse-transformed power production (see also Equation 1).

$\mathbf{x}_i, \mathbf{z}_i$: Vectors of input variables (NWP forecasts); $i = 1, ..., n$.

$q_\pi(y_i|\mathbf{x}_i)$: π-quantile of y_i given the regressor variables $\mathbf{x}_i$.

Note that the inverse-transformed power production (v_i) can be interpreted as wind speed censored at cut-in and nominal wind speed (see Figure 4). That means:

$$v_i = pc^{-1}(pr_i) = \begin{cases} v_{CI} & v_i^* \leq v_{CI} \\ v_i^* & v_{CI} < v_i^* < v_N \\ v_N & v_i^* \geq v_N \end{cases} \tag{1}$$

3. Regression models

To obtain probabilistic forecasts of power production, we consider a range of different regression models that lead either to conditional quantiles or full predictive distributions (from which conditional quantiles can be easily extracted). More formally, all models yield predictions of specific quantiles $q_\pi(pr_i|\mathbf{x}_i)$ of power production pr_i given a vector of regressor variables $\mathbf{x}_i$ (e.g., forecasts of wind speed etc.). We divide the models into parametric and nonparametric models. All models except some benchmark models are estimated in wind space. That means that quantiles $q_\pi(v_i^*|\mathbf{x}_i)$ of wind speed given some regressor variables are first estimated. Subsequently they are transformed to quantiles of transformed power by considering cut-in and nominal wind speed of the turbine and finally transformed into quantiles of power production by employing the power curve of the turbine:

$$q_\pi(v_i|\mathbf{x}_i) = \min(v_N, \max(v_{CI}, q_\pi(v_i^*|\mathbf{x}_i))) \tag{2}$$

$$q_\pi(pr_i|\mathbf{x}_i) = pc(q_\pi(pc^{-1}(pr_i)|\mathbf{x}_i)) = pc(q_\pi(v_i|\mathbf{x}_i)) \tag{3}$$

3.1. Parametric models

For parametric models, it is assumed that the response follows a specific distribution and here the normal (or Gaussian) distribution is used. If such an assumtion is appropriate these models are easy to estimate and with every forecast a full predictive ditribution is given. Arbitrary quantiles are very easy to compute by inverting this distribution. The main disadvantage of parametric models is that it is sometimes difficult to find an appropriate parametric distribution.

Tobit model

The tobit model was first introduced by Tobin (1958) and is a widely used linear model for censored data. For this model, it is assumed that the true wind speed v_i^* follows a normal distribution with a mean μ_i that depends linearly on some input variables $\mathbf{x}_i$ and typically a constant variance $\sigma_i = \gamma$:

$$v_i^* \sim N(\mu_i, \sigma_i^2) \tag{4}$$
$$\mu_i = \mathbf{x}_i^\top \beta \tag{5}$$
$$\sigma_i = \gamma \tag{6}$$

However, as outlined above, the wind speed obtained by transforming the observed power production (v_i) is censored at cut-in and nominal wind speed (Equation 1). Thus the coefficients β and σ are not estimated with standard least squares regression but with maximum likelihood estimation with the likelihood function

$$L(\beta, \gamma | v_i, \mathbf{x}_i) = \prod_{i=1}^{n} f(v_i|\mathbf{x}_i, \beta, \gamma)^{I(v_{CI} < v_i < v_N)} P(v_i = v_{CI}|\mathbf{x}_i, \beta, \gamma)^{I(v_i = v_{CI})} P(v_i = v_N|\mathbf{x}_i, \beta, \gamma)^{I(v_i = v_N)} \tag{7}$$

where the indicator function $I(a)$ is 1 if the argument a is true and is 0 if it is not. Furthermore

$$P(v_i = v_{CI}|\mathbf{x}_i, \beta, \gamma) = P(v_i^* \leq v_{CI}|\mathbf{x}_i) = \Phi\left(\frac{v_{CI} - \mathbf{x}_i^\top \beta}{\sigma_i}\right) \tag{8}$$
$$P(v_i = v_N|\mathbf{x}_i, \beta, \gamma) = P(v_i^* \geq v_N|\mathbf{x}_i) = 1 - \Phi\left(\frac{v_N - \mathbf{x}_i^\top \beta}{\sigma_i}\right) \tag{9}$$
$$f(v_i|\mathbf{x}_i, \beta, \gamma) = \frac{1}{\sigma_i}\phi\left(\frac{v_i - \mathbf{x}_i^\top \beta}{\sigma_i}\right) \tag{10}$$

where Φ and ϕ are the cumulative distribution function and the probability density function of the standard normal distribution, respectively. With this model, conditional quantile forecasts for v_i^* can be computed with

$$q_\pi(v_i^*|\mathbf{x}_i) = \mathbf{x}_i^\top \beta + \sigma_i \Phi^{-1}(\pi). \tag{11}$$

Heteroskedastic tobit model

The standard tobit model assumes a constant residual variance σ_i over all $i = 1, \ldots, n$. This assumption can be relaxed with an additional regression equation for the standard deviation σ_i. Thus, Equation 6 is generalized to

$$\log(\sigma_i) = \mathbf{z}_i^\top \gamma \tag{12}$$

where $\mathbf{z}_i$ is an additional vector of input variables, not necessarily equal to $\mathbf{x}_i$. The log link is used to assure positive variances. All remaining Equations 4–11 can still be applied as before.

The heteroskedastic version of the tobit model is used less frequently in the literature. However, for example, Thorarinsdottir and Gneiting (2010) proposed a closely related model with the main difference being that the parameters are estimated by minimizing the continuous ranked probability score (CRPS, Wilks 2006) instead of maximizing the likelihood function. Their method is a modified version of Gneiting, Raftery, Westveld, and Goldman (2005) considering the truncation of wind speed at zero. The method of Gneiting *et al.* (2005) has proven to perform very well for temperature and precipitation forecasts (Wilks and Hamill 2007).

3.2. Nonparametric models

Nonparametric models are more flexible than parametric ones since no distribution of the response has to be assumed. Therefore, they are preferable when no good approximation of the response distribution is known. The price for this flexibility is that only specific quantiles can be estimated and that the model has to be fitted separately for each quantile. If more than one quantile is required, this means that more parameters have to be estimated.

Quantile regression

Similar to the mean in least squares regression, specific quantiles can be estimated with quantile regression. Instead of the quadratic loss function in least squares regression, Koenker and Bassett Jr (1978) proposed to weight residuals above or belowthe quantile differently, namely

$$\rho_\pi(u) = \begin{cases} u\pi & \text{if } u \geq 0 \\ u(\pi - 1) & \text{otherwise} \end{cases} \tag{13}$$

The π-quantile can be estimated by

$$q_\pi(v_i^*|\mathbf{x}_i) = \mathbf{x}_i^\top \beta_\pi \tag{14}$$

with parameters β_π minimizing

$$\sum_{i=1}^{n} \rho_\pi(v_i - q_\pi(v_i^*|\mathbf{x}_i)) \tag{15}$$

Although the censoring of the transformed power production v_i is not considered explicitly in this model, we employ it for comparison to assess the importance of censoring in the regression. Additionally, we use several benchmark models based on quantile regression for observed power production pr_i directly as these are used frequently in the wind energy literature (Bremnes 2004, 2006; Nielsen *et al.* 2006; Moller, Nielsen, and Madsen 2008). See Section 3.3 for details on the different models.

Censored quantile regression

As for the parametric models it is also possible to consider censoring with quantile regression. As suggested by Powell (1986), Equations 13 and 14 still apply and in Equation 15, $q_\pi(v_i^*|\mathbf{x}_i)$ is replaced by $q_\pi(v_i|\mathbf{x}_i)$ from Equation 2. Note that further approaches to estimate censored quantile regression exist (Portnoy 2003; Peng and Huang 2008; Lin, He, and Portnoy 2012) besides the approach of Powell (1986).

Model		Response	Regressors
tobit1	Tobit model	v_i^*	$\mathbf{x}_i = v_{NWP,i}$
tobit3	Tobit model	v_i^*	$\mathbf{x}_i = (v_{NWP,i}, v_{NWP,i}^2, v_{NWP,i}^3)$
htobit1	Heteroskedastic tobit model	v_i^*	$\mathbf{x}_i = v_{NWP,i}$, $\mathbf{z}_i = \sigma(\mathbf{v}_{EPS,i})$
htobit3	Heteroskedastic tobit model	v_i^*	$\mathbf{x}_i = (v_{NWP,i}, v_{NWP,i}^2, v_{NWP,i}^3)$, $\mathbf{z}_i = \sigma(\mathbf{v}_{EPS,i})$
rq3	Quantile reg.	v_i^*	$\mathbf{x}_i = (v_{NWP,i}, v_{NWP,i}^2, v_{NWP,i}^3)$
crq1	Censored quantile reg.	v_i^*	$\mathbf{x}_i = v_{NWP,i}$
crq3	Censored quantile reg.	v_i^*	$\mathbf{x}_i = (v_{NWP,i}, v_{NWP,i}^2, v_{NWP,i}^3)$
rq3p	Quantile reg. in power space	pr_i	$\mathbf{x}_i = (pc(v_{NWP,i}), pc(v_{NWP,i})^2, pc(v_{NWP,i})^3)$
srq3p	Quantile reg. in power space	pr_i	$\mathbf{x}_i = 3$ spline basis functions of $pc(v_{NWP,i})$
srq4wp	Quantile reg. in power space	pr_i	$\mathbf{x}_i = 4$ spline basis functions of $v_{NWP,i}$

Table 1: List of models considered. The first seven models are all estimated in wind space and all except *rq3* incorporate censoring information. The remaining models are either estimated entirely in power space (*srq3p*, *srq4wp*) or in power-by-wind space (*srq4wp*).

3.3. Choice of regressors

In wind space (see Figure 4), a simple linear model that uses NWP wind speed forecasts ($v_{NWP,i}$) as the sole regressor is certainly justifiable. However, despite the inverse transformed response, some slight remaining nonlinearities at the lower and upper end appear to remain. These are much weaker than the nonlinearities in the untransformed power-by-wind space (see Figure 2) and can be captured very well by a low-dimensional polynomial. Therefore, we consider a number of models that employ not only the linear term $v_{NWP,i}$ but additionally the corresponding squared and cubic terms, i.e., a polynomial of order 3. In addition to these regressors for the mean/quantiles of the predicted wind distribution, the heteroskedastic model also allows for regressors for the standard deviation of the wind distribution. A natural candidate is the ensemble standard deviation of the 10 meter wind speed ($\sigma(\mathbf{v}_{EPS,i})$).

Combining these ideas, we consider a number of models listed in Table 1. The tobit model with NWP wind speed forecasts as single regressor variable is the simplest model and already produces a reasonable fit of the data (see *tobit1* in Figure 5). Adding the 2nd and 3rd powers to the regressor improves the fit somewhat (*tobit3*). Neither polynomials with higher powers nor the inclusion of further NWP variables as regressors lead to further significant improvements for the data considered. Hence, we confine ourselves to linear functions and order 3 polynomials in $v_{NWP,i}$ for all models in wind space. Only in power space or power-by-wind space, stronger nonlinearities may have to be accounted for by using spline basis functions for (transformed) NWP wind speed. More specifically, we assess three benchmark quantile hmodels (*rq3p*, *srq3p*, *srq4wp*) for pr_i (i.e., replacing v_i and v_i^* with pr_i in Equations 14 and 15). As regressor variables they either use 3 polynomial basis functions of transformed NWP wind speed forecasts (*rq3p*), spline basis functions (for details see Nielsen *et al.* 2006) of transformed wind speed forecasts with 3 degrees of freedom (*srq3p*), or spline basis functions of wind speed forecasts with 4 degrees of freedom (*srq4wp*).

4. Verification

In this section, several measures are described to compare the performance of the different models. First a unique skill score is introduced in Section 4.1 to measure the value of a

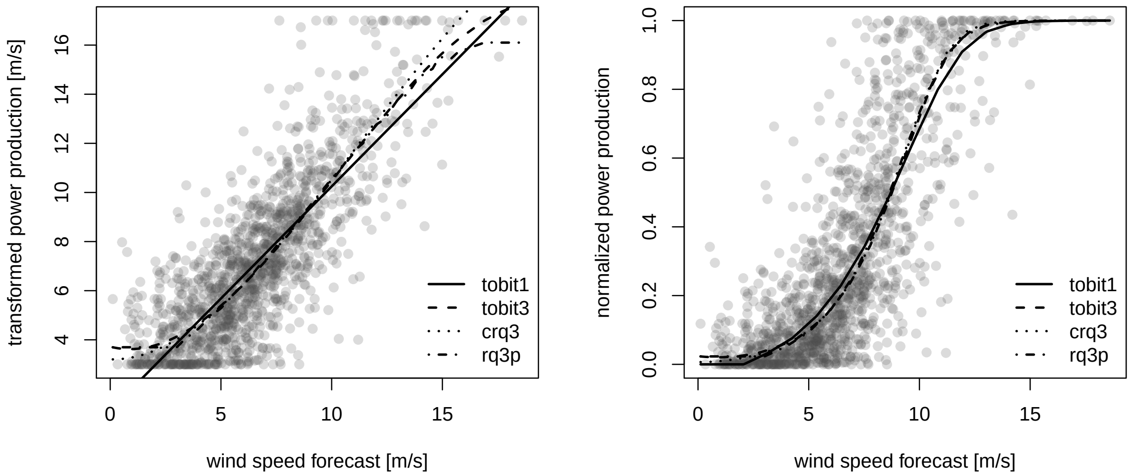

Figure 5: Different model fits (median) in wind space (left) and power-by-wind space (right).

forecast in a simplified energy market. A unique skill score is very convenient to compare the performance of different forecast methods but unfortunately cannot fully characterize the performance of a forecast (Wilks 2006). Therefore the three important properties of quantile forecasts, reliability, sharpness, and resolution, are discussed in the following subsections. Reliability is the crucial property of a good forecast that the forecast probabilities match the observed relative frequencies. A test to check whether this property is fulfilled is presented in Section 4.2. For two reliable forecasts, the one with the narrower predictive distribution is preferable. This property is termed sharpness here. Resolution is the property that the width of the predictive distribution varies between forecasts. Measures of both, sharpness and resolution are defined in Section 4.3.

We computed all measures by taking 250 bootstrap samples of our data where the bootstrap sample served as training data and the out-of-bootstrap sample aw verification data. With this approach, we get 250 values for each verification measure.

One problem when using quantile regression is that quantile crossing may occur. This would imply nonsense negative probabilities to fall between two quantiles. To avoid these negative probabilities that can cause problems in the evaluation we therefore sort the quantiles before verification. For example if the 0.2 quantile is higher than the 0.3 quantile they are interchanged.

4.1. Unique skill score – A simple market model

Since one important application for wind power forecasts is energy trading, the value of a forecast in an energy market can serve as a direct indicator of forecast performance. Instead of a real energy market, we use a simplified market model (Bremnes 2004; Roulston, Kaplan, Hardenberg, and Smith 2001): First the provider has to bid an amount $\widehat{pr}_i$ of energy. The actual production though is pr_i. The provider always receives a fee c for the energy pr_i he eventually produces. If less than the bid $\widehat{pr}_i$ is produced, a penalty c_- for each missing energy unit has to be payed. If too much is produced, each kW of surplus energy is penalized with c_+. Thus, this simple market can be described by the expected income or revenue

$$R(pr_i, \widehat{pr}_i, c, c_+, c_-) = pr_i c - \begin{cases} (\widehat{pr}_i - pr_i)c_- & \text{if } pr_i < \widehat{pr}_i \\ (pr_i - \widehat{pr}_i)c_+ & \text{if } pr_i > \widehat{pr}_i \end{cases} \tag{16}$$

In Bremnes (2004) it is shown that the expected income is maximized when $\widehat{pr}_i = q_\pi(pr_i|\mathbf{x}_i)$, with $\pi = c_+/(c_+ + c_-)$. When dividing Equation 16 by $(c_+ + c_-)$, replacing $\widehat{pr}_i$ by $q_\pi(pr_i|\mathbf{x}_i)$, and using $\pi = c_+/(c_+ + c_-)$ it can be seen that for a specific price combination c, c_- and c_+ the best forecast is the one that minimizes

$$S_{i,\pi} = (1-\pi)(q_\pi(pr_i|\mathbf{x}_i) - pr_i)^{I(pr_i < q_\pi(pr_i|\mathbf{x}_i))} + \pi(pr_i - q_\pi(pr_i|\mathbf{x}_i))^{I(pr_i > q_\pi(pr_i|\mathbf{x}_i))} \tag{17}$$

Note that this equation is equivalent to the loss function used for quantile regression (Equation 13).

A simple performance measure for wind power forecasts would be to compute the income of a specific forecast for a test data set (e.g., as in Bremnes 2004). However, to do so specific market prices have to be assumed. Because prices can vary over different markets and days we use a score that considers several price combinations:

$$S_i = \sum_{j=1}^{9} S_{i,\frac{j}{10}} \tag{18}$$

Here, small values of S_i denote good performance. The mean value of S_i over the test dataset is denoted as $\overline{S}$. Note that this score also fits into the framework of Pinson *et al.* (2006) and Gneiting and Raftery (2007) for a unique skill score.

4.2. Reliability

Reliability is the property of the forecast probabilities to be in accordance with the observed relative frequencies. For example, 75% of the observations should be on average beyond the 0.75-quantile. The set of quantile forecasts $q_{1/10}(v_i^*|\mathbf{x}_i), q_{2/10}(v_i^*|\mathbf{x}_i), ..., q_{9/10}(v_i^*|\mathbf{x}_i)$ form 10 intervals with nominal probability of 1/10 for an observations v_i to fall into one of these intervals. To test the reliability, the relative frequencies of observations falling into specific intervals can be compared with their nominal probability by a Pearson's χ^2 test as proposed by Bremnes (2006).

A problem occurs for the censored regression models when the observation falls on one of the censoring points (zero or nominal power). If one or more quantiles are below cut-in or above nominal wind speed respectively it is not clear in which interval the observation falls. Thus, in the χ^2 test such censored observations are split up proportionally into the intervals from which they may stem. To illustrate this split-up strategy, consider the following example (see also Figure 6): If the uncensored wind quantiles are $q_{1/10}(v_i^*|\mathbf{x}_i) = 2.5m/s$ and $q_{2/10}(v_i^*|\mathbf{x}_i) = 4.5m/s$ and the observation is censored at cut-in wind speed $v_{CI} = 3$ (i.e, that v_i for which $pc^{-1}(pr_i = 0)$), then it may either come from the first decile (10%) or the first quarter of the second decile ($2.5\% = (3-2.5)/(4.5-2.5)*10\%$). Thus, the first decile receives weight $0.8 = 0.1/(0.1+0.025)$ and the second decile receives $0.2 = 0.025/(0.1+0.025)$ for this event.

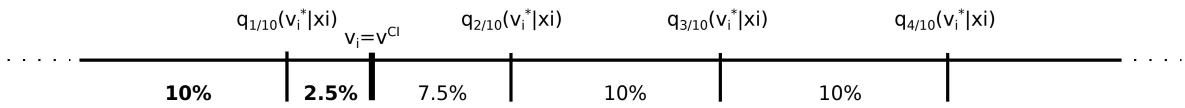

Figure 6: Schematic figure how censored observations are split up in a χ^2 test.

Note that this analysis is done in wind space (before applying Equations 2 and 3). Exceptions are the models that are estimated in power space for which the analysis is also done in power space. Like in Bremnes (2006) we declare forecasts to be unreliable if the p-value of the χ^2 test is below 0.05.

4.3. Sharpness and resolution

Sharpness and resolution are further properties that can be used to characterize forecast performance. Here, we follow the approach of Pinson *et al.* (2006) of sharpness and resolution: Define a central prediction interval as

$$\delta_{\alpha,i} = q_{(1-\alpha/2)}(pr_i|\mathbf{x}_i) - q_{\alpha/2}(pr_i|\mathbf{x}_i) \tag{19}$$

The probability of the verification to fall within this interval is α. Given a reliable forecast, it is preferable that this prediction interval is as narrow as possible whereas its width should vary strongly between different events. Narrow prediction intervals are related to a small forecast uncertainty, while highly variable prediction intervals signify that this uncertainty is situation dependent. The mean width of the prediction interval over the dataset $\overline{\delta}_\alpha$ is hereafter denoted as sharpness, while the resolution is given by the standard deviation $\sigma(\delta_{\alpha,i})$. Because large $\overline{\delta}_\alpha$ are related to large $\sigma(\delta_{\alpha,i})$, the resolution should only be compared for forecasts with similar sharpness.

5. Results

In this section the verification measures, introduced in the previous section are used to compare the performance of the different models. Since reliability is the crucial property for a

	12h	24h	36h	48h
tobit1	0.03	0.17	0.03	0.19
tobit3	0.03	0.07	0.02	0.18
htobit1	0.08	0.19	0.16	0.27
htobit3	0.07	0.16	0.13	0.23
rq3	0.05	0.05	0.04	0.07
crq1	0.18	0.12	0.15	0.15
crq3	0.19	0.10	0.14	0.11
rq3p	0.00	0.04	0.00	0.06
srq3p	0.00	0.04	0.00	0.06
srq4wp	0.02	0.06	0.01	0.09

Table 2: Median p-values of reliability test for models listed in Table 1 for different lead times.

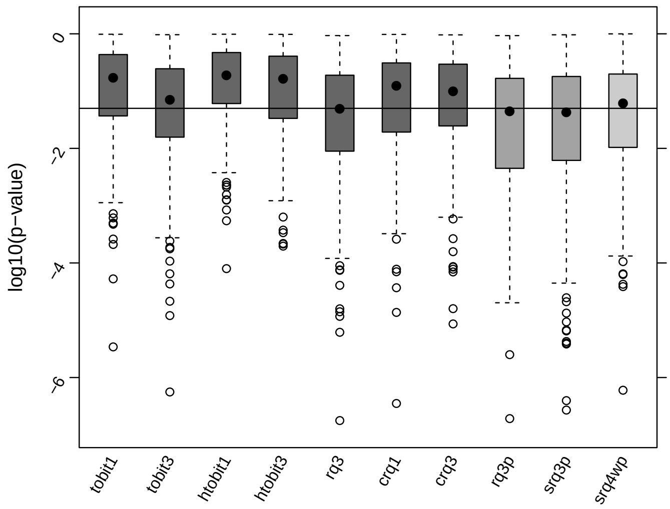

Figure 7: Reliability p-values of different models (see Table 1) for lead time 24 hours. A horizontal line is plotted for 0.05.

good probabilistic forecast, it is assessed first. Table 2 shows the medians of the 250 reliability p-values from bootstrapping of all tested models and lead times. First, it can be seen that all tobit models have worse p-values for lead times 12 and 36 hours than for 24 and 48 hours. While the heteroskedastic tobit model is still reliable for these lead times the p-value of the standard tobit model drops beyond the 0.05 level. This worse performance for 12 and 36 hours might stem from the worse performance of the ECMWF model for daytime forecasts (Drechsel *et al.* 2012). Beside the heteroskedastic tobit model, censored quantile regerssion is also reliable for all lead times. When comparing this to uncensored quantile regression it can be seen that not considering the censoring clearly deteriorates the reliability. Finally it can be seen from Table 2 that all models in the power space seem to have problems with reliability.

Similar features are shown in Figure 7 where a more detailed picture of reliability at lead time 24 hours is plotted. As in Table 2, it can be seen that all censored models in wind space (i.e., using the inverse power curve transformation) are rather reliable while the uncensored quantile regression in wind space (*crq3*) and all models in power space are not.

In Figure 8, the market score for different lead times is plotted. Not surprisingly, the market score increases with lead time. As already apparent in the reliability, the models predict more poorly for 12 and 36 hours (daytime) than for 24 and 48 hours. The reason for this are the worse forecasts of the ECMWF model for daytime. When comparing the models among each other, the differences are small and mostly not significant when compared to the uncertainty. All in all the heteroskedastic tobit model (*htobit3*) seems to be one of the best models throughout all lead times.

The sharpness and resolution for two different prediction intervals and lead time 24 hours is shown in Figure 9. One feature of this figure is that the nonparametric models have clearly better sharpness, especially for the small 0.4 prediction interval. This suggests that the assumption of a normal distribution in the parametric models does not apply perfectly.

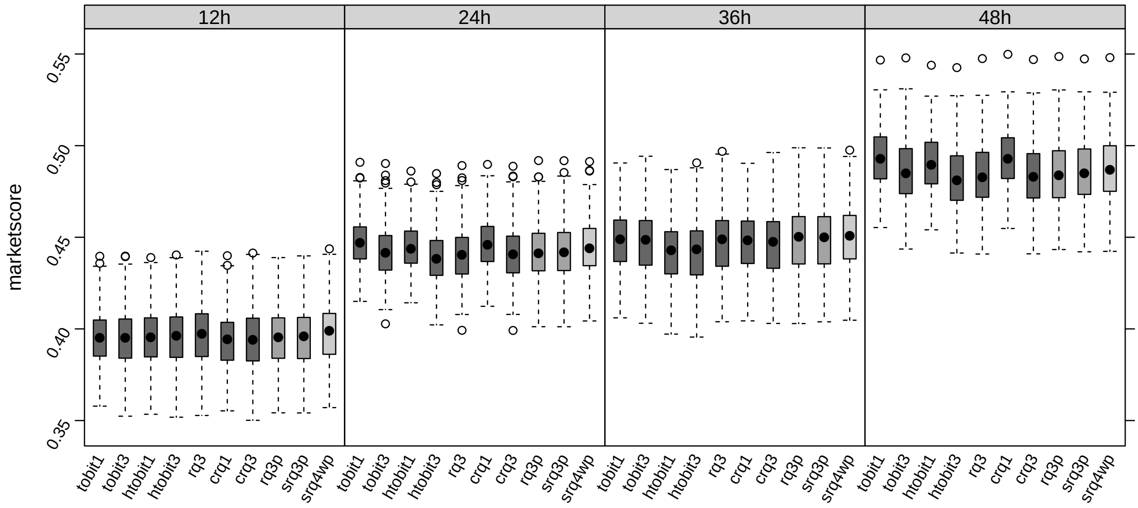

Figure 8: Market score ($\overline{S}$; smaller is better) for different models and lead times.

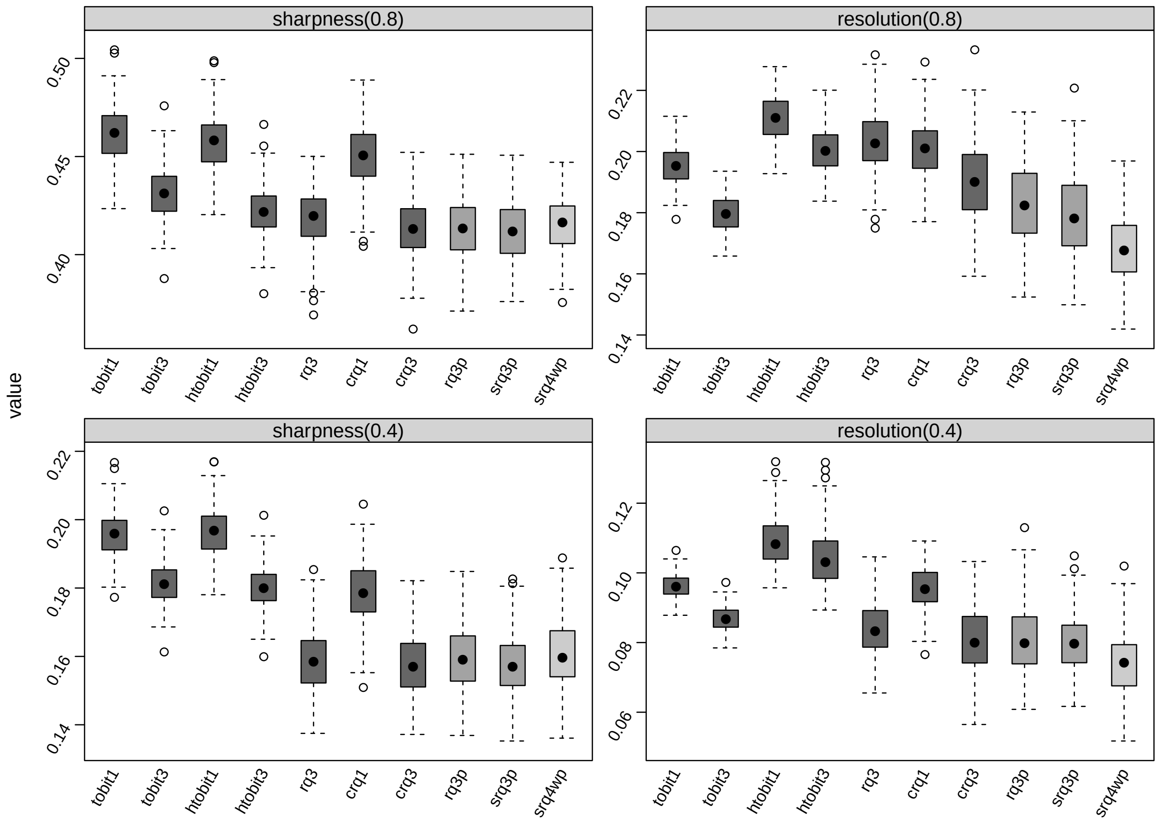

Figure 9: Sharpness (smaller is better) and resolution (larger is better) of interval forecasts with interval probabilities $\alpha = 0.4$ (bottom) and $\alpha = 0.8$ (top) for different models.

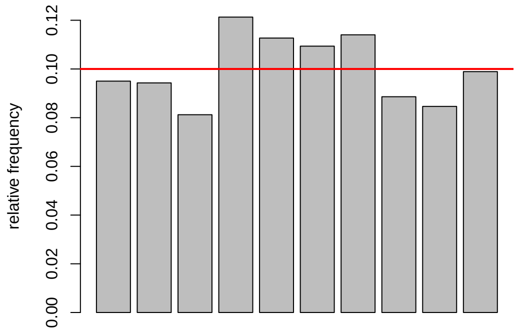

Figure 10: Relative frequencies of observations v_i falling into intervals $(-\infty, q_{1/10}(pr_i|\mathbf{x}_i)], [q_{1/10}(pr_i|\mathbf{x}_i)), q_{2/10}(pr_i|\mathbf{x}_i)], \ldots$ for the heteroskedastic tobit model (*htobit3*).

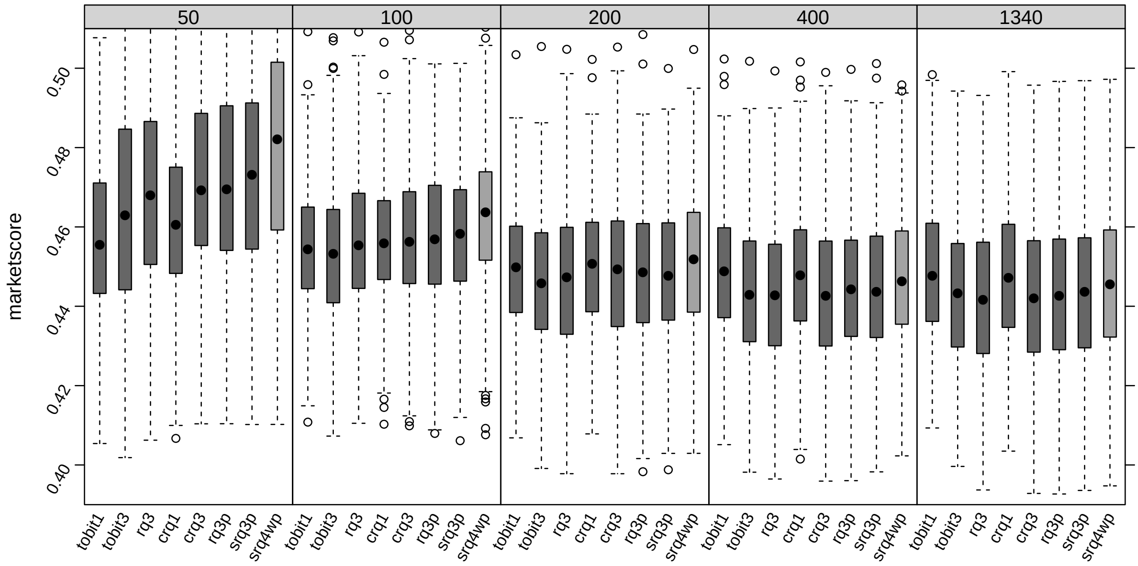

Figure 11: Marketscore ($\overline{S}$; smaller is better) for different training sample sizes and lead time 24 hours.

Figure 10 shows the relative frequencies of observations falling into the intervals formed by the predicted deciles for the heteroskedastic tobit model. It can be seen that the observations fall slightly too often into intervals in the center and too rarely into intervals in the margins. This suggests that in fact the response follows a distribution with somewhat heavier tails than the normal distribution. Probably because too few data are available, the differences in the distribution cannot be seen in the market score and are too small for the null hypothesis of reliability to be rejected.

Note that the resolution can only be compared for models with similar sharpness like the tobit and heteroskedastic tobit models. Comparing the resolution of these two models it can be seen that the resolutions of the heteroskedastic tobit models are clearly better. This is not surprising since the size of the prediction interval of the tobit model is basically constant and varies solely because of the censoring. On the other hand the heteroskedastic tobit model has variable interval sizes by construction. Censored quantile regression and the models in power space also have similar sharpnesses, so that their resolutions can be compared as well. Censored quantile regression is slightly better than the other models, especially for the large 0.8 interval.

Finally, we show a plot of market scores for different training sample sizes in Figure 11. Clearly the performance increases with a larger training sample. Fewer parameters have to be estimated for the parametric models . Therefore it is not surprising that they perform better than the nonparametric models if only few data are available for fitting. The spline model with the completely untransformed data (*rqs4wp*) has the most degrees of freedom and is therefore the worst of all models for small training sample sizes. While for very small training sample sizes the simplest tobit model (*tobit1*) seems to be the best, the heteroskedastic tobit model is already best for training sample sizes ≥ 100. However, as for the full dataset, the differences are mostly not significant compared to the uncertainty.

6. Conclusion

A combination of new approaches for improving probabilistic wind power forecasts is proposed: (1) Exploit the readily available information from the power curve of the turbine to transform observed power production to wind speed (inverse power curve transformation). (2) Respect the censoring of power production between zero and nominal power (in power space) or between cut-in and nominal wind speed (in wind space) with censored regression models. The resulting combined strategy has the advantage that almost all nonlinearity and heteroskedasticity of the observations is directly captured. Consequently, relatively simple linear regression models with normally distributed responses can be used.

To assess this new strategy, a wide range of combinations of parametric and nonparametric regression models, with and without inverse power curve transformation, with and without censoring information is considered for data from a wind turbine in Austria. For all models, wind speed forecasts and its transformations are used as regressor variables and, furthermore, some heteroskedasticity models additionally use the standard deviation of the ECMWF ensemble forecasts. It is shown that the censored regression models obtained in wind space with the inverse-transformed power production are more reliable than uncensored regression models in all spaces considered (i.e., in wind space, power space, and power-by-wind space). As for the comparison of parametric vs. nonparametric censored models in wind space, it can be shown that the more parsimonious parametric models already perform well for relatively small training samples while the nonparametric models perform somewhat better in large training samples. However, the performance of the parametric models may potentially be improved in future work by using a response distribution with heavier tails (e.g., logistic or Student-t instead of Gaussian) so that resolution and sharpness can be enhanced.

In addition to data from the wind turbine presented in this manuscript, data from another wind turbine were assessed but not presented as they lead to very similar outcomes. Hence, similar results can be expected for other turbines/regions but, of course, this still has to be tested in future work. One special feature of the tested turbines is that the wind speeds are relatively small and thus right-censoring (at nominal speed/power) does not play an important role although it is supported by our models. Furthermore, switching off the turbine because of too high wind speed did never happen in our data and is therefore not considered in our models. For turbines where this plays a role it would generally be possible to consider an additional right censoring. Instead of using the manufacturer's power curve, an empirical power curve computed from observation data (Cabezon, Marti, San-Isidro, and Perez 2004) could be used as well. This is particularly important if forecasts for entire wind parks are required which consist of different types of turbines. Different to most other studies we used a global numerical weather prediction (NWP) model instead of a limited area model. However regarding Louka, Galanis, Siebert, Kariniotakis, Katsafados, Pytharoulis, and Kallos (2008) and Müller (2011) we do not expect large differences in the results.

Computational details

Our results were obtained on Ubuntu and Debian GNU/Linux using R 2.15.1 (R Core Team 2012) and packages quantreg 4.81 (Koenker 2012) for (censored) quantile regression, dcrq 1.0 (Lin *et al.* 2012) for doubly censored quantile regression, and numerical optimization of the likelihood for the (heteroskedastic) tobit models via `optim()` with `method = "BFGS"`. A proper package for the latter is under development but the code is also available upon request in the meantime.

Acknowledgments

This study was supported by the Austrian Science Fund (FWF): L615-N10. We are also very grateful to *WEB Windenergie AG* for providing the wind turbine data.

References

Bremnes JB (2004). "Probabilistic Wind Power Forecasts Using Local Quantile Regression." *Wind Energy*, **7**(1), 47–54. ISSN 1095-4244.

Bremnes JB (2006). "A Comparison of a Few Statistical Models for Making Quantile Wind Power Forecasts." *Wind Energy*, **9**(1-2), 3–11. ISSN 1099-1824.

Cabezon D, Marti I, San-Isidro M, Perez I (2004). "Comparison of Methods for Power Curve Modeling." In *Proceedings to Global Windpower 2004*. Chicago.

Drechsel S, Mayr GJ, Messner JW, Stauffer R (2012). "Wind Speeds at Heights Crucial for Wind Energy: Measurements and Verification of Forecasts." *Journal of Applied Meteorology and Climatology*, **51**(9), 1602–1617. ISSN 1558-8424.

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C (2011). "The State-Of-The-Art in Short-Term Prediction of Wind Power – A Literature Overview." *Technical report*, ANEMOS. plus.

Glahn H, Lowry D (1972). "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting." *Journal of Applied Meteorology*, **11**, 1203–1211.

Gneiting T, Raftery AE (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association*, **102**(477), 359–378. ISSN 0162-1459.

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. ISSN 0027-0644.

Juban J, Fugon L, Kariniotakis G (2007). "Probabilistic Short-Term Wind Power Forecasting Based on Kernel Density Estimators." In *Proceedings of the 2007 European Wind Energy Conference*, pp. 7–10. Milan.

Koenker R (2012). *quantreg: Quantile Regression*. R package version 4.91, URL http://CRAN.R-project.org/package=quantreg.

Koenker R, Bassett Jr G (1978). "Regression Quantiles." *Econometrica*, **46**(1), 33–50. ISSN 0012-9682.

Lange M (2005). "On the Uncertainty of Wind Power Predictions – Analysis of the Forecast Accuracy and Statistical Distribution of Errors." *Journal of Solar Energy Engineering*, **127**(2), 177–184.

Lin G, He X, Portnoy S (2012). "Quantile Regression with Doubly Censored Data." *Computational Statistics & Data Analysis*, **56**(4), 797–812. ISSN 01679473.

Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Pytharoulis I, Kallos G (2008). "Improvements in Wind Speed Forecasts for Wind Power Prediction Purposes Using Kalman Filtering." *Journal of Wind Engineering and Industrial Aerodynamics*, **96**(12), 2348–2362. ISSN 01676105.

Moller J, Nielsen H, Madsen H (2008). "Time-Adaptive Quantile Regression." *Computational Statistics & Data Analysis*, **52**(3), 1292–1303. ISSN 01679473.

Müller MD (2011). "Effects of Model Resolution and Statistical Postprocessing on Shelter Temperature and Wind Forecasts." *Journal of Applied Meteorology and Climatology*, **50**(8), 1627–1636. ISSN 1558-8424.

Nielsen H, Madsen H, Nielsen T (2006). "Using Quantile Regression to Extend an Existing Wind Power Forecasting System with Probabilistic Forecasts." *Wind Energy*, **9**(1-2), 95–108. ISSN 1099-1824.

Nielsen HA, Madsen H, Nielsen TS, Badger J, Giebel G, Landberg L, Sattler K, Feddersen H (2004). "Wind Power Ensemble Forecasting." In *Proceedings of the 2004 Global Wind Power Conference*. Chicago.

Peng L, Huang Y (2008). "Survival Analysis with Quantile Regression Models." *Journal of American Statistical Association*, **103**, 637–649.

Pinson P, Kariniotakis G, Nielsen TS, Madsen H (2006). "Properties of Quantile and Interval Forecasts of Wind Generation and their Evaluation." In *Proceedings of the 2006 European Wind Energy Conference*. Athens.

Portnoy S (2003). "Censored Quantile Regression." *Journal of American Statistical Association*, **98**, 1001–1012.

Powell JL (1986). "Censored Regression Quantiles." *Journal of Econometrics*, **32**(1), 143–155.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Roulston M (2003). "Using Medium-Range Weather Forcasts to Improve the Value of Wind Energy Production." *Renewable Energy*, **28**(4), 585–602. ISSN 09601481.

Roulston M, Kaplan D, Hardenberg J, Smith L (2001). "Value of the ECMWF Ensemble Prediction System for Forecasting Wind Energy Production." In *Proceedings of the 2001 European Wind Energy Conference*. Copenhagen.

Roulston MS, Smith L (2003). "Combining Dynamical and Statistical Ensembles." *Tellus A*, **55**(1), 16–30. ISSN 0280-6495.

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: Series A*, **173**(2), 371–388. ISSN 09641998.

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

Wilkes J, Moccia J, Dragan M (2012). "Wind in Power: 2011 European Statistics." *Technical report*, European Wind Energy Association (EWEA).

Wilks D (2006). *Statistical Methods in the Atmospheric Sciences*. 2nd edition. Academic Press.

Wilks DS, Hamill TM (2007). "Comparison of Ensemble-MOS Methods Using GFS Reforecasts." *Monthly Weather Review*, **135**(6), 2379–2390. ISSN 0027-0644.

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

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/

Jochen Broecker
Department of Mathematics and Statistics
University of Reading
Whiteknights
Reading RG6 6AX, United Kingdom
E-mail: J.Broecker@reading.ac.uk

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

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

2013-01 **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

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:** To reward the best or to punish the worst? A comparison of two tournament mechanisms with 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
- 2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints?
- 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 *forthcoming in Kyklos*

University of Innsbruck

Working Papers in Economics and Statistics

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

Abstract
Forecasting wind power is an important part of a successful integration of wind power into the power grid. Forecasts with lead times longer than 6 hours are generally made by using statistical methods to postprocess forecasts from numerical weather prediction systems. Two major problems that complicate this approach are the nonlinear relationship between wind speed and power production and the limited range of power production between zero and nominal power of the turbine. In practice, the nonlinearity is often tackled by using nonlinear nonparametric regression methods while the limited range is typically not addressed explicitly. However, such an approach ignores valuable and readily available information: the power curve of the turbine's manufacturer. Much of the nonlinearity can be directly accounted for by transforming the observed power production into wind speed via the inverse power curve so that simpler linear regression models can be used. Furthermore, the limited range of the transformed power production can be easily exploited by adopting censored regression models. In this study, we evaluate quantile forecasts from a range of methods: (a) using parametric and nonparametric models, (b) with and without the proposed inverse power curve transformation, and (c) with and without censoring. The results show that with our inverse (power-to-wind) transformation, simpler linear regression models with censoring perform equally or better than nonlinear models with or without the frequently used wind-to-power transformation.

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