DOI: https://doi.org/10.37075/EA.2025.1.06 Economic Alternatives, 2025, Issue 1, pp. 109-129

Interval, Quantile and Density Forecasts

Received: 04.10.2023
Available online: 31.03.2025

Mihail Yanchev*

Abstract

This text is a detailed review of the academic literature on interval, quantile, and density forecasting. Confidence and prediction intervals and different approaches to their estimation are discussed. The concept of quantile regression is examined as a standalone method as well as an initial step to generating density forecasts. Various methods for generation and evaluation of density forecasts and some noteworthy applications are considered.

Keywords: Economic forecasting, Density Forecasting, Interval Forecasting, Quantile Regression

JEL: B40, C10, C40

1. Introduction

Towards the end of the 19th century, there was a transition from point estimates to distribution estimates in the field of statistics according to Stigler (1975). Gneiting (2008) describes a similar shift in interest from point forecasts to probabilistic forecasts across many fields, and economics makes no exception. The review of forecasting literature done by Diebold and Lopez in 1996 reveals that when it comes to forecast evaluation, the topic of point forecast evaluation dominated the field at the time. Few articles were concerned with the evaluation of prediction intervals (Chatfield, 1993; Christoffersen, 1998) or probability forecasts (Wallis, 1993; Clemen et al., 1995). Furthermore, Diebold et al. (1998) believe that until the advent of quantitative finance and risk management, there was little demand for interval or density forecasts within the field of economics. The practice of forecasting in itself is an attempt to study an uncertain future, and probabilistic forecasts expressed as probability distributions over expected future realizations are a prime way to measure the degree of uncertainty.

According to the engineering and machine learning literature, uncertainty can be classified into two broad categories when it comes to forecasting in general – aleatoric and epistemic (see Hora, 1996, Faber, 2005, Dutta, 2013, Shaker and Hüllermeier, 2020, and Hüllermeier and Waegeman, 2021). Aleatoric uncertainty relates to the inherent uncertainty in the data-generating process and the uncertainty in its measurement. Its main property is its irreducibility in principle. Epistemic uncertainty relates to the limitations and knowledge of the forecaster and can be separated into two main subcategories – model uncertainty and approximation uncertainty (Yanchev, 2023). Approximation uncertainty relates to the uncertainty surrounding the model parameters and can be expressed as the difference between a chosen hypothesis or model and the optimal hypothesis within the chosen hypothesis

* Sofia University St Kliment Ohridski – FEBA

space (family of models). Model uncertainty refers to the choice of the hypothesis space or family of models in general and can be expressed as the difference between the ground truth (or the population model) and the optimal hypothesis within the hypothesis space. Epistemic uncertainty can be reduced by acquiring more knowledge and information, and in the context of forecasting, this might include accumulating more data and a greater number of predictor variables. The forecasting methods discussed in the rest of the text will be considered with respect to this classification.

The aim of this text is to perform a detailed review of methods for generating interval, quantile and density forecasts. Confidence and prediction intervals and different approaches to their estimation are discussed. The concept of quantile regression is examined as a standalone method as well as an initial step to generating density forecasts. Various methods for generation and evaluation of density forecasts are considered as well as some noteworthy applications.

2. Confidence and Prediction Intervals

Currently, in the field of forecasting the simplest way to measure uncertainty related to a forecast is via confidence and prediction intervals. Chatfield (1993) describes interval forecasts as consisting of upper and lower limits associated with a predefined probability. These upper and lower limits define the range in which a future value of the random variable would fall with some level of confidence. Hansen (2006) elaborates that interval forecasts are often constructed around point forecasts as an additional measure of uncertainty. Indeed, as many forecasting methods are tailored towards generating point forecasts, calculating intervals is a straightforward way to quantify the uncertainty around such forecasts.

Going a step back, it is appropriate to review the terminology concerning confidence and prediction intervals. According to Hyndman (2013)[1], the confidence interval is associated with a parameter (which can be a model parameter or a coefficient) and contains with some degree of confidence the true parameter of the population. Based on this description we could assume that the confidence interval is a measure of epistemic uncertainty and more specifically what was called earlier approximation uncertainty since it describes how certain we are about the range of values a certain parameter might take and in general the model's ability to approximate the data. On the other hand, prediction intervals are associated with a random variable to be observed in the future, with a specified probability of the random variable lying within the interval (Hyndman, 2013). Prediction intervals account for both the aleatoric uncertainty inherent in the data and the epistemic approximation uncertainty because it already contains the confidence interval for the mean response[2], but also an additional term accounting for the variability of the dependent variable observed in the sample. According to the literature on the subject, the term prediction interval concerning

[1] The explanation is taken from Hyndman's blog post: https://robjhyndman.com/hyndsight/intervals/

[2] Mean response and predicted outcome are the values of the dependent variables calculated from the regression parameters and the independent variable. The values of the two model outputs are the same, but their hypothesized variances are different. This is the reason for the differences in the widths of the confidence and prediction intervals.

a point forecast should be equivalent to the term interval forecast.

The summary outlined by Dybowski and Roberts (2001) will be used for the formulations. To be more specific let us define a linear regression forecasting task in matrix notation:

$$\hat{Y} = X\beta \quad (1)$$

Where X is a vector of k predictors and a first column of ones, β is a vector of k+1 regression parameters, and $\hat{Y}$ is the vector of the target variable Y. The confidence interval for the estimated coefficients β would be respectively:

$$\boldsymbol{\beta} \pm \boldsymbol{t}_{(\alpha/2, n-(k+1))} \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}} \quad (2)$$

Where $(X^T X)^{-1}_{jj}$ is the j-th diagonal of $(X^T X)^{-1}$. In this way, one defines the uncertainty around the parameter which is reflected in both the confidence intervals around the mean response and the prediction interval. The confidence interval around a mean response, which is a confidence interval around the estimated value $\hat{y}$ given a vector of specific values of the independent variables X_0 can be estimated in the following way:

$$\hat{Y}_0 \pm t_{(\alpha/2, n-(k+1))} \sqrt{\hat{\sigma}^2 \, (X_0^T (X^T X)^{-1} X_0)} \quad (3)$$

where X_0 is a column vector of specific values of the independent variables with a first element equal to one and $\hat{\sigma}^2$ is formulated as:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-(k+1)} (Y_t - \hat{Y}_t)^T (Y_t - \hat{Y}_t) \\ &= \frac{1}{n-(k+1)} e^T e \\ &= MSE \end{aligned} \quad (4)$$

and $t_{(\alpha/2, n-(k+1))}$ is the critical value for the t-distribution for a specified significance level α. The assumptions for using this formulation of the confidence interval include linearity and independently, identically and normally distributed errors, but deviations from the normality assumption are acceptable in the presence of large samples.

Similarly, to incorporate the aleatoric uncertainty around the target variable, another MSE term is added when calculating the prediction interval:

$$\hat{Y}_0 \pm t_{(\alpha/2, n-(k+1))} \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2 \, (X_0^T (X^T X)^{-1} X_0)} \quad (5)$$

The additional term added in the prediction interval guarantees that the prediction interval is always wider than the confidence interval for the mean response. The assumptions for constructing the prediction interval in this way include linearity and independently, identically and normally distributed errors, but the formula depends strongly on the normality assumption.

This constitutes the most basic way to construct confidence intervals for simple linear regression. The two confidence intervals for the mean response and the prediction intervals are illustrated in Figure 1 below. The solid represents the regression line, the inner dashed lines represent the 95% confidence interval based on equation (3) and the outer dotted lines represent the 95% prediction intervals based on equation (5).

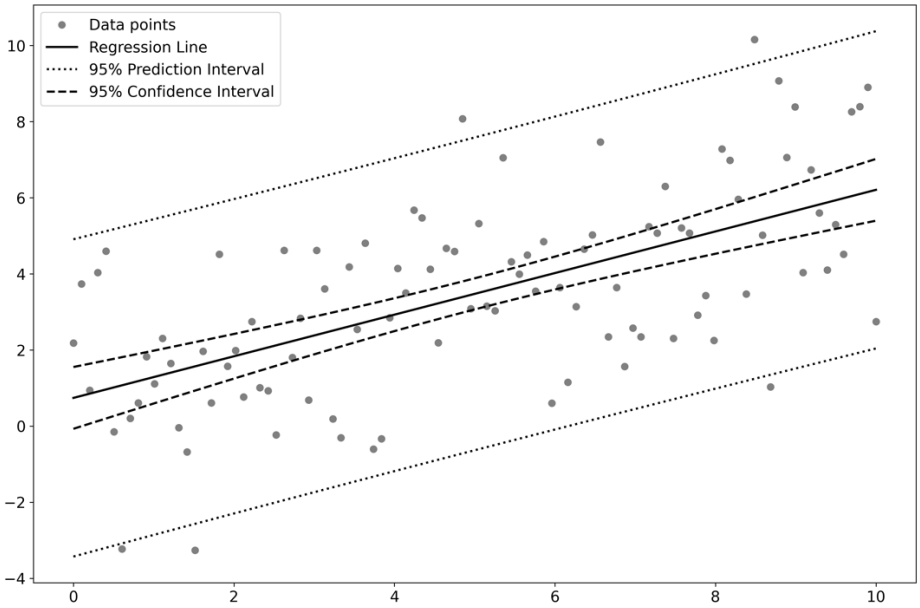

Figure 1. Confidence and Prediction Intervals for Simple Linear Regression Using Randomly Generated Data

Source: Author

Numerous studies present generalized ways to deal with confidence and prediction intervals in different contexts. Cox (1975) established the calculation of parametric prediction intervals for large samples with applications in constructing approximate confidence intervals in empirical Bayes estimation. Chatfield (1993) does a comprehensive review and comparison of different strategies for generating prediction intervals including using analytical solutions derived from fitted statistical models, approximate prediction interval formulas, analysis of the empirical forecasting errors, and via bootstrapping or simulation. Heskes (1996) provides a technique for calculating prediction and confidence intervals for feed-forward artificial neural networks. Dybowski and Roberts (2001) also present a technique for calculation confidence and prediction intervals for feed-forward network models but also do a comprehensive bottom-up review of methods from the classical foundations, which in itself can be rather useful. Hansen (2006) focuses on refining the construction of interval forecasts, which incorporate parameter uncertainty and are applied to models where the error term is independent of the predictors. Lee and Scholtes (2014) develop a procedure for generating prediction intervals from Box-Jenkins (ARIMA) processes including additional external predictors.

However, while intervals are a quantification of uncertainty they do provide limited information compared to probability distributions and do not allow for separating aleatoric and epistemic uncertainty in a straightforward way. Moreover, the calculation of intervals often requires strong parametric assumptions about the errors or residuals. Williams and Goodman (1971) argue that in practice often these assumptions do not

hold. Therefore, non-parametric approaches like bootstrapping and simulations might be preferable when dealing with calculating prediction intervals. One such example is discussed in Staszewska-Bystrova (2011) for vector autoregressive models.

When it comes to evaluating and comparing interval forecasts, Christoffersen (1998) established a framework with optimality conditions for evaluation especially when some of the generic assumptions are violated. Askanazi et al. (2018) revisit the topic and discuss the difficulty of comparing and evaluating interval forecasts. They then proceed to advise abandoning the practice of interval forecasting in favor of density forecasting, which provides richer information and can be more readily compared using proper scoring rules.

3. Quantile Regression

Quantile regression is a concept, which dates back to the 18th century. However, it was more recently re-introduced by Koenker and Bassett (1978) and applied in economic analysis in various studies. According to Fitzenberger et al., (2002) who presented a number of economic studies which utilized quantile regression, it was not until the 1990s that the technique gained larger popularity among economists and econometricians.

More recent studies utilizing the method for forecasting are Ma and Pohlman (2008), who used quantile regression to forecast returns on financial markets and define an alternative approach to portfolio construction, Gaglianone and Lima (2011) construct forecasts using quantile regression on a sample of unemployment forecasts from the Survey of Professional Forecasters (SPF), Huang et al. (2011), who utilized quantile regression to predict the volatility of exchange rate on data from various countries, Maciejowska et al. (2016), who studied accuracy improvements from quantile regression forecasting of electricity spot prices on data from the British market, Wan et al. (2016), who performed wind power generation forecasting on data from Denmark. A disproportionate part of the literature related to using quantile regression for forecasting is on topics related to energy generation and load forecasting.

As Koenker and Bassett (1978) show, a task of sorting can be turned into an optimization problem. Just as finding a sample mean can be done by minimizing the sum of squared errors, finding the median can result from minimizing the sum of absolute errors. Koenker and Bassett (1978) further elaborate to show that an asymmetrical loss function which gives different penalties to positive and negative residuals, can yield any quantile for a given sample. Solving for the following equation (6) yields the τ-th quantile as its solution:

$$\boldsymbol{min}_{\xi \in \mathbb{R}} \sum_{i=0}^{n} \boldsymbol{\rho_\tau}(\boldsymbol{y_i} - \boldsymbol{\xi}) \tag{6}$$

Where $0 < \tau < 1$ and $\rho_\tau(\cdot)$ is the tilted absolute value function, which can be seen in Figure 2, for a sample of size *n*. According to this equation, if τ is equal 0.5, the equation will yield the median. Therefore, if the scalar ξ in equation (6) is replaced with a parametric function $\xi(x_i, \beta)$ and τ is set to equal 0.5, one could estimate the conditional median function.

$$\boldsymbol{min}_{\xi \in \mathbb{R}} \sum_{i=0}^{n} \boldsymbol{\rho_\tau}(\boldsymbol{y_i} - \boldsymbol{\xi}(\boldsymbol{x_i}, \boldsymbol{\beta})) \tag{7}$$

Setting τ to different values will lead to the estimation of different conditional quantiles.

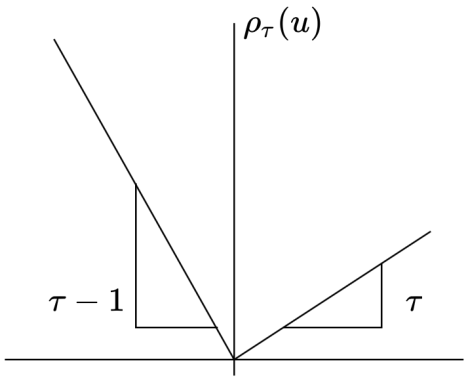

Figure 2. Tilted absolute value function
Source: Author, based on Koenker (2005)

In general, we would model the relation between the conditional quantile of the predicted output y_{t+h} and a vector of predictors X, for a given period t and a forecasting horizon *h*. To estimate the quantile regression of y_{t+h} on X, the regression coefficients β_τ for a given parameter τ is chosen to minimize the weighted absolute value of errors:

$$\hat{\beta}_\tau = \operatorname{argmin}_{\beta_\tau \in \mathbb{R}^k} \sum_{t=1}^{T-h} (\tau \cdot 1_{(y_{t+h} \geq X\beta_\tau)} |y_{t+h} - X\beta_\tau| + (1-\tau) \cdot 1_{(y_{t+h} < X\beta_\tau)} |y_{t+h} - X\beta_\tau|) \quad (8)$$

where $\mathbf{1(\cdot)}$ the indicator function, which subsets negative and positive errors, and T is the total length of the time series. The output of the model is the quantile of y_{t+h} conditional on the model input X:

$$\hat{Q}_{y_{t+h}|X}(\tau|\mathrm{X}) = X\beta_\tau \quad (9)$$

This method allows one to estimate a quantile regression model to estimate any arbitrary quantile, conditional on the predictors. However, if one would like to estimate several different quantiles, one might run into the so-called crossing problem, which multiple scholars have run into and tried to address in one way or another (see Koenker, 1984; Cole and Green, 1992; He, 1997; Bondell et al., 2010; Rodrigues and Pereira 2020). Among the more interesting solutions are the ones proposed by Bondell et al. (2010) and Rodrigues and Pereira 2020.

Bondell et al. (2010) propose an additional term to the tilted cost function, which directly addresses the crossing quantile problem. The additional penalty for crossing quantiles can be described in the following way:

$$\sum_{j=1}^{J-1} max(0, X\hat{\beta}_{\tau_j} - X\hat{\beta}_{\tau_{j+1}}) \quad (10)$$

where J is the number of quantiles sorted by the increasing value of τ. This term can be added to the loss function described in equation (8).

Rodrigues and Pereira (2020) propose a multi-output deep learning approach for estimating multiple conditional quantiles jointly again in order to address the problem of crossing quantiles. The authors' proposition is to aggregate the loss function for the separate quantiles and evaluate it for all conditional quantiles jointly at every step of the optimization process. As the authors further elaborate in this case one could have a model that outputs an arbitrary number of quantiles, which allows one to construct a prediction interval of arbitrary coverage. However, one has to be careful, because following the authors' method the prediction interval would account for aleatoric uncertainty alone, as it does not implicitly or explicitly contain information about model or approximation uncertainty.

It is outlined how one could use the conditional quantiles generated using a quantile regression as a measure of aleatoric uncertainty. However, it is possible to use the estimated conditional quantiles in a two-

step procedure, to generate a full predictive distribution. A study mentioned above by Gaglianone and Lima (2011) constructs conditional quantiles based on ensemble forecasts from the SPF and then in a second step estimates nonparametrically a density forecast either using a method provided by Koenker (2005) or through quantile interpolation via a kernel. The kernel function used by the authors is the Epanechinikov kernel.

In a seminal paper, Adrian et al. (2019) used a two-step procedure of estimating conditional quantiles using quantile regression and fit a probability distribution to the estimated quantiles. The authors studied the distribution of economic growth in the USA with a focus on financial conditions and their dynamics during economic downturns. The authors pointed out a number of stylized facts about the conditional distribution of economic growth in the USA among which a strong negative correlation between the conditional mean and variance and a strong link between current financial conditions and future shifts in the lower tail of the distribution. Similar conclusions were confirmed by De Santis and Van der Veken (2020), who performed a similar empirical study including data from the beginning of 2020 and a separate dataset covering the Spanish flu pandemic period across a number of countries. Figueres and Jarociński (2020) also confirmed the same stylized facts identified by Adrian et al. (2019) for the Euro Area.

Quantiles of the conditional distribution of GDP growth in this framework are expressed as functions of the observed predictors. After generating the conditional quantiles, one could fit a probability distribution function to them to generate a density forecast. Adrian et al. (2019) propose using a skewed t-distribution for this purpose. To estimate the four parameters related to the skewed t-distribution, the problem can be formulated as a least squares optimization problem, using the estimated conditional quantiles[3] and the inverse cumulative probability function:

$$\{\hat{\mu}_{t+h}, \hat{\sigma}_{t+h}, \hat{\nu}_{t+h}, \hat{\alpha}_{t+h}\} = \underset{J}{\operatorname{argmin}_{\mu,\sigma,\nu,\alpha}} \sum_{j=1} \left(\hat{Q}_{y_{t+h}|X}(\tau_j|X) - F^{-1}(\tau_j; \mu, \sigma, \nu, \alpha) \right)^2 \quad (11)$$

where $\hat{\mu}_{t+h} \in \mathbb{R}$ (mean or location shift), $\hat{\sigma}_{t+h} \in \mathbb{R}^+$ (standard deviation or scaling parameter), $\hat{\nu}_{t+h} \in \mathbb{R}$ (skewness parameter), and $\hat{\alpha}_{t+h} \in \mathbb{R}^+$ (kurtosis or tail weight parameter). F^{-1} is the inverse cumulative distribution function and $\hat{Q}_{y_{t+h}|X}(\tau_j|X)$ is the estimated quantile of y_{t+h} for a given τ and conditional on X. This method can be used to estimate a density based on the conditional quantiles, or observed (unconditional) quantiles of a given economic indicator.

The established procedure used by Adrian et al. (2019) is a two-step procedure. Its first step consists of a quantile regression model with a loss function similar to equation (8), which is used to generate conditional quantiles. The second step uses the conditional quantiles as an input and performs a least squares optimization between the input and the inverse CDF of the distribution of choice (in this case the skewed t-distribution). Adrian et al. (2019) apply two alternative approaches to demonstrate that the results from the two-step procedure are reliable. They use fully parametric and fully non-parametric approaches in order to compare and find very similar characteristics of the resulting

[3] The .05, .25, .75, and .95 quantiles are used for the estimation of the conditional distribution.

conditional distributions. They conclude that the two-step procedure is less parametric, less hard-coded, and perhaps offers greater versatility.

The advantage of the approach by Adrian et al. (2019) compared to Gaglianone and Lima (2011) is that by assuming a specific family of distribution one can more concretely analyze the dynamics of the distribution parameters like variance, skewness, and kurtosis. The nonparametric approach allows for a more limited analysis in this respect. On the other hand, the nonparametric approach of Galianone and Lima (2011) does not necessitate making an assumption about the probability distribution of the data-generating process.

It is important to mention that the concept of the quantile is not only useful in the context of quantile regression. Quantile analysis is useful in many fields like economics, finance, and risk management and is a primary way to analyze an arbitrary distribution of data. For example, the estimation of unconditional quantiles can be used to construct intervals of arbitrary coverage around a random variable. Also, if bootstrapping is performed on a certain statistical parameter or model output, the quantiles of the resulting distribution can be used to construct confidence intervals. This only shows how intertwined the three general concepts discussed in this paper are.

By reviewing the literature on quantile regression and its application in economics, it was established that the estimation of conditional quantiles can be used both to construct arbitrary upper and lower bounds for a prediction interval or as an input for the estimation of density functions. Quantile regression seems to be somewhat underutilized in economics compared to other traditional methods, however, its practical benefits and versatility are undeniable.

4. Density forecasts

Early works in meteorology were the first to recognize the benefits of probabilistic forecasting. Cooke (1906) and Von Myrbach (1913) advocated for the need to attach a degree of confidence to meteorological forecasts. Later on in a seminal paper, Brier (1944) elaborated on the need for probabilistic forecasts, which led to the way to gradual adoption of probabilistic methods in meteorology. For several decades, meteorology researchers worked towards developing methods for generating probabilistic forecasting, as well as refining the ways to evaluate them. Murphy and Winkler (1984) provide a great summary of the important developments in the field for most of the 20th century.

A density forecast is a forecast expressed as a probability distribution, instead of a single value, which would be considered a point forecast. Point forecasts are often a central feature of a probability distribution like the conditional mean or conditional median, which can be arrived at by optimizing the loss functions of respectively mean squared errors and mean absolute errors. Density forecasts can be expressed as the parameters that describe a probability distribution or as they are formally called the moments of the distribution. For many families of probability distributions like the normal skewed distribution, the skewed t-distribution, and the Sinh-Arcsinh distribution these include the mean (or location), the variance (or scale), the skewness and kurtosis (the last two are sometimes referred to as shape parameters). In the time series context, which is predominant in economics and

econometrics a density forecast over horizon *h*, is expressed as the forecasted moments of a probability distribution for each time within the horizon.

Following Yanchev (2023), a density forecast can be defined in the following way. $\mathcal{D}^T = (Y, X, X_{T+h})$ is the information set available at the time of forecasting. Y is vector of the target variable or dependent variable. X is a vector of the predictors or the independent variables up to time T and X_{T+h} is a vector of predictors available after time T in order to generate forecasts Y_{T+h} for a forecast horizon with length h. The predictive distribution $f(Y_{T+h}|\mathcal{D}^T)$ is what is referred to as a density forecast since it describes the distribution of the target variable Y for future values and is conditional on the information set that is available to the forecaster.

Following Bassetti et al. (2019), a basic method for generating a density forecasting could be demonstrated by using a multiple linear regression model without an intercept for convenience:

$$Y_t = X_t^T \beta + \varepsilon_t \tag{12}$$

where *t = 1, ..., T* and $\varepsilon_t \sim i.i.d.(0, \sigma^2)$. β is a (m × 1) vector of coefficients, σ^2 is the variance of the error term ε_t, X_t is a (m × 1) vector of covariates or predictors, which can include exogenous variables z_t and lagged values of the dependent variable, $Y_{t-p}, p > 0$. A direct method to compute a density forecast is to assume the distribution for the error term and ignore parameter uncertainty. A usual assumption is the one of normality - $\varepsilon_t \sim N(0, \sigma^2)$. This would account for aleatoric uncertainty but ignore the epistemic one. The *h*-step ahead density forecast, conditional on the information available in the information set up to time *T* would be:

$$f(Y_{T+h}|\mathcal{D}^T) = N(X_T^T \hat{\beta}, \hat{\sigma}^2) \tag{13}$$

where $\hat{\beta}$ and $\hat{\sigma}^2$ can be computed either analytically or numerically. In this case, the variance is fixed for a given estimation of the model. As Bassetti et al. (2019) point out there are several ways to account for parameter uncertainty in this case. Hansen (2006) offers a closed-form solution for a linear model. One could use a residual-based bootstrapping following Davidson and Hinkley (1997), which essentially resamples the residuals from the initial estimation. The main disadvantage of this method is that it treats the residuals as i.i.d., which is often not the case. In order to account for autocorrelation and heteroskedasticity in the residuals, the block wild bootstrapping method proposed by Yeh (1998) can be used. The basic premise behind block bootstrapping is to divide the time series into a number of blocks with a block length based on the lag length of the autocorrelation. The wild block bootstrapping proposed by Yeh (1998) allows the block length to vary instead of being fixed.

An alternative approach to accounting for parameter uncertainty involves using Bayesian inference (Dunson et al., 2007; Bassetti et al. 2019). Based on Bayes theorem one could formulate prior distribution on the parameters, which multiplied by the likelihood results in the parameter posterior distributions. Based on the linear regression in equation (12), the objective of Bayesian inference would be to calculate the forecasted density:

$$f(Y_{T+h}|\mathcal{D}^T) = \int p(Y_{T+h}, X_T, \theta|\, \mathcal{D}^T) d\theta$$
$$= \int l(Y_{T+h}|X_T, \theta, \mathcal{D}^T) \mathrm{p}(\theta|\, \mathcal{D}^T) d\theta \tag{14}$$

where $\theta = (\beta, \sigma^2)$ is a ((m+1) × 1) vector of model parameters, $\mathcal{D}^T = (Y, X_T)$ is the information set, $l(Y_{T+h}|X_T, \theta, \mathcal{D}^T)$ is the likelihood of the model for time T+h, $\mathrm{p}(\theta|\, \mathcal{D}^T)$ is the parameter marginal distribution conditional

on the information set. Depending on the choice of the parameter prior distribution, if the prior is conjugate then the parameter posterior distribution and the predictive distribution can be computed analytically. For non-conjugate prior, the posterior and predictive distributions need to be evaluated using numerical methods like Monte Carlo simulation methods. This approach would account for both aleatoric uncertainty and epistemic uncertainty. However, due to the initial setup of the model the variance parameter is fixed across time, which is not suitable for non-normal data-generating processes with skewness or heavy tails.

A seminal work in economics by Engle (1982) aimed to model and forecast volatility in the stock market. The autoregressive conditional heteroskedasticity (ARCH) model is explicitly modeling the conditional variances. The model expresses conditional variance as a linear function of squares of past observations and in this way generates forecasts with time-varying conditional variances. Bollerslev (1986) defined a generalized version of the ARCH model – the GARCH model. In its simplest form, the generalized model can be formulated in the following way in scalar notation:

$$y_t = \sigma_t \epsilon_t \tag{15}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{16}$$

where $\epsilon_t \sim i.i.d.(0,1)$. This formulation assumes the time series are stationary with zero mean. However, this formulation of the conditional variance can be coupled with a variety of models for the conditional mean – a constant mean model as in the formulation above or any version of an ARIMA model. Since the seminal paper by Engle, there have been numerous extensions of the model (see Nelson, 1991; Higgins and Bera, 1992; Ding, Engle, and Granger, 1993; Glosten, Jagannathan, and Runkle, 1993; Zakoian, 1994).

A significant benefit of the GARCH model is that the assumption of symmetric normality is not necessary. One could use a family of skewed distribution as well. However, its main setback is that fundamentally it is a univariate framework for modeling and forecasting. There have been extensions to accommodate exogenous regressors (see Sharma et al., 1996; Engle and Patton, 2001), although the literature on such extensions is limited.

A different approach of modeling explicitly all moments of the distribution of a target variable is defined by the generalized additive models for location, scale, and shape - GAMLSS (Rigby and Stasinopoulos, 2005). In this model, the moments of a given distribution are modeled via separate equations, which are estimated jointly either via maximum likelihood estimation or Bayesian methods. Following the formulation of Stasinopoulos et al. (2018), the GAMLSS model can be described as follows in scalar notation:

$$g_1(\mu) = X_1\beta_1 + s_{11}(x_{11}) + \cdots + s_{1J_1}(x_{1J_1}) \tag{17}$$

$$g_1(\sigma) = X_2\beta_2 + s_{21}(x_{21}) + \cdots + s_{1J_2}(x_{2J_2}) \tag{18}$$

$$g_1(\upsilon) = X_3\beta_3 + s_{31}(x_{31}) + \cdots + s_{1J_3}(x_{3J_3}) \tag{19}$$

$$g_1(\tau) = X_4\beta_4 + s_{41}(x_{41}) + \cdots + s_{1J_4}(x_{4J_4}) \tag{20}$$

assuming that the target follows a four-parameter distribution $y \sim D(\mu, \sigma, \upsilon, \tau)$, υ and τ are shape parameters associated with skewness and kurtosis. The X_1, X_2, X_3, X_4 are sets of regressors or predictors, $s()$ stands for smoothing non-parametric functions applied to some of the explanatory variables. As the

authors point out the use of the smoothing function defines a data-driven approach to determine the relationship between the explanatory variables and the target, instead of enforcing a specific type of relationship.

There is a limited but growing body of applications of this type of model within the field of economics. Serinaldi (2011) uses it for short-term forecasting of electricity prices based on data from the USA and Italy. Gilchrist et al. (2011) used GAMLSS to forecast movie box-office revenues. Mikis et al. (2021) applied the GAMLSS to study the Greek-German bond yield spreads in a period up to the onset of the sovereign bond crisis in Europe. Ziel (2021) outlines the use of the GAMLSS in the M5 competition and specifically for the probabilistic forecasting of the sales of Walmart's retail goods. Regis et al. (2022) model asset prices on the Brazilian stock market. Lastly, Umlauf et al. (2017) extend the GAMLSS and define the Bayesian version of the GAMLSS called the BAMLSS, which utilizes the Markov chain Monte Carlo (MCMC) simulation techniques.

Similar to the GAMLSS framework, one could model the moments of a target variable's distribution using artificial neural networks. TensorFlow released a separate package dealing specifically with probabilistic forecasting, which according to Ziel (2021) is very similar to the core idea of the GAMLSS models. Since artificial neural networks allow for a flexible definition of inputs and outputs of the network, one could define a multi-output network, which outputs the moments of the distribution of the target variable, and network parameters are estimated using maximum likelihood. In this spirit,

Gal and Ghahramani (2016) propose a theoretical framework called Monte Carlo dropout (MC dropout) using a regularization technique called dropout (see Srivastava et al. 2014) as approximate Bayesian inference in Gaussian processes. This technique is generally used only during training to turn off neurons at random within a layer of the network, which is a strategy for preventing overfitting. However, the authors extend the use of dropout during inference, which can be used to generate a sample of predictions, hence its reference to Monte Carlo simulations. One could then fit a distribution to this sample and thus generate a density forecast. The simplicity of the idea and its implementation is tempting, although there have been researchers who criticize the claim that it approximates Bayesian inference (see Osband, 2016; Folgoc et al. 2021).

Salinas et al. (2019) present the DeepAR model which is an autoregressive recurrent neural network, which performs probabilistic forecasting and is specifically tailored to forecast a large number of time series. The model is estimated by applying gradient descent through time optimization (see Rumelhart, Hinton, Williams, 1986; Williams, Zipser, 1992) to maximize log-likelihood. The authors make an assumption for Gaussian likelihood for real-valued continuous target variables, and negative-binomial likelihood for positive count data. Other recent advancements in probabilistic forecasting using deep learning can be found in Wen et al. (2018), who use a mixed architecture between recurrent and convolutional layers to jointly estimate an arbitrary number of quantiles, Wang et al. (2019) combine a local method with a global neural network with the aim of forecasting a large number of time series, and Rangapuram et al. (2018) parametrize a linear state-space model, implemented via a Kalman filter in the context of a recurrent neural network architecture. Alexandrov et

al. (2020) in presenting their forecasting package for Python called GluonTS, mention implementations of the transformer (Vaswani et al., 2017) and wavenet (van den Oord et al., 2016) architectures, which have been very successful in natural language processing, to the task of time series probabilistic forecasting.

Due to the fact that these innovations in deep learning have been relatively recent, the economics literature is yet to implement them at scale. Few studies can be found that implement any of the mentioned models or similar methods based on deep learning in the context of economic forecasting. He and Li (2018) apply a quantile regression neural network in combination with kernel density estimation to forecast wind power generation on data from Canada. Alghamdi et al. (2021) use deep learning and MC dropout to forecast Amazon and Apple stock prices. Barbaglia et al. (2021) use news information as an additional source of information and apply the DeepAR to model the spread between the 10-year and 3-month government treasury bills. Similarly, Consoli et al. (2022) extract information from news and use the DeepAR model to forecast the Spanish IBEX-35 stock market index. The last two studies use text data from news articles as well, which is expected to become a growing trend in economics. As Goulet Coulombe et al. (2022) point out the non-linearity which many machine learning methods allow for is where the main benefit is of applying such methods to economic modeling and forecasting. Therefore, one could only expect that their adoption would become gradually more widespread, especially as new tools come out to increase their explainability and transparency (see Bharghava and Gupta, 2022).

Another approach for generating density forecasts, which is used in economics and finance relies on copula theory. Copulas allow for the decomposition of a joint probability distribution of a number of correlated variables into uncorrelated marginals and a function, which specifies the correlation between the variables – namely the copula. Patton (2013) provides an extensive overview of the methodology for forecasting using copula methods. More recent applications of these methods can be found in Bessa et al. (2011) on the topic of wind power forecasting and He et al. (2017) on the topic of power load density forecasting, although these methods are often used in conjunction with other traditional modeling techniques.

Lastly, a somewhat popular procedure for generating density forecasts often used in economics is finding a probability distribution, which is the best fit to a sample of pre-existing forecasts. Considering a panel of separate forecasts like the forecasts published by the SPF, one could fit a probability distribution function around the panel of forecasts (see Diebold et al. 1998; Gaglianone and Lima, 2011). This method necessitates however that one is in possession of such a dataset, which is usually available only for important aggregate indicators like GDP, unemployment, and inflation. A similar exercise can be followed if one is in possession of multiple forecasts based on different methods or procedures for the same indicator (see Barnard, 1963, Roberts, 1965, Bates and Granger, 1969). One could use the distribution of the individual forecasts as a density forecast.

5. Evaluation of Density Forecasts

Forecast evaluation for point forecasting is somewhat straightforward. When evaluating the performance of a forecasting method, one

is in possession of some ground truth values and the generated forecasts. See Gneiting (2011) for a discussion on the topic of the evaluation of point forecasts. The evaluation of the performance relies on choosing an appropriate performance metric, which is usually based on the distance between the ground truth and the forecast. Such metrics are the mean squared errors or the mean absolute errors and while there are some variations of these, the main premise is the same.

With density forecasting the task of evaluating a forecast is harder, because one needs to compare forecast or predictive densities with a single ground truth value. Therefore, numerous studies have been focusing on developing and refining methods for density forecast evaluation. As Gneiting et al. (2007) define two separate aspects in which density forecast needs to be evaluated – calibration and sharpness. Calibration refers to the statistical consistency between the density forecasts and the observed, while sharpness is understood as the concentration of the density forecasts. A higher concentration of the density forecasts is better, subject to calibration, as more concentrated forecasts imply higher precision, lower forecast uncertainty, and a higher degree of model confidence. In an extreme example, the "sharpest" possible forecast would be a deterministic forecast, where the predictive distribution assigns the total likelihood to a single value and no likelihood to any other value. Therefore, a more concentrated forecast is desirable, but only if calibration is fulfilled as well. As Mitchell and Wallis (2011) point out, sharpness is a property of the predictive distributions alone, while calibration is a property of the forecast-observation pairs.

Following Gneiting et al. (2007), a basic theoretical framework is laid out in order to define the context of the task of evaluating density forecasts. At time steps *t = 1,...,T* there is a distribution G_t, which we think of as the true data-generating process, and the forecaster generates a probabilistic forecast in the form of a predictive CDF F_t. The ground truth outcome x_t is a random number with distribution G_t. Gneiting et al. (2007) refer to the distribution of the data-generating process as being "picked by nature" and therefore it is assumed that the forecaster's set of information is at most that of nature. Therefore, if:

$$F_t = G_t \tag{21}$$

for all time steps then F_t is an ideal forecast. The true distribution G_t is hypothetical and remains unknown. Dawid (1984) and Diebold et al. (1998) proposed the use of the probability integral transform (PIT):

$$p_t = F_t(x_t) \tag{22}$$

for making this comparison. If the forecasts are ideal and F_t is continuous, then p_t has a uniform distribution. This can be examined visually via a histogram of the PIT or a uniformity test. Gneitling et al. (2007) summarize how one should interpret deviations from uniformity in the histogram. Concave-down-shaped (hump-shaped) histograms indicate overdispersed predictive distributions with greater dispersion on average. This would be equivalent to forecasts lacking enough sharpness. Concave-up-shaped (U-shaped) histograms often point to predictive distributions that are excessively concentrated or narrow, which means they are too sharp. Triangle-shaped histograms, which are skewed in one direction or the other, indicate biased predictive distributions.

In terms of formal statistical testing, Smith (1985) proposes using the inverse CDF (quantile function) on the PIT values and test for normality, since there are a number of established tests for normality. If $z_t = \Phi^{-1}(p_t)$ and $p_t \sim U(0,1)$ then $z_t \sim N(0,1)$. On the other hand, the goodness of fit tests like Pearson's chi-squared test can assess how well the PIT histogram fits a uniform distribution using the p_t or the transformed z_t values. Other suitable tests include the Kolmogorov-Smirnov (KS) test, and its Anderson-Darling (AD) modification (Mitchell and Wallis, 2011).

Thus, the uniformity of the PIT is a necessary condition for the forecaster to be ideal, and checks for its uniformity have become a fundamental forecast evaluation technique. However, Hamill (2001) presented an example of uniformly distributed PIT values derived from forecasts where every single forecast is biased. The example aimed to show that the uniformity of the PIT values is a necessary, but not a sufficient condition for the forecaster to be ideal Gneiting et al. (2007). This would imply that checking for uniformity in the PIT does not allow one to compare between an ideal forecaster and competitor forecasters, since all of them can exhibit necessary uniformity. Mitchell and Wallis (2011) argue that this example is unrealistic and misleading to the existing literature on evaluating density forecasts.

Scoring rules assign numerical scores to probabilistic forecasts based on the predictive distribution and the realization of the forecasted variable. They conveniently summarize the predictive performance, when the quality of a probabilistic forecast is evaluated. They address both calibration and sharpness (Gneitling et al., 2007). Assuming that a score $s(F,x)$ is a penalty that the forecaster would like to minimize, a score is considered proper if its value is minimized for an observation x drawn from G when $F = G$. A strictly proper score would be one with a unique minimum (Gneitling et al., 2007). According to Mitchell and Wallis (2011), sharpness became an established term in the forecasting literature with the decomposition of the Brier score (Brier, 1950) into two components by Sanders (1963) - one measuring "validity" and the other "sharpness".

The logarithmic score is the negative of the logarithm of the PDF of the predictive density evaluated at the observation (Good, 1952; Bernardo, 1979). The logarithmic score is proper according to the definition above (Roulston and Smith, 2002), but according to Selten (1998) and Gneiting and Raftery (2006), it lacks robustness. According to Selten (1998), sometimes it can exhibit excessive sensitivity with respect to differences between very small probabilities, while in other cases, it would be not sensitive enough with respect to the distance between the truth and the prediction.

$$logS(x_t) = logf_t(x_t) \quad (23)$$

One could consider the logarithmic score as the logarithm of the predictive likelihood, and thus two forecasts can be compared using the Bayes factor or the difference in their logarithmic scores (see Geweke and Amisano, 2010). An alternative scoring rule, which is often used in literature is the continuous ranked probability score (CRPS), which is directly defined in terms of the predictive CDF:

$$crps(F,x) = \int_{-\infty}^{\infty} \{F(y) - \mathbf{1}(y \geq x)\}^2 dy \quad (24)$$

Where $1()$ is the indicator function and according to Gneiting et al. (2007) is a more robust alternative to the logarithmic score. Gneiting and Raftery (2006) propose an

alternative representation, which is useful when dealing with forecast ensembles and shows that the CRPS generalizes the absolute error. According to Gneiting et al. (2007), the CRPS is a proper score and competing forecasts can be ranked based on their average:

$$CRPS = \frac{1}{T}\sum_{t=1}^{T} crps(F_t, x_t) = \int_{-\infty}^{\infty} BS(y)dy \quad (25)$$

where

$$BS(y) = \frac{1}{T}\sum_{t=1}^{T}\{F_t(y) - \mathbf{1}(x_y \leq y)\}^2 \quad (26)$$

denotes the Brier score (1950) for probabilistic forecasts of the binary event at the threshold value $y \in \mathbb{R}$. Like all proper scoring rules for binary probability forecasts (Gneiting et al., 2007), the Brier score allows for the distinction of a calibration and sharpness components (see Murphy, 1972; DeGroot and Fienberg, 1983; Dawid, 1986). A discussion of calibration-sharpness decomposition of the CRPS can be found in Candille and Talagrand (2005). Both the logarithmic score and the CRPS can be used for comparing probabilistic forecasts, but the latter has some advantages in terms of robustness and decomposition into calibration and sharpness.

Several studies employ an alternative method for comparing probabilistic and quantile forecasts called the pinball loss (see Maciejowska et al., 2016; Abramova and Bunn, 2020; Berrisch and Ziel, 2022). The pinball loss is essentially equivalent to the tilted loss function, also called the quantile loss function. The pinball loss function is always positive and results in greater values the further away the estimate quantile is from the observation, one is comparing the forecast against (Abramova and Bunn, 2020). For the case of a predictive density

$$PS_t(q_\tau, y) = (\tau - 1)\sum_{y_t < \Phi_\tau^{-1}(p_t)} (y_t - \Phi_\tau^{-1}(p_t)) + \tau \sum_{y_t \geq \Phi_\tau^{-1}(p_t)} (y_t - \Phi_\tau^{-1}(p_t)) \quad (27)$$

If one is evaluating a full predictive density, one could use the inverse CDF to generate the continuum of quantiles on the basis of which to calculate the pinball loss. The generated quantiles are considered the best fit when the values of the pinball loss are the lowest. This could similarly to the PIT histogram point out deficiencies related to bias, overdispersion, or too narrow predictive distributions, but in the context of comparing between forecasts.

Lastly, it is worth evaluating the accuracy of the central features of the predictive distributions like the mean and median. This can be performed by treating these central features like point forecasts. Therefore, traditional metrics for point-forecast accuracy like MSE, RMSE, or MAE are valid in this respect.

6. Conclusion

A clear trend towards probabilistic forecasting is observed in the scientific literature, due to the evident advantages over point-forecasts discussed as early as the late 60s of the 20th century (Anscombe, 1968), but even more often after 2000 (Granger and Pesaran, 2000, Tay and Wallis, 2000). Many studies in economics and even in machine learning still focus on point forecasting. On the other hand, many researchers as well as institutions like central banks have already recognized the advantage of using density forecasts. However, the practical tools for generating and evaluating density forecasts

are still not as accessible and established as the instruments now widely used for point forecasting.

In the economics literature, the prevailing approaches to probabilistic forecasting include generating intervals for traditional point forecasts, quantile regression, Bayesian methods, and various modifications of the GARCH model. The use of machine learning and deep learning methods, copula methods, and other mixed methods like quantile-based density forecasting seems to be still a road less traveled. Two distinct sub-fields where novel methods are often applied are energy and financial economics, where it seems the demand for ever more accurate forecasting drives innovation.

References

Abramova, E., Bunn, D., 2020. Forecasting the intra-day spread densities of electricity prices. *Energies*, 13(3), 687.

Adrian, T., Boyarchenko, N., Giannone, D., 2019. Vulnerable Growth. *American Economic Review*, 109, pp. 1263-1289.

Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., Maddix, .D., Rangapuram, S., Salinas, D., Schulz, J., Stella, L., Türkmen, A. C., Wang, Y., 2020. Gluonts: Probabilistic and neural time series modeling in python. *The Journal of Machine Learning Research*, 21(1), pp. 4629-4634.

Alghamdi, D., Alotaibi, F., & Rajgopal, J., 2021. A Novel Hybrid Deep Learning Model For Stock Price Forecasting. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Anscombe, F. 1967. Topics in the investigation of linear relations fitted by the method of least squares. *Journal of the Royal Statistical Society*, Series B, 29(1), pp. 1-52

Asimakopoulos, S., Paredes, J., Warmedinger, T., 2013. Forecasting fiscal time series using mixed frequency data. European Central Bank, Working Papers Series, No 1550 / May 2013

Askanazi, R., Diebold, F. X., Schorfheide, F., Shin, M., 2018. On the Comparison of Interval Forecasts. *Journal Time Series Analysis* 39, pp. 953–965.

Barbaglia, L., Consoli, S., & Manzan, S., 2021. Exploring the predictive power of news and neural machine learning models for economic forecasting. In Mining Data for Financial Applications: 5th ECML PKDD Workshop, MIDAS 2020, Ghent, Belgium, September 18, 2020, Revised Selected Papers 5 (pp. 135-149). Springer International Publishing.

Barnard, G. A., 1963. New methods of quality control. *Journal of the Royal Statistical Society*. Series A (General), 126(2), pp. 255-258.

Bassetti, F., Casarin, R., Ravazzolo, F., 2019. Density Forecasting, No BEMPS59, BEMPS - Bozen Economics & Management Paper Series, Faculty of Economics and Management at the Free University of Bozen

Bates, J. M., Granger, C. W., 1969. The combination of forecasts. *Journal of the Operational Research Society*, 20(4), pp. 451-468.

Bernardo, J. M., 1979. Expected information as expected utility. the Annals of Statistics, pp. 686-690.

Bessa, R. J., Mendes, J., Miranda, V., Botterud, A., Wang, J., Zhou, Z., 2011. Quantile-copula density forecast for wind power uncertainty modeling. 2011 IEEE Trondheim PowerTech, Trondheim, Norway, pp. 1-8

Bokan, N., Ravnik, R., 2012. Estimating Potential Output in the Republic of Croatia Using a Multivariate Filter, Croatian National Bank Working Papers W-35.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), pp. 307-327.

Bondell, H. D., Reich, B. J., Wang, H., 2010. Noncrossing quantile regression curve estimation. – *Biometrika*, 97, pp. 825-838.

Brier, G., 1944. Verification of a forecaster's confidence and the use of probability statements in weather forecasting. Research Paper No. 16, U.S. Weather Bureau.

Brier G., 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78, pp. 1–3.

Candille, G., & Talagrand, O., 2005. Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(609), pp. 2131-2150.

Chatfield, C., 1993. Calculating interval forecasts. *Journal of Business & Economic Statistics*, 11(2), pp. 121-135.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International economic review, pp. 841-862.

Clemen, R. T., Murphy, A. H., Winkler, R. L., 1995. Screening probability forecasts: contrasts between choosing and combining. *International Journal of Forecasting*, 11(1), pp. 133-145.

Cole, T. J., Green, P. J., 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, 11(10), pp. 1305–1319.

Consoli, S., Tiozzo Pezzoli, L., Tosetti, E., 2022. Neural forecasting of the Italian sovereign bond market with economic news. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(Supplement_2), S197-S224.

Cooke, W. E., 1906. Weighting Forecasts, *Monthly Weather Review*, 34, pp. 274-275

Cox, D. R., 1975. Partial Likelihood. *Biometrika*, 62(2), pp. 269–276.

Dawid, A. P., 1984. Statistical theory: the prequential approach (with discussion). *Journal of the Royal Statistical Society*, A, 147(2), pp. 278-292

Dawid, A. P., 1986. Probability forecasting. In Encyclopedia of Statistical Sciences, vol. 7 (eds S. Kotz, N. L. Johnson and C. B. Read), pp. 210–218. New York: Wiley.

DeGroot, M. H., Fienberg, S. E., 1983. The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society*: Series D (The Statistician), 32(1-2), pp. 12-22.

De Santis, R. A., Van der Veken, W., 2020. Forecasting macroeconomic risk in real time: great and Covid 19 recessions. – European Central Bank Working Paper Series No 2436 / July 2020.

Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4), pp. 863-883

Diebold, F. X., Lopez, J. A., 1996. 8 Forecast evaluation and combination. Handbook of statistics, 14, pp. 241-268.

Ding, Z., Granger, C. W., Engle, R. F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1), pp. 83-106.

Dunson, D. B., Pillai, N., Park, J. H., 2007. Bayesian density regression. *Journal of the Royal Statistical Society Series B*: Statistical Methodology, 69(2), pp. 163-183.

Dutta, P., 2013. An Approach to Deal with Aleatory and Epistemic Uncertainty within the Same Framework: Case Study in Risk Assessment. IJCA 80, pp. 40–45

Dybowski, R., Roberts, S. J., 2001. Confidence Intervals and Prediction Intervals for Feed-Forward Neural Networks In Dybowski R. & Gant ,V. (eds.), Clinical Applications of Artificial Neural Networks, Cambridge University Press 2001, pp. 298-326.

Engle, R. F., Patton, A. J., 2001. What good is a volatility model?. *Quantitative Finance*, 1(2), 237.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, pp. 987-1007.

Faber, M.H., 2005. On the Treatment of Uncertainties and Probabilities in Engineering Decision Analysis. *Journal of Offshore Mechanics and Arctic Engineering* 127, pp. 243–248.

Fitzenberger, B., Koenker, R., Machado, J. A. (Eds.). 2001. Economic applications of quantile regression. Springer Science & Business Media.

Folgoc, L. L., Baltatzis, V., Desai, S., Devaraj, A., Ellis, S., Manzanera, O. E. M., Nair, A., Qui, H., Schnabel, J., Glocker, B, 2021. Is MC Dropout Bayesian?. arXiv preprint arXiv:2110.04286.

Gaglianone, W.P., Lima, L.R., 2012. Constructing Density Forecasts from Quantile Regressions. *Journal of Money, Credit and Banking* 44, pp. 1589–1607

Geweke, J., Amisano, G., 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2), pp. 216-230.

Gilchrist, R., Stasinopoulos, D., Rigby, R., Sedgwick, J., Voudouris, V., 2011. Forecasting film revenues using GAMLSS. In 26th International Workshop on Statistical Modelling (IWSM), València.

Gneiting, T., 2008. Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society Series* A, Royal Statistical Society, Vol. 171(2), pp. 319-321, April.

Gneiting, T., 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), pp. 746-762.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society Series B*, 69(2), pp. 243-268

Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), pp. 359-378.

Good I., 1952. Rational decisions. *Journal of the Royal Statistical Society B* 14, pp. 107–114.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., Surprenant, S., 2022. How is machine learning useful for macroeconomic forecasting?. *Journal of Applied Econometrics*, 37(5), pp. 920-964.

Granger, C., Pesaran, M., 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7), pp. 537-560

Hamill T., 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review* 129, pp. 550–560.

Hansen, B. E., 2006. Interval forecasts and parameter uncertainty. *Journal of Econometrics*, 135(1-2), pp. 377-398.

He, X., 1997. Quantile Curves without Crossing. *The American Statistician*, 51(2), pp. 186–192.

He, Y., Li, H., 2018. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. *Energy conversion and management*, 164, pp. 374-384.

He, Y., Liu, R., Li, H., Wang, S., Lu, X., 2017. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory. Applied Energy 185, pp. 254–266.

Heskes, T., 1996. Practical confidence and prediction intervals. *Advances in neural information processing systems*, 9.

Hora, S., 1996. Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. *Reliability Engineering and System Safety*, 54(2–3), pp. 217–223

Hüllermeier, E., Waegeman, W., 2021, Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine Learning 110, pp. 457–506

Koenker, R., 2005. Quantile Regression. Cambridge University Press. pp. 146

Koenker, R., Bassett, G., 1978. Regression Quantiles. – Econometrica, 46, p. 33.

Koenker, R., Hallock, K. F.. 2001. Quantile Regression. *Journal of Economic Perspectives*, 15 (4), pp. 143-156.

Lee, Y. S., Scholtes, S., 2014. Empirical prediction intervals revisited. *International Journal of Forecasting*, 30(2), pp. 217-234.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pp. 347-370.

Maciejowska, K., Nowotarski, J., Weron, R., 2016. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *International Journal of Forecasting*, 32(3), pp. 957-965.

Mikis, D. S., Robert, A. R., Nikolaos, G., Fernanda, D. B., 2022. Principal component regression in GAMLSS applied to Greek–German government bond yield spreads. *Statistical Modelling*, 22(1-2), pp. 127-145.

Mitchell, J., Wallis, K. F., 2011. Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, 26(6), pp. 1023-1040.

Murphy, A. H., 1972. Scalar and vector partitions of the probability score: Part I. Two-state situation. *Journal of Applied Meteorology* (1962-1982), pp. 273-282.

Murphy, A. H., Winkler, R. L., 1984. Probability forecasting in meteorology. *Journal of the American Statistical Association*, 79(387), pp. 489-500.

Osband, I., 2016. Risk versus uncertainty in deep learning: Bayes, bootstrap and the dangers of dropout. In NIPS workshop on Bayesian deep learning (Vol. 192).

Patton, A., 2013. Copula Methods for Forecasting Multivariate Time Series, in: Handbook of Economic Forecasting. Elsevier, pp. 899–960

Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., Januschowski, T., 2018. Deep state space models for time series forecasting. Advances in neural information processing systems, 31.

Regis, R. O., Ospina, R., Bernardino, W., Cribari-Neto, F., 2022. Asset pricing in the Brazilian financial market: five-factor GAMLSS modeling. *Empirical Economics*, 1-37.

Rigby, R. A., Stasinopoulos, D. M., 2005. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C* (Applied Statistics), 54(3), pp. 507-554.

Roberts, C. F., 1965. On the use of probability statements in weather forecasts (No. 1). Environmental Science Services Administration, Weather Bureau.

Rodrigues, F., Pereira, F., 2020. Beyond Expectation: Deep Joint Mean and Quantile Regression for Spatiotemporal Problems. – IEEE Transactions on Neural Networks and Learning Systems, 31, pp. 5377-5389.

Roulston, M. S., Smith, L. A., 2002. Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, 130(6), pp. 1653-1660.

Sanders F., 1963. On subjective probability forecasting. *Journal of Applied Meteorology* 2, pp. 191–201.

Selten, R., 1998. Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1, pp. 43-61.

Serinaldi, F., 2011. Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. *Energy Economics*, 33(6), pp. 1216-1226.

Shaker, M.H., Hüllermeier, E., 2020. Aleatoric and Epistemic Uncertainty with Random Forests, in: Berthold, M.R., Feelders, A., Krempl, G. (Eds.), Advances in Intelligent Data Analysis XVIII, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 444–456.

Sharma, J. L., Mougoue, M., Kamath, R., 1996. Heteroscedasticity in stock market indicator return data: volume versus GARCH effects. *Applied Financial Economics*, 6(4), pp. 337-342.

Smith J., 1985. Diagnostic checks of non-standard time series models. *Journal of Forecasting* 4, pp. 283–291.

Smith, J., Wallis, K. F., 2009. A simple explanation of the forecast combination puzzle. Oxford bulletin of economics and statistics, 71(3), pp. 331-355.

Stasinopoulos, M. D., Rigby, R. A., Bastiani, F. D., 2018. GAMLSS: A distributional regression approach. *Statistical Modelling*, 18(3-4), pp. 248-273.

Stigler, S. M., 1975. The transition from point to distribution estimation. Bull. Int. Statist. Inst. 46, pp. 332-340

Staszewska-Bystrova, A., 2011. Bootstrap prediction bands for forecast paths from vector autoregressive models. *Journal of Forecasting*, Vol. 30, Issue 8, pp. 721–735.

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

Tkacz, G., Hu, S., 1999. Forecasting GDP Growth Using Artificial Neural Networks, Staff Working Papers 99-3, Bank of Canada.

Umlauf, N., Klein, N., Zeileis, A., 2018. BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, 27(3), pp. 612-627.

Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.,Kavukcuoglu, K., 2016. Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. 2017. Attention is all you need. Advances in neural information processing systems, 30.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., 2010. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion 38. *Journal of Machine Learning Research*, Volume 11, Number 110, pp. 3371-3408

Von Myrbach, O., 1913. Verlasslichkeit der Wetterprognosen (Ein Voschlag), Meteorologische Zeitschrift, 30, pp. 496-497

Wallis, K. F., 1993. Comparing Macroeconometric Models: A Review Article. *Economica*, 60(238), pp. 225–237.

Wan, C., Lin, J., Wang, J., Song, Y., Dong, Z. Y., 2016. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation. *IEEE Transactions on Power Systems*, 32(4), pp. 2767-2778.

Williams, W. H., Goodman, M., 1971. A simple method for the construction of empirical confidence limits for economic forecasts. *Journal of the American Statistical Association*, 66(336), pp. 752-754.

Yan, X., 2009. Linear Regression Analysis: Theory and Computing, World Scientific, pp. 1–2,

Yanchev, M., 2023. Uncertainty - Definition and Classification for the Task of Economic Forecasting, Bulgarian Economic Papers bep-2023-03, Faculty of Economics and Business Administration, Sofia University St Kliment Ohridski – Bulgaria, Center for Economic Theories and Policies at Sofia University St Kliment Ohridski, revised Mar 2023.

Yeh, A. B., 1998. A bootstrap procedure in linear regression with nonstationary errors. *Canadian Journal of Statistics*, 26(1), pp. 149-160.

Zakoian, J. M., 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5), pp. 931-955.