Journal of Statistics: Advances in Theory and Applications
Volume 5, Number 1, 2011, Pages 15-35

QUANTILE VERSIONS OF HOLT-WINTERS FORECASTING ALGORITHMS

A. ALEXANDRE TRINDADE and YANXUN XU

Department of Mathematics & Statistics
Texas Tech University
Lubbock, TX 79409-1042
USA
e-mail: alex.trindade@ttu.edu

Abstract

We propose new versions of Holt-Winters (HW) and seasonal Holt-Winters (SHW) time series forecasting algorithms. The exponential smoothing construct is identical to HW/SHW, except that the coefficients are estimated by minimizing a given quantile error criterion, instead of the usual squared errors. We call these versions quantile HW/SHW (QHW/QSHW), which amounts to performing HW/SHW under an asymmetric error loss function. We discuss best linear prediction (BLP) for ARIMA models, and highlight some models, where various versions of exponential smoothing are known to be optimal (in the BLP sense). This serves as a guide to models that we should focus on for a simulation study. The simulations compare scaled prediction errors between BLP and QHW, with models driven by Gaussian and Laplace noise. The results show that in most cases QHW gives similar forecasts to BLP. The advantage of QHW over BLP is that the user does not have to a-priori decided on a model for the data. The methodology is illustrated on some real data sets of interest in climatology and finance.

2010 Mathematics Subject Classification: Primary 60G10, 60G25; Secondary 62F25, 62G15.

Keywords and phrases: exponential smoothing, best linear prediction, quantile forecast, asymmetric loss.

Received January 13, 2011

1. Introduction

Exponential smoothing has been shown through the years to be very useful in many forecasting situations. It originated in Brown's work in the 1940s. During the early 1950s, Brown extended *simple exponential smoothing* (SES) to discrete data and developed methods for trends and seasonality (Brown [4]). At approximately the same time, Holt developed a similar method to exponential smoothing for use on non-seasonal time series showing no trend (Holt [12]). He later extended this to a procedure that does handle trends. Winters [28] generalized the method to include seasonality, and this became known as the Holt-Winters (HW) forecasting system. See Gardner [8, 10] for a more detailed discussion on the early history of the methodology.

The most important reason for the popularity of exponential smoothing is the surprising accuracy that can be obtained with minimal effort in model identification. There has been substantial work in connecting exponential smoothing and autoregressive integrated moving average (ARIMA) modelling. Muth [21] was among the first to prove that SES is "optimal" for an ARIMA (0,1,1) model, in the sense that it produces minimum mean squared error predictions. Two large empirical studies involving Box-Jenkins methodology by Makridakis and Hibon [17] and Makridakis et al. [18], noted little difference in forecast accuracy between exponential smoothing and ARIMA models. Non-seasonal HW has been shown to be optimal for two generating processes: A linear growth state-space model and an ARIMA (0,2,2) process (Harrison [11], Nerlove and Wage [22], Theil and Wage [26]).

The seasonal version of HW (SHW) is optimal for a certain seasonal ARIMA process derived by McKenzie [20]. As Chatfield [5] observes, the additive version of SHW would probably never be identified through Box-Jenkins procedures; and the multiplicative model does not appear to have an ARIMA equivalent. Thus, in general, HW models are not special cases of the ARIMA class. Makridakis et al. [18] showed that HW was robust at short horizons, but had a tendency to overshoot the data at longer horizons.

Since Gardner [8] the special case argument has been changed completely, and we now know that exponential smoothing methods are optimal for a very general class of state-space models that is in fact, broader than the ARIMA class (Gardner [10]). The basic elements of this new "taxonomy" of describing state-space equivalent models are introduced by Hyndman et al. [13]. All linear exponential smoothing methods have equivalent ARIMA representations, although most are so complex that, it is unlikely they would ever be identified through Box-Jenkins methodology. Gardner and McKenzie [9] found that the non-seasonal models contained at least six ARIMA models as special cases, with different parameters choices.

While much attention has been paid to *point* forecasts, few studies address the issue of *interval* forecasts, where a range of plausible predicted values with a given confidence level is desired. The issue is reviewed by Chatfield ([6], Chapter 7), who remarks that there is no consensus on what to call these: Forecast limits, prediction bounds, confidence intervals, forecast regions, or prediction intervals. One approach has been to estimate quantiles of the conditional probability distribution of future values, and as such is synonymous with the *value at risk* forecasting methodology so popular in the financial literature (e.g., Jorion [14]). Taylor and Bunn [23] apply quantile regression (Koenker [15]) to the empirical fit errors from a version of exponential smoothing using simple power functions of the lead time as regressors, in order to produce quantile models for the forecast error. Taylor [24, 25] proceeds similarly, but now uses exponentially weighted quantile regression for quantile forecasting.

In this paper, we propose a version of quantile forecasting by simply minimizing the empirical quantile criterion for the errors derived from HW and SHW. In other words, for a given quantile, estimates of the smoothing coefficients in HW and SHW are produced by minimizing empirical quantile errors. We call this *quantile HW/SHW* (QHW/QSHW). In the terminology of Makridakis and Hibon [19], who investigate the effect of different initial values and loss functions on exponential smoothing forecast accuracy, the approach may be viewed as utilizing an asymmetric loss function on forecast errors for optimal determination of the smoothing coefficients.

The remainder of our paper is organized as follows. Section 2 summarizes the standard best linear prediction theory for autoregressive integrated moving average (ARIMA) processes, and associated quantile forecast versions in the context of models driven by both Gaussian and Laplace noise. Section 3 highlights well-known connections between ARIMA and HW-optimal forecasts, and suggests some models for in-depth investigation in a simulation study. The proposed empirical quantile versions of the HW and SHW algorithms are presented in Section 4. Simulations in Section 5 compare the predictions from these algorithms with those derived from quantile versions of ARIMA models. We conclude in Section 6 by illustrating the proposed algorithms on three real data sets: Global annual average temperature, arctic sea ice extent, and the S&P 500 index.

2. Prediction for ARMA and ARIMA Models

We give here a brief overview of classical best linear prediction theory in the context of stationary autoregressive moving average (ARMA) models and non-stationary autoregressive integrated moving average (ARIMA) models. This standard material, which can be found in, e.g., Brockwell and Davis [3], is necessary here in order to discuss quantile prediction and thus make a connection with the proposed QHW/QSHW method.

Consider the time series $\{X_t\}, t = \ldots, -1, 0, 1, \ldots$ driven by a sequence of zero-mean independent and identically distributed (IID) random variables $\{Z_t\}$ with variance σ^2. Letting B denotes the backward shift operator, $BX_t = X_{t-1}$, the ARMA(p, q) model, $\phi(B)X_t = \theta(B)Z_t$, where

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \tag{1}$$

denote the AR and MA polynomials, respectively, can be expressed in the equivalent forms

$$X_t = \frac{\theta(B)}{\phi(B)} Z_t \equiv \psi(B) Z_t, \quad \text{and} \quad Z_t = \frac{\phi(B)}{\theta(B)} X_t \equiv \pi(B) X_t.$$

If the process is *causal*, then $\psi(z) = \psi_0 + \psi_1 z + \psi_2 z^2 + \cdots,$ for some absolutely summable sequence of constants $\{\psi_j\}$, so that the process can be expressed in a form that is independent of the future,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \tag{2}$$

If the process is *invertible*, then $Z_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$, for some sequence of absolutely summable coefficients $\{\pi_i\}$. (Equivalently, the series is causal, if and only if $\phi(z) \neq 0$ for $|z| \leq 1$, and invertible, if and only if $\theta(z) \neq 0$ for $|z| \leq 1$.)

The ARIMA(p, d, q) model satisfies the equations $\phi^*(B) X_t = \theta(B) Z_t$, where $\phi^*(z) = (1 - z)^d \phi(z)$. For $d \geq 1$, this is a non-stationary model, but the integrated process $(1 - B)^d X_t$ is a stationary ARMA (p, q) with AR and MA polynomials as in (1). As in the ARMA case, we have the equivalent infinite order MA representation,

$$X_t = \frac{\theta(B)}{\phi^*(B)} Z_t \equiv \psi^*(B) Z_t, \quad \psi^*(z) = \theta(z) / \phi^*(z).$$

Consider the classical best linear h-step ahead prediction problem, i.e., the linear function of observations $X_1, \ldots, X_n$ from a causal invertible ARMA driven by IID noise that minimizes the mean squared prediction error. To simplify the exposition, we will focus on the h-step ahead best linear predictor (BLP), $\widetilde{X}_n(h)$, based on the infinite past. Then, standard results from, e.g., Brockwell and Davis [3], gives

$$\widetilde{X}_n(h) = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j},$$

whence the prediction error is

$$\tilde{e}(h) \equiv X_{n+h} - \tilde{X}_n(h) = \sum_{j=0}^{h-1} \psi_j Z_{n+h-j},$$

with variance,

$$\tilde{\sigma}(h)^2 \equiv \mathbb{E}(X_{n+h} - \tilde{X}_n(h))^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2.$$

As is well known, the BLP on the infinite past coincides with the *best* predictor on the infinite past, $\mathbb{E}(X_{n+h} \mid X_n, X_{n-1}, \dots)$. In applications, and if n is large, the difference between the BLPs on the infinite vs. finite pasts is negligible.

For an ARIMA with underlying causal and invertible ARMA, the BLP expressions are analogous:

$$\tilde{X}_n(h) = \sum_{j=h}^{\infty} \psi_j^* Z_{n+h-j}, \quad \tilde{e}(h) = \sum_{j=0}^{h-1} \psi_j^* Z_{n+h-j}, \quad \tilde{\sigma}(h)^2 = \sigma^2 \sum_{j=0}^{h-1} (\psi_j^*)^2.$$

Let $\tilde{X}_{n+h} = \tilde{x}_n(h) + \tilde{e}(h)$ denote the predictive distribution of X_{n+h} conditional on a fixed predicted value of $\tilde{X}_n(h) = \tilde{x}_n(h)$. If $\tilde{e}_\nu(h)$ denotes the ν quantile of the prediction error $\tilde{e}(h)$, then we have

$$P(\tilde{x}_n(h) + \tilde{e}_{\nu/2}(h) < \tilde{X}_{n+h} < \tilde{x}_n(h) + \tilde{e}_{1-\nu/2}(h)) = 1 - \nu,$$

whence, the end terms of the inequality can be taken to be $(1-\nu)100\%$ prediction bounds. In particular, the ν quantile of the predictive distribution of X_{n+h} is given by

$$\begin{aligned} Q_\nu(\tilde{X}_{n+h}) &= Q_\nu(\tilde{x}_n(h) + \tilde{e}(h)) = \tilde{x}_n(h) + Q_\nu(\tilde{e}(h)) \\ &= \tilde{x}_n(h) + \tilde{e}_\nu(h) \equiv \tilde{x}_n(h, \nu), \end{aligned} \tag{3}$$

where $Q_\nu(\cdot)$ is generic notation for the ν quantile function.

In certain special cases, it is possible to obtain explicit expressions for (3). Consider the following:

Gaussian noise. $Z_t \sim \mathrm{N}(0, \sigma^2)$. Then $\tilde{e}(h) \sim \mathrm{N}(0, \tilde{\sigma}(h)^2)$, and thus, $\tilde{e}_\nu(h) = \Phi_\nu^{-1}\tilde{\sigma}(h)$, where Φ_ν^{-1} is the ν quantile of a standard normal.

Laplace noise. $Z_t \sim \mathrm{AL}(\theta, \kappa, \tau)$. Then $\tilde{e}(h)$ is a linear combination of asymmetric Laplace (AL) distributions, whence its quantiles, $\tilde{e}_\nu(h)$, can be explicitly computed according to the algorithm of Trindade et al. ([27]), Proposition 1). The AL is characterized by location (θ), scale (τ), and skewness $(\kappa > 0)$ parameters. Values of κ in the intervals $(0, 1)$ and $(1, \infty)$, correspond to positive (right) and negative (left) skewness, respectively (Kotz et al. [16]).

If the model contains a deterministic trend, μ_t, this term is subtracted from the model before obtaining the predictions, and then added back in. For example, in the Laplace case,

$$\tilde{x}_n(h, \nu) = \mu_{n+h} + \tilde{x}_n(h) + \tilde{e}_\nu(h).$$

3. Connections Between ARIMA and Holt-Winters Optimal Forecasts

This section gives an overview of the additive versions of exponential smoothing and its extensions to accommodate trends and seasonality. We will also highlight existing connections with optimal ARIMA and seasonal ARIMA models. This paves the way for deciding what models would make good candidates for extensive examination in order to assess forecast accuracy from the proposed QHW/QSHW method.

Consider the observed time series $X_1, \dots, X_n$. In describing the algorithms that constitute the various versions of exponential smoothing, we adopt the notation of Chatfield [6] and define the following terms applicable to the various algorithms: Simple exponential smoothing (SES), Holt-Winters (HW), and seasonal Holt-Winters (SHW).

Description of term	Term	Parameter	Applicable Algorithms
Level at time t	L_t	α	SES, HW, SHW
Slope at time t	T_t	β	HW, SHW
Seasonality of period s at time t	I_t	δ	SHW

Algorithm 1 (Simple exponential smoothing)**.** For smoothing parameter $0 < \alpha < 1$, the SES update is:

$$L_{t+1} = \alpha X_{t+1} + (1 - \alpha) L_t,$$

with initial conditions $L_1 = X_1$. The h-step forecast is, $\hat{X}_n(h) = L_n$.

It is well-known that the form of the predictors produced by SES coincides with the minimum mean-square error (MSE) predictors on the infinite past (henceforth, *optimal*) obtained from the ARIMA (0,1,1) model

$$(1 - B)X_t = Z_t - (1 - \alpha)Z_{t-1}, \quad Z_t \sim \mathrm{IID}(0, \sigma^2).$$

See, e.g., Chatfield ([6], Chapter 4).

Algorithm 2 (Holt-Winters)**.** For smoothing parameters $0 < \alpha < 1$ and $0 < \beta < 1$, the HW algorithm updates are:

$$L_{t+1} = \alpha X_{t+1} + (1 - \alpha)(L_t + T_t),$$

$$T_{t+1} = \beta(L_{t+1} - L_t) + (1 - \beta)T_t,$$

with initial conditions $L_2 = X_2$ and $T_2 = X_2 - X_1$. The h-step forecast is, $\hat{X}_n(h) = L_n + hT_n$.

Algorithm 3 (Seasonal Holt-Winters)**.** For smoothing parameters $0 < \alpha < 1$, $0 < \beta < 1$, and $0 < \delta < 1$, the SHW algorithm updates are:

$$L_{t+1} = \alpha(X_{t+1} - I_{t+1-s}) + (1 - \alpha)(L_t + T_t),$$

$$T_{t+1} = \beta(L_{t+1} - L_t) + (1 - \beta)T_t,$$

$$I_{t+1} = \delta(X_{t+1} - L_{t+1}) + (1 - \delta)I_{t+1-s},$$

with initial conditions $L_{s+1} = X_{s+1}$, $T_{s+1} = (X_{s+1} - X_1)/s$, and $I_i = X_i - [X_1 + (i-1)T_{s+1}]$, for $i = 1, \ldots, s+1$. The h-step forecast is, $\hat{X}_n(h) = L_n + hT_n + I_{n-s+h}$.

For SHW, the optimal model is a complicated seasonal ARIMA,

$$(1 - B)(1 - B^s)X_t = \theta_1 Z_{t-1} + \ldots + \theta_{s+1} Z_{t-s-1} + Z_t,$$

with specified dependence of the θ_j on the smoothing parameters (see Equation (6.15) in Abraham and Ledolter [1]).

Abraham and Ledolter [1] show that HW provides optimal forecasts when the underlying true model is the following:

Model 1. HW forecasts are optimal according to the ARIMA (0,2,2) model:

$$(1 - B)^2 X_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t, \quad Z_t \sim \mathrm{IID}(0, \sigma^2),$$

with $\theta_1 = \alpha(1 + \beta) - 2$ and $\theta_2 = (1 - \alpha)$.

We present two other models, where neither HW nor SHW are optimal, but where the strong trend and/or lack of seasonality would suggest use of HW. These will be used in the simulation study of Section 5 in order to compare the forecasts obtained from QHW/QSHW to minimum mean-square error (MSE) predictors based on the fitted model.

Model 2. An ARIMA (2,1,0) model, where neither HW nor SHW forecasts are optimal:

$$(1 - 0.5B)(1 + 0.9B)(1 - B)X_t = Z_t, \quad Z_t \sim \mathrm{IID}(0, \sigma^2).$$

Model 3. An AR (1) model with deterministic linear trend, where neither HW nor SHW are optimal:

$$X_t = 0.3t + W_t,$$

$$W_t = 0.7W_{t-1} + Z_t, \quad Z_t \sim \mathrm{IID}(0, \sigma^2).$$

Figure 1 displays some simulated realizations from these three models.

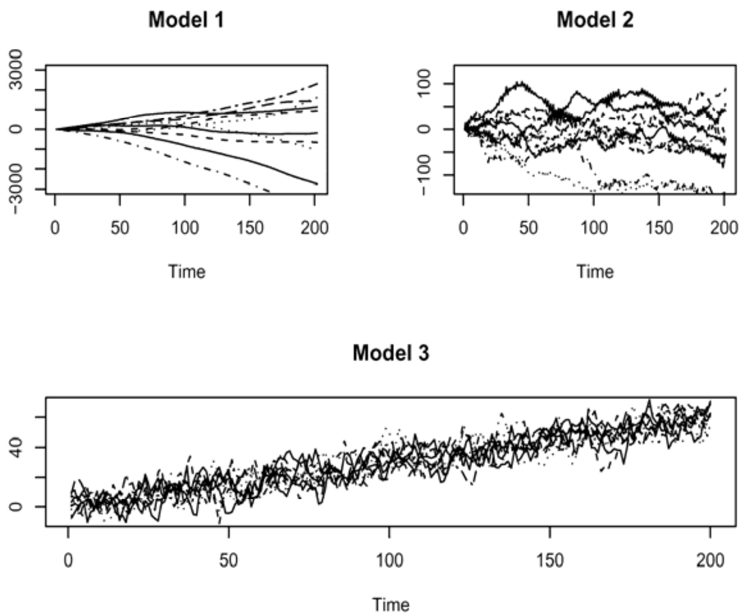

Figure 1. Simulated realizations of length $n = 200$ from Models 1, 2, and 3, driven by Gaussian noise with variance $\sigma^2 = 25$, i.e., $Z_t \sim$ IID N(0, 25).

4. Quantile Versions of Holt-Winters Algorithms

In both HW and SHW, the smoothing parameters $\{\alpha, \beta, \delta\}$ can either be chosen arbitrarily, or estimated by minimizing some (*loss*) function of successive one-step prediction errors, $\hat{e}_{t-1}(1) = X_t - \hat{X}_{t-1}(1)$,

$$\sum_{t=s+2}^{n} \ell(\hat{e}_{t-1}(1)), \tag{4}$$

where s is the period in SHW, and $s = 1$ in HW. Typically squared loss, $\ell(z) = z^2$, is used, resulting in the mean-squared error (MSE) criterion. Another common choice is $\ell(z) = |z|$, mean absolute deviation (MAD).

We propose versions of HW/SHW that estimates the smoothing coefficients by minimizing the ν quantile error criterion. That is, use the loss function

$$\ell_{\nu}(z) = z(\nu - I(z < 0)), \tag{5}$$

in (4), where $I(z < 0)$ takes on the value 1 if $z < 0$, and 0 otherwise. We call these *Quantile HW/SHW* (QHW/QSHW).

The rationale for this choice of loss function is that, for a random variable Z with quantile function $Q(\cdot)$, minimization of $\mathbb{E}\ell_{\nu}(Z - \xi)$ over all $\xi \in \mathbb{R}$ results in the optimal value, $\xi_{\nu} = Q(\nu)$, the ν quantile of Z. The empirical analogue for a sample of data $z_1, \ldots, z_n$, is minimization of $n^{-1}\sum_{i=1}^{n} \ell_{\nu}(z_i - \hat{z})$, over all $\hat{z} \in \mathbb{R}$, resulting in $\hat{z}$ being the ν sample quantile (e.g., Koenker [15]). Heuristically, this criterion then "trains" the one-step predictions $\hat{X}_{t-1}(1)$ to closely track the ν quantile of the predictive distribution of X_{t+1}.

In general, for h-step prediction based on observations $x_1, \ldots, x_n$, let us denote the (empirically) obtained QHW/QSHW optimal value by $\hat{x}_n(h, \nu)$. We then propose empirical quantile versions of Algorithms 1-3, where the smoothing parameters are obtained by minimizing the loss function $\ell_{\nu}(\cdot)$. For example, QHW is as follows:

Algorithm 4 (Empirical quantile Holt-Winters)**.** For observations $x_1, \ldots, x_n$, and chosen quantile $0 < \nu < 1$, let the smoothing parameters $0 < \hat{\alpha} < 1$ and $0 < \hat{\beta} < 1$ be such that (4) is minimized with loss function (5). The QHW algorithm updates are then given by:

$$L_{t+1} = \hat{\alpha} x_{t+1} + (1 - \hat{\alpha})(L_t + T_t),$$

$$T_{t+1} = \hat{\beta}(L_{t+1} - L_t) + (1 - \hat{\beta})T_t,$$

with initial conditions $L_2 = x_2$ and $T_2 = x_2 - x_1$. The h-step forecast is, $\hat{x}_n(h, \nu) = L_n + hT_n$.

To the best of our knowledge, this particular variant of exponential smoothing has not been investigated. Makridakis and Hibon [19] explored different loss functions on exponential smoothing forecast accuracy, most of them are symmetric, although they did consider an asymmetric loss function, which weighted positive errors less than negative ones. This loss is, however, quite different from the one we propose, and was not motivated from the point of view of quantile forecasting. On the other hand, Taylor and Bunn [23] and Taylor [24, 25], do start from the perspective of quantile forecasting in the context of exponential smoothing, but apply variants of quantile regression directly to the series X_t. None of these approaches subsumes ours as a special case.

5. Simulations

If quantile forecasting is desired in a given practical setting, one might resort to the quantile BLP values, $\hat{x}_n(h, \nu)$ defined in (3), in the context of the classical ARIMA model. However, this forces one to think about appropriate models for the data and as such is hindered by the "model selection" problem. The proposed QHW algorithms offer a more "automatic" alternative through the corresponding $\hat{x}_n(h, \nu)$ values. In light of the discussion in Sections 3 and 4, it therefore makes sense to compare the following values:

$$\hat{x}_n(h, \nu) \approx \tilde{x}_n(h, \nu). \tag{6}$$

In this section, we carry out some simulation studies to assess the closeness of the QHW values to the quantile BLP values, suggested by (6). To this end, we simulate realizations from Models 1-3. For a given model, we repeat the following loop, $m = 100$ times:

- Simulate $X_1, \ldots, X_n$. (If the model is an ARIMA (p, d, q), then we start the simulation by taking the $(p + d)$ pre-sample values equal to zero.)

- Simulate the next value of the series: $X_{n+1} = x_{n+1}$.

• Compute the QHW 1-step prediction: $\hat{x}_n(1, \nu)$, for $\nu = 0.05, 0.1, 0.15, \ldots, 0.95$.

• Fit the model to the series via maximum likelihood. Compute the BLP 1-step prediction: $\tilde{x}_n(1, \nu)$, for $\nu = 0.05, 0.1, 0.15, \ldots, 0.95$. (This prediction uses the estimated model parameters.)

In addition, each series is of length $n = 200$, with the following distributions for the serially independent white noise process Z_t: Normal, symmetric Laplace $(\kappa = 1)$, and asymmetric Laplace $(\kappa = 0.5)$. The location and scale parameters are chosen such that $\mathbb{E}Z_t = 0$ and $\mathrm{Var}(Z_t) = \sigma^2 = 25$ in all cases.

A visual summary of these simulation results can be seen in Figure 2, which displays boxplots of the scaled quantile forecast differences (SQD) between QHW and BLP, as a function of the quantile (ν),

$$\mathrm{SQD}(\nu) \equiv \frac{\hat{x}_n(1, \nu) - \tilde{x}_n(1, \nu)}{\sqrt{2\sigma^2}}.$$

An appropriate specification of the scale factor appearing in the denominator of SQD is a difficult problem. A heuristic argument for our choice is as follows. Ignoring the covariance between QHW and BLP, and assuming the variance of QHW is approximately the same as that of the BLP, the expressions from Section 2 give

$$\begin{aligned} \mathrm{Var}[\hat{x}_n(1, \nu) - \tilde{x}_n(1, \nu)] &= \mathrm{Var}[\hat{e}_\nu(1) - \tilde{e}_\nu(1)] \\ &\approx \mathrm{Var}[\hat{e}(1) + \tilde{e}(1)] \\ &\approx 2\mathrm{Var}[\tilde{e}(1)] \\ &= 2\tilde{\sigma}(1)^2 = 2\sigma^2\psi_0^* = 2\sigma^2, \end{aligned} \tag{7}$$

since $\psi_0 = \psi_0^* = 1$ for all models.

The boxplots in Figure 2, each based on 100 values, show that there is generally close agreement between QHW and BLP quantile forecasts. In fact, most SQD values lie between ± 1, with few outside ± 3. There seems to be changing variability across some plots, but this just alludes to the fact that the standard error expression derived in (7) may actually depend upon ν. This is particularly evident in Model 1, where more variability is seen for low and high values of ν. Given this, it may be more important to focus on whether there is a systematic bias in the QHW and BLP differences at a given quantile, as indicated by median SQD values away from zero. This bias is practically zero for the normal-based plots, but is not so negligible for the Laplace-based ones (although, it does average out to zero across all quantiles in a given plot). The only bad case is Model 2-normal, which exhibits SQD values between ± 20 (but little bias).

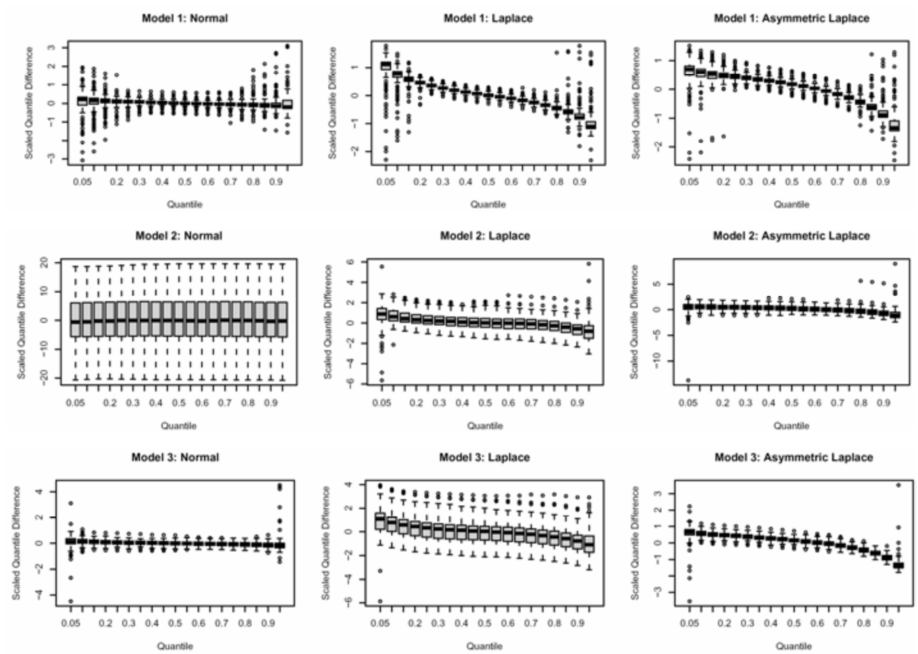

Figure 2. Scaled differences between QHW and BLP quantile one-step forecasts for simulated realizations of length $n = 200$ from Models 1-3, driven by independent noise with mean zero and variance 25. The noise distribution is normal, Laplace, and asymmetric Laplace ($\kappa = 0.5$), on the left, middle, and right panels, respectively.

6. Applications

In this section, we showcase some applications of the proposed QHW/QSHW algorithms.

Figure 3 displays the famous global annual average temperature series from 1856 to 2005 (degrees celsius), expressed as anomalies from the average over the period 1961-1990. (Source: Climatic Research Unit, University of East Anglia.) The absence of seasonality suggests HW and QHW as forecasting algorithms. The points extending from 2006-2025 are predictions by using HW and QHW for quantiles ranging from 0.1-0.9. For quantiles $\nu > 0.5$, the difference between QHW and HW is negligible.

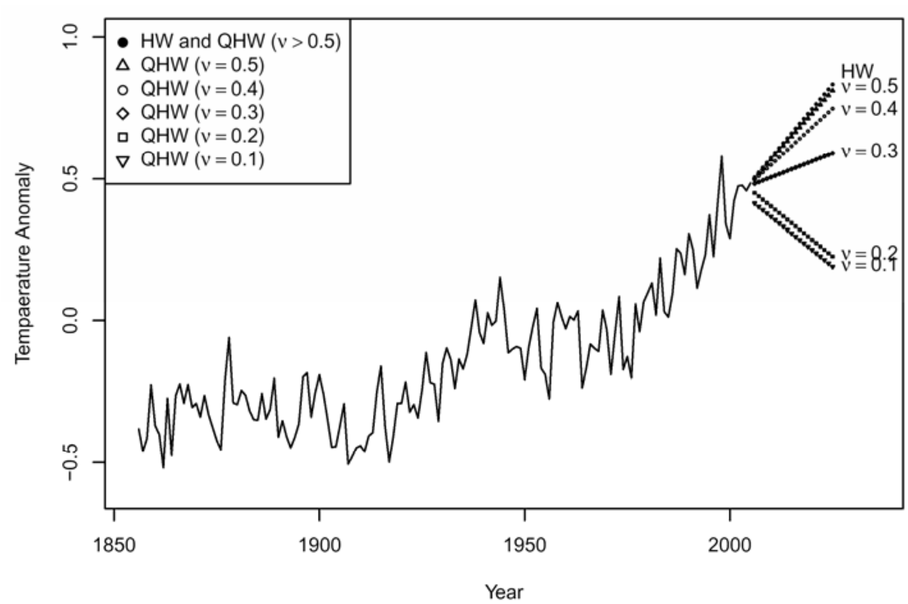

Figure 3. Time series of global annual average temperatures from 1856-2005 (degrees celsius), expressed as anomalies from the average over the period 1961-1990. Points extending from 2006-2025 are predictions by using HW and QHW for quantiles ranging from 0.1-0.9.

The second time series is monthly Arctic Sea ice extent from January 1979 to December 2005 (million square kilometers). (Source: The National Snow and Ice Data Center, University of Colorado. Two missing values in the original data were filled in by simple interpolation.) The presence of seasonality suggests SHW and QSHW as forecasting algorithms with period 12. The data from January 1979 to December 2003 (first 300 months) were used to predict sea ice extent for the period January 2004 to December 2005 (next 24 months). For illustrative purposes, we show only the QSHW forecasts obtained using $\nu = 0.1$ and $\nu = 0.9$, thus corresponding to a plausible envelope of lower and upper 10% prediction limits. Figure 4 shows the actual observed data subtracted from the forecasts: SHW (solid), QSHW $\nu = 0.1$ (dash), and QSHW $\nu = 0.9$ (dash-dot). The reason for showing the forecasts as departures from observed data (forecast-observed), is that the presence of a very regular seasonal cycle would otherwise make it difficult to visually distinguish the lines.

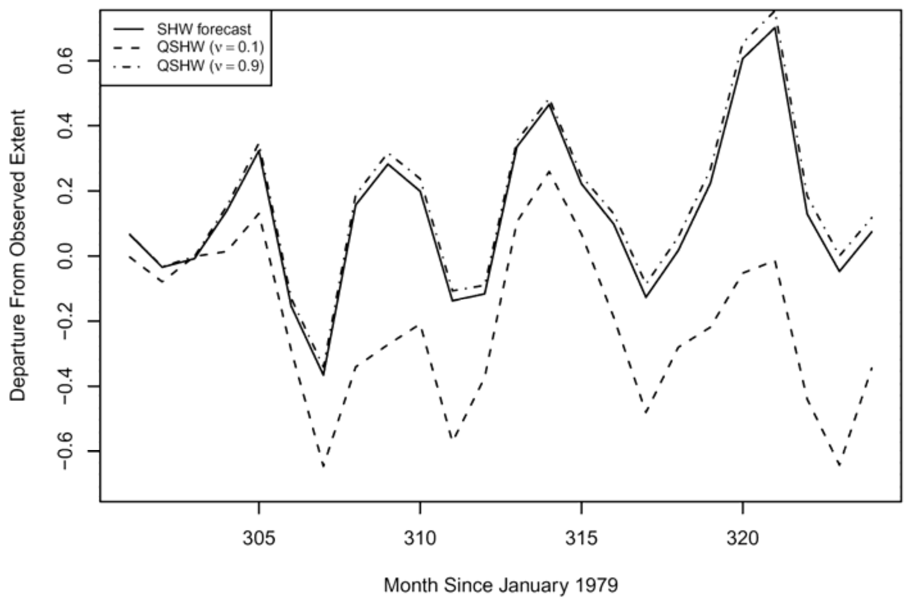

Figure 4. Predictions for monthly Arctic Sea ice extent (million km^2). The observed data spans the period from January 1979 to December 2005 (324 months). The first 300 months were used to predict sea ice extent for the next 24 months, via SHW (solid), QSHW with quantile 0.1 (dash), and QSHW with quantile 0.9 (dash-dot). The forecasts shown are departures from the observed data values.

Our last series consists of daily closing prices for the S&P 500 index from January 1 to October 15, 2009; a total of $n = 199$ observations. We compare one-step-ahead moving window forecasts based on two classes of model: Random walk, and QHW with $\nu = 0.1, 0.2, \ldots, 0.9$. Letting X_t be the value of the series on day t, we used the first 10 values to predict the next value at $t = 11$. Then, the first 11 values were used to predict at $t = 12$, etc. For the random walk, the predicted price index for the current day was the previous day's price.

Table 1 contains various summary measures of forecast accuracy for the S&P 500 data. The *underage* counts the proportion observations that fell below their corresponding predicted values. If $\text{QHW}(\tau)$ is correctly predicting the τ- quantile, we expect underages to be approximately τ. In this particular application, all underages fall in a narrow band of roughly 0.37 to 0.47, with the closest agreement between empirical and theoretical occurring for $\text{QHW}(\tau = 0.4)$, which has an underage of 0.3968. This does not necessarily discredit the validity of QHW, but merely suggests that, it may not be an appropriate model for capturing quantiles of the predictive distribution that deviate substantially from 0.4. We also notice some evidence of non-monotonicity in the underages, for example, the underage of 0.4603 for $\text{QHW}(\tau = 0.8)$ is actually larger than 0.4550, which corresponds to $\text{QHW}(\tau = 0.9)$. The possibility for such "quantile-crossings" was already noted by Koenker [15] in the context of quantile regression, and is usually an indication that the fitted model is not sufficiently rich to capture all the nuances of the data.

Table 1. Empirical summary measures of forecast accuracy for the daily closing prices of the S&P 500 data. The underage tallies the proportion of observations that fell below their corresponding predicted values

Forecast Method	Underage	MAPE	RMSPE
Random walk	0.4497	11.75	15.60
QHW (τ = 0.1)	0.3651	12.38	16.13
QHW (τ = 0.2)	0.3915	12.13	15.92
QHW (τ = 0.3)	0.3968	12.01	15.86
QHW (τ = 0.4)	0.3968	11.99	15.85
QHW (τ = 0.5)	0.4709	11.77	15.69
QHW (τ = 0.6)	0.4709	11.81	15.77
QHW (τ = 0.7)	0.4762	11.85	15.83
QHW (τ = 0.8)	0.4603	11.87	15.87
QHW (τ = 0.9)	0.4550	11.90	15.91

The other summary measures reported on Table 1 are the *mean absolute prediction error* (MAPE) and *root mean square prediction error* (RMSPE). If $\hat{x}_k(1)$ denotes a generic one-step-ahead prediction based on observations $x_1, \ldots, x_k$, these measures are defined as

$$\text{MAPE} = \frac{1}{n-10}\sum_{k=11}^{n}|x_k - \hat{x}_{k-1}(1)|, \quad \text{and} \quad \text{RMSPE} = \sqrt{\frac{1}{n-10}\sum_{k=11}^{n}[x_k - \hat{x}_{k-1}(1)]^2}.$$

Interestingly, both measures are minimized here at $\tau = 0.5$ among all the QHW forecasts. However, these values are lower still for the random walk predictions, suggesting that this may be a more appropriate model for these data. These two forecasts are illustrated in Figure 5, which displays departures from observed values (forecast-observed). It is clear that QHW ($\tau = 0.5$) closely tracks the random walk forecasts. Also discernible is a period of higher volatility in the first 75 or so forecasts.

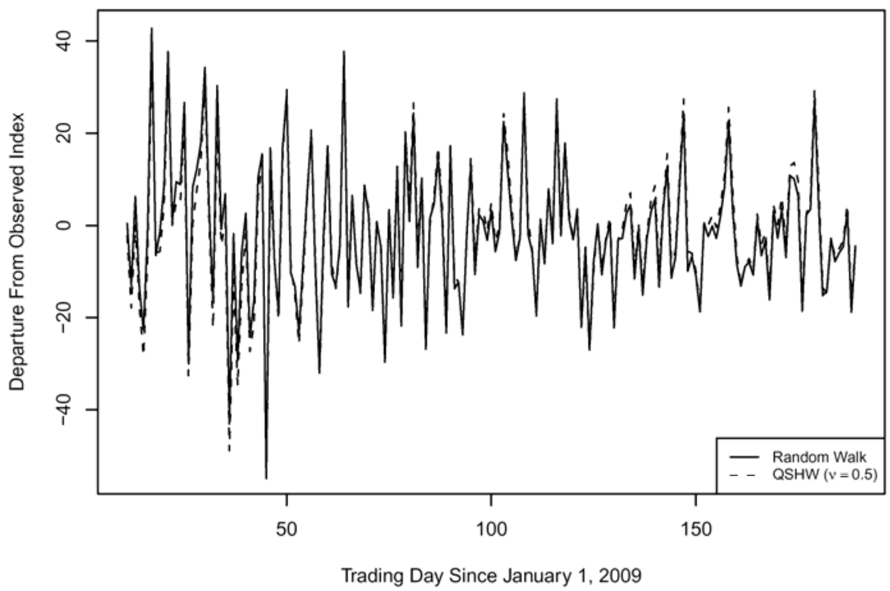

Figure 5. One-step-ahead moving window predictions for the S&P 500 index starting on the 11th trading day after January 1, 2009. The solid line indicates forecasts derived from a random walk model. QHW forecasts with quantile 0.5 are represented by a dashed line. The forecasts shown are departures from the observed index values.

References

[1] B. Abraham and J. Ledolter, Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures, International Statistical Review 54 (1986), 51-66.

[2] B. Abraham and J. Ledolter, On a theory of computation over the real numbers; NP completeness, recursive functions and universal machines, A. M. S. Bulletin 21 (1989), 1-46.

[3] P. J. Brockwell and R. A. Davis, Time Series: Theory and Methods, 2nd ed., Springer, New York, 1991.

[4] R. G. Brown, Smoothing, Forecasting and Prediction, Prentice-Hall, Englewood Cliffs, 1963.

[5] C. Chatfield, Some recent developments in time series analysis, Journal of Royal Statistical Society 140 (1977), 492-510.

[6] C. Chatfield, Time-Series Forecasting, Chapman & Hall/CRC, Boca Raton, 2001.

[7] E. S. Gardner, A simple method of computing prediction intervals for time series forecasts, Management Science 34 (1977), 541-546.

[8] E. S. Gardner, Exponential smoothing: The state of the art (with discussion), Journal of Forecasting 4 (1985), 1-38.

[9] E. S. Gardner and E. McKenzie, Model identification in exponential smoothing, Journal of the Operational Research Society 39 (1988), 863-867.

[10] E. S. Gardner, Exponential smoothing: The state of the art - Part II, International Journal of Forecasting 22 (2006), 637-666.

[11] P. J. Harrison, Exponential smoothing and short-term sales forecasting, Management Science 13 (1967), 821-842.

[12] C. C. Holt, Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages, ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957.

[13] R. J. Hyndman, A. B. Koehler, R. D. Snyder and S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting 18 (2002), 439-454.

[14] P. Jorion, Value at Risk: The New Benchmark for Managing Financial Risk, 3rd ed., McGraw-Hill, New York, 2006.

[15] R. Koenker, Quantile Regression, Cambridge University Press, New York, 2005.

[16] S. Kotz, T. J. Kozubowski and K. Podgórski, The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance, Birkhäuser, Boston, 2001.

[17] S. Makridakis and M. Hibon, Accuracy of forecasting: An empirical investigation, (with discussion), Journal of the Royal Statistical Society (A) 142 (1979), 97-145.

[18] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler, The accuracy of extrapolation (time series) methods: Results of a forecasting competition, Journal of Forecasting 1 (1982), 111-153.

[19] S. Makridakis and M. Hibon, Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy, International Journal of Forecasting 7 (1991), 317-330.

[20] E. McKenzie, A comparison of some standard seasonal forecasting systems, The Statistician 25 (1976), 3-14.

[21] J. F. Muth, Optimal properties of exponentially weighted forecasts, Journal of American Statistical Association 55 (1960), 299-306.

[22] M. Nerlove and S. Wage, Some observations on adaptive forecasting, Management Science 10 (1964), 107-224.

[23] J. W. Taylor and D. W. Bunn, A quantile regression approach to generating prediction intervals, Management Science 45 (1999), 225-237.

[24] J. W. Taylor, Forecasting daily supermarket sales using exponentially weighted quantile regression, European Journal of Operational Research 178 (2007), 154-167.

[25] J. W. Taylor, Using exponentially weighted quantile regression to estimate value at risk and expected shortfall, Journal of Financial Econometrics 6 (2008), 382-406.

[26] H. Theil and S. Wage, Some observations on adaptive forecasting, Management Science 10 (1964), 198-206.

[27] A. A. Trindade, Y. Zhu and B. Andrews, Time series models with asymmetric Laplace innovations, Journal of Statistical Computation and Simulation 80 (2010), 1317-1333.

[28] P. R. Winters, Forecasting sales by exponentially weighted moving averages, Management Science 6 (1960), 324-342.

■