A Quantile Regression Approach to Generating Prediction Intervals

James W. Taylor • Derek W. Bunn
London Business School, Sussex Place, Regents Park, London NW1 4SA, United Kingdom
jtaylor@lbs.ac.uk
dbunn@lbs.ac.uk

Exponential smoothing methods do not involve a formal procedure for identifying the underlying data generating process. The issue is then whether prediction intervals should be estimated by a theoretical approach, with the assumption that the method is optimal in some sense, or by an empirical procedure. In this paper we present an alternative hybrid approach which applies quantile regression to the empirical fit errors to produce forecast error quantile models. These models are functions of the lead time, as suggested by the theoretical variance expressions. In addition to avoiding the optimality assumption, the method is nonparametric, so there is no need for the common normality assumption. Application of the new approach to simple, Holt's, and damped Holt's exponential smoothing, using simulated and real data sets, gave encouraging results.
(*Exponential Smoothing*; *Predictive Distribution*; *Empirical Approach*; *Quantile Regression*)

1. Introduction

It is often important to associate an assessment of uncertainty with a point forecast to provide some idea of the likely variation. Despite the fact that the predictive distribution is in many cases difficult to generate, forecasters are frequently encouraged to produce a close approximation. Reasons for this include the increasing use of risk simulation and the general embedding of forecasting methods in decision support systems that take account of uncertainty.

Exponential smoothing is a widely used, robust forecasting approach. It is one of a set of forecasting methods that do not involve a formal procedure for identifying the underlying data generating model. The question then arises as to whether prediction intervals should be calculated by some empirical approach or by assuming that the method is optimal in some sense. A normality assumption is usually made. In reviewing the prediction interval literature, Chatfield (1993) observes that, when theoretical formulae are not available or there are doubts about model assumptions, the use of empirically based methods should be considered as a general purpose alternative. However, he suggests that more research on this sort of approach is needed.

In this paper we present a quantile regression approach to the construction of predictive distributions and prediction intervals. The method is a hybrid of the empirical and theoretical approaches in that it uses the empirical fit errors and produces forecast error quantile models which are functions of the lead time, k, as suggested by the theoretically derived variance expressions. The optimality assumption of the theoretical approach and the normality assumption are avoided by the application of quantile regression to the empirical fit errors.

Section 2 is a discussion of the existing approaches to generating prediction intervals for exponential smoothing. Section 3 describes the theory of quantile regression before presenting our proposed approach in §4. Section 5 describes a simulation study that compared the performance of the new approach to the

0025-1909/99/4502/0225$05.00

main established methods for three exponential smoothing methods. In §6, we report the results of a study which used real data. Section 7 offers some concluding comments.

2. Prediction Intervals for Exponential Smoothing

2.1. Theoretical Approaches

Simple exponential smoothing is well known to be optimal for an ARIMA(0, 1, 1) model or for the structural no-trend model. Holt's two-parameter (nonseasonal) smoothing is optimal for an ARIMA(0, 2, 2) model or for a particular linear growth structural model. The three-parameter additive Holt-Winters method is optimal for a seasonal ARIMA model. There is no equivalent ARIMA or structural model to the multiplicative Holt-Winters, because the forecasts are not a linear combination of past observations. The method put forward in the literature for calculating theoretical prediction intervals for exponential smoothing is to assume that the optimal ARIMA or state-space model is the true underlying model. This is straightforward for one- or two-parameter exponential smoothing. It is also possible to calculate prediction intervals for the additive Holt-Winters method, with or without damping, by assuming that the equivalent optimal ARIMA model is true or alternatively (and equivalently) by assuming that the 1-step-ahead forecast errors are uncorrelated and homoscedastic (Yar and Chatfield 1990). Chatfield and Yar (1991) address the multiplicative case and offer forecast error variance expressions corresponding to alternative assumptions for the 1-step-ahead forecast error variance.

For illustration, consider the simple exponential smoothing k-step-ahead predictor

$$\hat{X}_t(k) = S_t \quad \text{where } S_t = \alpha X_t + (1 - \alpha)S_{t-1}. \tag{1}$$

X_t is the actual observation, S_t is the smoothed level, and α is the smoothing parameter that is usually constrained so that $0 < \alpha < 1$. The method is optimal for the ARIMA(0, 1, 1) process:

$$X_t = X_{t-1} + e_t - \psi e_{t-1}.$$

The condition for optimality is $\psi = 1 - \alpha$. If we let $e_t(k) = X_{t+k} - \hat{X}_t(k)$ denote the k-step-ahead forecast error and σ_e^2 denote the variance of the 1-step-ahead forecast errors, then both the ARIMA(0, 1, 1) model and structural no-trend model lead to

$$\text{var}(e_t(k)) = \sigma_e^2[1 + (k - 1)\alpha^2]. \tag{2}$$

Assuming normality, prediction intervals can then be constructed based on this expression.

The derivation of (2) assumes complete knowledge of the model including values of α and σ_e. Yar and Chatfield (1990) state that the variance of the 1-step-ahead forecast errors can be estimated from the data in the fit period. Chatfield (1993) argues that, even though a model has not been formally identified, it is not unreasonable to use (2), provided that the observed 1-step-ahead forecast errors show no obvious autocorrelation and there are no other obvious features of the data (e.g., trends) that need to be modelled. This theoretical approach to prediction interval estimation is used by several of the widely used forecasting packages, including Forecast Pro for Windows (Business Forecast Systems 1994). However, a problem with the theoretical method is that the variance expression may be far from correct if the generating process is not the particular ARIMA model for which the exponential smoothing method is optimal. The normality assumption is another potential weakness.

2.2. Empirical Approaches

Gardner's (1988) empirical method can be applied to time series forecasting models. It proceeds by fitting the model's parameters, which for exponential smoothing is traditionally on the basis of 1-step-ahead fit. One pass through the data is then made to compute the standard deviation of the fitted errors at 1-step-ahead. Then a second pass is made to compute the standard deviation at two-steps-ahead and so on. The model is not reestimated each time, and the empirically computed standard deviations at different lead times are based on within-sample fitted errors rather than post-sample forecast errors. The advantage of this method of computing variances is that the validity of the model and the form of the generating function are irrelevant.

A multiplier based on the Chebyshev inequality is then applied to each standard deviation to deliver the desired prediction interval. Gardner uses these limits in preference to a normality assumption because his method is based on fit errors and, as is commonly reported (e.g., Makridakis and Winkler 1989), true post-sample forecast errors tend to be larger than model-fitting errors. Bowerman and Koehler (1989) point out that the use of the Chebyshev inequality may give very wide prediction intervals in some cases and, therefore, may be of little practical use. Chatfield (1993) comments that they may be unnecessarily wide for reasonably stable series for which the usual normal values would be adequate. His preference is for a normality assumption, which is the approach adopted by Yar and Chatfield (1990). They observe that Gardner's procedure requires considerable computation and is subject to sampling variability particularly for short series.

The technique of Williams and Goodman (1971) is similar in spirit to Gardner's approach, but relies on the empirical distributions of post-sample forecast errors. A model is fitted to part of the time series, and forecasts are made from the end of this part-series. Post-sample errors at different lead times may then be found. The model is then refitted with one additional observation and the procedure is repeated until the end of the series. Williams and Goodman constructed prediction intervals using the percentiles of the empirical distribution, thereby avoiding any distributional assumptions. They found that the distribution of forecast errors approximated to a gamma distribution. Yar and Chatfield (1990) point out that it is more computationally intensive than Gardner's method and that it is hardly used in practice.

3. Quantile Regression

Consider the following rather general model of systematic heteroscedasticity,

$$y_t = \mu_t(\boldsymbol{x}_t) + \sigma_t(\boldsymbol{x}_t)e_t,$$

where $\mu_t(\boldsymbol{x}_t)$ may be thought of as the conditional mean of the regression process, $\sigma_t(\boldsymbol{x}_t)$ as the conditional scale, and e_t as an error term independent of vector $\boldsymbol{x}_t$. The θth quantile of e_t is the value, $Q_e(\theta)$, for which $P(e_t < Q_e(\theta)) = \theta$. Note that having μ_t and σ_t depend on the same vector $\boldsymbol{x}_t$ is solely for notational convenience. The conditional quantile functions of y_t are then simply,

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \mu_t(\boldsymbol{x}_t) + \sigma_t(\boldsymbol{x}_t)Q_e(\theta).$$

Consider the case where μ_t and σ_t are linear functions of $\boldsymbol{x}_t$ which has 1 as first element,

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \boldsymbol{x}_t\boldsymbol{\beta} + (1 + \boldsymbol{x}_t\gamma)Q_e(\theta), \tag{3}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of coefficients. (Setting all the elements of $\boldsymbol{\gamma}$ to zero is equivalent to assuming that the error term of y_t is i.i.d.). Equation (3) can be rewritten as

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \boldsymbol{x}_t\beta(\theta), \tag{4}$$

where $\boldsymbol{\beta}(\theta)$ is a vector of coefficients dependent on θ. The *θth regression quantile* $(0 < \theta < 1)$ is defined as any solution, $\boldsymbol{\beta}(\theta)$, to the *quantile regression* minimisation problem

$$\min_{\boldsymbol{\beta}}\Big[\sum_{t \mid y_t \geq \boldsymbol{x}_t\boldsymbol{\beta}} \theta \mid y_t - \boldsymbol{x}_t\beta \mid + \sum_{t \mid y_t < \boldsymbol{x}_t\boldsymbol{\beta}} (1-\theta) \mid y_t - \boldsymbol{x}_t\boldsymbol{\beta} \mid \Big]. \tag{5}$$

Koenker and Bassett (1982) showed that if y_t and x_t are specified as dependent and independent variables respectively, then quantile regression delivers coefficients that asymptotically approach the coefficients, $\boldsymbol{\beta}(\theta)$, in (4) as the number of observations increases.

The common procedure for building an explanatory model for a variable is to look for a relationship between past observations of that variable and past observations of potential explanatory variables. This is not a feasible procedure for building a model for the quantiles of a variable because, although past observations of that variable may be available, past observations of the quantiles will not. The appeal of quantile regression is that past observations of the quantiles of the dependent variable are not required. Instead the dependent variable itself is regressed on the explanatory variables to produce a model for the quantile.

4. Estimating the Predictive Distribution Using Quantile Regression

Expression (2) gives the variance of the k-step-ahead forecast errors for an ARIMA(0, 1, 1) model, with the MA parameter equal to $1 - \alpha$ (the forecast is unbiased since it is assumed that α is known). The forecast error variance is a function of k, with α and σ_e constant parameters. It seems intuitively reasonable that the forecast error variance is not constant but is a function of k. A quick glance at the variance expressions for other ARIMA models confirms this. For example, the expression for an AR(1) model with parameter ϕ is

$$\operatorname{var}(e_t(k)) = \sigma_e^2(1 - \phi^{2k})/(1 - \phi^2). \tag{6}$$

If we view the variation in a series of forecast errors from various steps-ahead as systematic heteroscedasticity, we can use quantile regression to regress the forecast error on chosen explanatory variables to give a model for the selected quantile of the forecast error. In view of the variance expressions in (2) and (6), obvious candidates for the explanatory variables are simple functions of k. A reasonable approximation to the correct theoretical expression for the quantile should be obtained if the explanatory variables are well chosen. Consequently, we present this new approach as a hybrid of empirical and theoretical methods. This method is not dependent on the forecasting method used as it focuses on the series of forecast errors. Before further considering the new approach, we present some background to the proposal.

4.1. Prediction Error Regression Quantiles: Background

Koenker and Bassett's (1978,1982) theory of quantile regression is concerned with the estimation of quantiles for an observation y_t. There are many other approaches in the literature to addressing this problem. A popular procedure is simply to use a forecast and the distribution of the historical forecast errors. This approach assumes independent and identically distributed errors, which, as Gélinas and Lefrançois (1993) note, is more likely to be true for the errors than for the time series observations themselves.

The approaches of Gorr and Hsu (1985), Bookbinder and Lordahl (1989), and Gélinas and Lefrançois (1993) use the time series itself, as opposed to the fitting errors. Gorr and Hsu (1985) present an adaptive filter that estimates regression quantiles with time-varying parameters through a damped negative feedback mechanism. The bootstrapping procedure of Bookbinder and Lordahl (1989) estimates a quantile as the weighted sum of two successive order statistics. Gélinas and Lefrançois (1993) apply exponential smoothing to order statistics derived from recursive estimation using a moving window of observations.

This paper is concerned with estimating the forecast error distribution associated with a point forecast. As already mentioned, many of the approaches to the estimation of the distribution of a future observation y_t use an approximation of the forecast error distribution. This does not seem unreasonable. However, it would certainly be incorrect to reverse this procedure and estimate the forecast error distribution by an approximation of the distribution of y_t. The latter is an assessment of the true potential variation in the random variable, y_t. This is different to the estimation uncertainty in a point forecast of y_t. Our proposal is to use quantile regression to estimate the forecast error distribution. However, we do not estimate the predictive distribution by an approximation of the distribution of y_t, as we do not apply quantile regression to the time series of y_t, but instead to the series of past forecast errors.

4.2. Implementation of the Quantile Regression Approach

The basic idea is to use errors from Gardner's (1988) method, described in §2.2, to estimate an expression for forecast error quantiles as a function of k, the lead time. The method involves collecting together, as a single series, all of the errors derived by Gardner's procedure. If we are interested in constructing predictive distributions for 1, 3, 6, 9, 12, 15, and 18-step-ahead forecasts, then this single series would have the 1-step-ahead errors first, followed by the 3-step-ahead errors, the 6-step-ahead errors, etc. This is the *forecast error series*. We then define the elements of the *k series* as taking a value of k when the corresponding element of the forecast error series is a k-step-ahead error. If constructing predictive distributions for 1, 3, 6, 9, 12, 15, and 18-step-ahead forecasts, the k series will consist of 1s, followed by 3s, then 6s, etc. Expressing the series as vectors, we then carry out quantile regressions of the forecast error vector on simple functions of the k vector, such as k, k^2, $k^{1/2}$, $k^{-1/2}$, to deliver quantile models for the forecast error.

For example, the 95th forecast error quantile could be estimated by using $\theta = 0.95$ in the quantile regression minimisation in (5) with the forecast error vector as dependent variable and vectors k and k^2 as independent variables. The result would be a model of the form

$$Q_{fe}(0.95) = a + bk + ck^2,$$

where a, b and c are coefficients estimated by quantile regression. Yar and Chatfield (1990) note that one advantage of the theoretical method is that the formulae give insight into how the forecast error variance varies with k. A similar point can be made for our new proposal.

While an exponential smoothing model is fitted using just 1-step-ahead forecast errors, the new proposal constructs predictive distributions using errors from several different step-ahead forecasts. Consequently, using errors from the fit period as proxy for post-sample errors is quite reasonable. The theoretical approach, on the other hand, uses only the 1-step-ahead fit errors.

In view of technological development, we feel that the computational intensity of the approach is unlikely to be a significant constraint. A limitation of the method is that, since it is empirically based, it may not be efficient for small data sets. From a theoretical perspective, there may be inefficiencies due to the likely correlation between dependent variable observations. Unlike the ordinary least squares case, it is not clear how to handle this in quantile regression. However, our research has an applied agenda and, to avoid overcomplicating a pragmatic proposal, we do not consider this further in this paper.

4.3. Analytical Motivation for the Choice of Regressors

The simple exponential smoothing formulae were presented in (1). The theoretical prediction intervals are based on the variance expression in (2) and the assumption of normality. We might motivate our choice of explanatory variables for the quantile regression by considering (2) a little further. Using (2), the θth k-step-ahead forecast error quantile might be estimated as

$$Q_{fe}(\theta) = Z_\theta \sigma_e [1 + (k-1)\alpha^2]^{1/2}. \qquad (7)$$

Z_θ is the θth quantile of the standard normal distribution. If $(k-1)\alpha^2 < 1$, one could expand as

$$Q_{fe}(\theta) = Z_\theta \sigma_e [1 + \tfrac{1}{2}(k-1)\alpha^2 - \tfrac{1}{8}(k-1)^2\alpha^4 + \cdots].$$

We might then expect k and k^2 to be reasonable regressors when $(k-1)\alpha^2 < 1$ is satisfied. Alternatively, one could rearrange (7) as

$$Q_{fe}(\theta) = Z_\theta \sigma_e k^{1/2} \alpha [1 + (1-\alpha^2)/(k\alpha^2)]^{1/2}.$$

If $(1-\alpha^2) < k\alpha^2$, one could expand this as

$$Q_{fe}(\theta) = Z_\theta \sigma_e k^{1/2} \alpha [1 + (1-\alpha^2)/(2k\alpha^2) - (1-\alpha^2)^2/(8k^2\alpha^4) + \cdots].$$

One might then expect $k^{1/2}$ and $k^{-1/2}$ to be reasonable regressors when $(1-\alpha^2) < k\alpha^2$.

When k and α do not satisfy either of the two inequalities, $(k-1)\alpha^2 < 1$ and $(1-\alpha^2) < k\alpha^2$, the expansion of expression (7) will be less simple. If the data are not generated by an ARIMA(0, 1, 1) process for which the parameter is known, then motivating our choice of explanatory variables by expanding expression (7) is less appealing. However, in these cases it still seems reasonable to assume that the true forecast error quantile expression will be some function of k, so using simple functions of k as explanatory variables seems sensible. Similar analysis can be used to motivate choice of explanatory variables for other exponential smoothing approaches.

4.4. Testing for Significance of Explanatory Variable Coefficients

The statistical package STATA (Stata 1993), performs quantile regression (using linear programming) and provides estimates for the standard errors of the coefficients based on a method of Koenker and Bassett (1982). Although adequate for homoscedastic errors, the method understates the standard errors for the heteroscedastic case (Rogers 1992). Consequently, STATA also allows the standard errors to be estimated by bootstrap resampling (Gould 1992). STATA provides the following measure of fit, which uses the loss function in expression (5),

$$\text{pseudo } R^2 = 1 - \frac{\text{sum of weighted deviations about estimated quantile}}{\text{sum of weighted deviations about raw quantile}},$$

where the raw quantile is given by quantile regression with no regressors.

The choice of explanatory variables to use in the quantile regressions could thus be based on an inspection of standard errors of the coefficients and, to a lesser extent, on the pseudo R^2 statistic. In our work we found that, on the whole, the models involving k alone, $k^{1/2}$ alone, k and $k^{1/2}$ together, and k and k^2 together gave reasonable t-values for a range of quantiles considered. The pseudo R^2 tended to be highest when using a pair of regressors, although the actual values that resulted were less than 10%. The values are low because the application of quantile regression to model forecast error only attempts to model heteroscedasticity.

5. Simulation Study

We carried out a simulation study to compare the quantile regression procedure to the theoretical method and to an empirical approach which uses Gardner's (1988) fit errors and a normality assumption. The motivation for using the controlled environment of simulated data is that one can monitor carefully the conditions under which one method outperforms another. Exponential smoothing is more suited to a simulation study than many forecasting methods because it does not involve a formal model identification procedure. Indeed, in inventory control, for example, exponential smoothing is sometimes employed with little or no prior inspection of the data. We considered simple, Holt's, and damped Holt's exponential smoothing. These are three of the most popular approaches, particularly in view of the common tendency to deseasonalise data before smoothing. The most fundamental requirement of a quantile estimate for a predictive distribution is that the probability of an observation falling below the θth quantile is θ% (Meade and Islam 1995). Our study used this criterion as a basis for comparison.

5.1. Simulation Procedure

Each iteration of the simulation involved generating a series of 98 observations from a prespecified process. The first 80 were used to build the exponential smoothing model which was then used to make 1, 3, 6, 9, 12, 15, and 18-step-ahead post-sample forecasts. Using the same 80 observations, the 75th and 95th forecast error quantiles corresponding to these seven forecasts were then estimated using the theoretical, empirical, and quantile regression methods. Subtracting the 1, 3, 6, 9, 12, 15, and 18-step-ahead forecasts from their respective generated actuals gave a post-sample forecast error for each step-ahead. We then recorded whether each forecast error was greater or less than each of the respective quantiles.

We used 400 iterations for each process. The quantile regressions used 60 fit errors from each step-ahead. The selection of lead times reflected those used by Yar and Chatfield (1990). The 75th and 95th quantiles were chosen to represent the body and tail of the distribution.

We considered three sets of quantile regression explanatory variables in the simulations: k alone, $k^{1/2}$ alone, k and k^2 together. In practice, the regressors could be chosen after inspection of t-values and pseudo R^2 for each time series considered. The method was thus applied in a rather naive way in order to make the simulation exercise feasible.

The simulations were carried out using Gauss for UNIX.[1] The smoothing parameters were estimated by the traditional method of minimising the sum of the squared *ex post* 1-step-ahead forecast errors (Gardner 1985). The initial values for the parameters were calculated by backforecasting (Ledolter and Abraham 1984). It is the convention of most of the popular commercial forecasting packages to use parameters between zero and one. Although wider parameter intervals can be justified (Gardner 1988), we were keen not to distance our study from common practice, so

[1] Gauss for UNIX uses Kinderman and Ramage's (1976) fast acceptance-rejection algorithm to generate normal random numbers.

we restricted the parameters to lie within these bounds. We only considered series for which the autocorrelation of the 1-step-ahead errors was not significant at the 1% level. This seems sensible as Chatfield (1993) proposes using the theoretical formula only when there is no autocorrelation in the residuals.

5.2. Simple Exponential Smoothing

We generated time series from simple ARIMA processes. Using both k and k^2 as regressors in the quantile regression method proved to be more successful than using k or $k^{1/2}$ alone. Figure 1 focuses on six of the generating processes and presents graphical comparisons for estimation of the 75th quantile, corresponding to quantile regression with k and k^2 as regressors. The average value derived for α is denoted by $\bar{\alpha}$ in each of the six plots. For each step-ahead considered (1, 3, 6, 9, 12, 15, and 18), the graphs display the proportion of times out of 400 that the simulated post-sample forecast error fell below the respective quantile estimate. This proportion would ideally be 0.75 and so each graph shows this ideal as a horizontal line.

Simple exponential smoothing is optimal for ARIMA(0, 1, 1) processes, and with 80 observations the parameter is estimated with reasonable accuracy. Thus, we would not expect the new procedure to improve on the theoretical method; however, it is important that it does not perform significantly worse. We found that quantile regression was a little less efficient than the theoretical and empirical methods, as can be seen from the plot in the top left corner of Figure 1. The quantile regression approach is suited to the more common situation where exponential smoothing is applied to data for which it is not optimal. For various simple nontrending ARIMA processes, quantile regression outperformed the theoretical method and matched the empirical approach, as illustrated in the next three plots of Figure 1.

A normal error term was used for all the processes except the bottom two in Figure 1, which used nonnormal error terms with the aim of producing nonnormal forecast error. It seems likely that the effect of the nonnormal error term on the shape of the forecast error is reduced beyond the short term. Chatfield (1993 §4.1) notes that the normal approximation does seem to improve as k increases, at least for an AR(1) model. In the short term, however, since simple smoothing is optimal for ARIMA(0, 1, 1) processes, the forecast error will be similar to the error term, and so a nonnormal error term will lead to a similarly nonnormal forecast error. Comparing the bottom two plots in Figure 1 with the plot in the top left, which was identical but with a normal error term, the theoretical and empirical estimates worsened, particularly in the short term, whilst the quantile regression estimates did not.

For the 95th quantile, quantile regression and, to a lesser extent, the empirical approach were noticeably worse than the theoretical method for ARIMA(0, 1, 1) processes beyond the first step-ahead. For the other processes considered, similar comments can be made to those for estimation of the 75th quantile, except that the superiority of quantile regression was smaller. To investigate whether the relative estimation quality was different for the lower tail of the distribution, we repeated the simulation exercise for the 25th and 5th quantiles using the bottom two processes in Figure 1. The results were similar to those for the 75th and 95th quantiles.

5.3. Holt's Linear Trend and Damped Trend Exponential Smoothing

We repeated the simulation exercise for the Holt's and damped Holt's methods using a range of data generating processes. As the results were similar for the two, we only report the damped Holt's analysis in detail.

At time t the k-step-ahead predictor $\hat{X}_t(k)$ is given by the damped Holt's formulation as

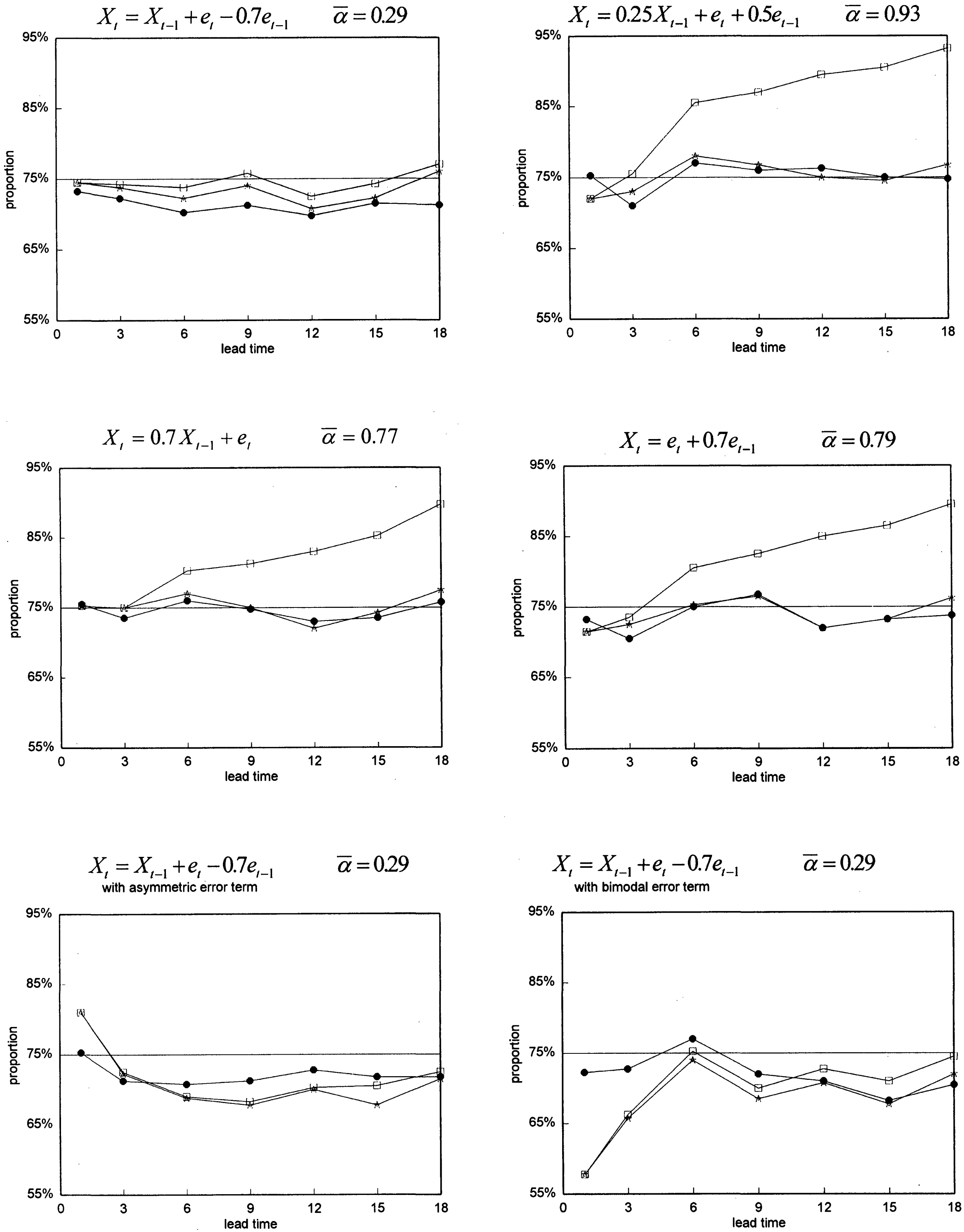
$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + \phi T_{t-1})$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)\phi T_{t-1}$$

$$\hat{X}_t(k) = S_t + \sum_{i=1}^{k} \phi^i T_i.$$

S_t and T_t are the smoothed level and trend respectively. This model was first proposed by Gardner and McKenzie (1985). Rosas and Guerrero (1994) remark that, although not frequently employed as yet, the

Figure 1 **Proportion of Post-sample Forecast Errors Below 75th Quantile Estimates for Simple Exponential Smoothing Applied to Simulated Data**
□ Theoretical ∗ Empirical • Quantile Regression Using k and k^2



damped exponential smoothing models usually perform better than their nondamped counterparts. Roberts (1982) showed that these forecasts are optimal for the ARIMA(1, 1, 2) process with the conditions that $\theta_1 = 1 + \phi - \alpha - \phi\alpha\gamma$ and $\theta_2 = -\phi(1 - \alpha)$.

$$X_t = (1 + \phi)X_{t-1} - \phi X_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}.$$

Assuming that this is the true underlying process and that the parameters are known, the theoretical expression for the variance of the k-step-ahead forecast errors is

$$\mathrm{var}(e_t(k)) = \sigma_e^2\Bigg[1 + (k-1)\alpha^2 + \frac{2\phi}{(1-\phi)}(k-1)\alpha^2\gamma + \frac{\phi^2}{(1-\phi)^2}(k-1)\alpha^2\gamma^2 + \frac{\phi^4(1-\phi^{2k-2})}{(1-\phi)^3(1+\phi)}\alpha^2\gamma^2\Bigg]. \quad (8)$$

The theoretical estimate of the θth k-step-ahead forecast error quantile is based on (8) and an assumption of normality. Similarly to the simple exponential smoothing case, we expanded the square root of expression (8) to motivate our choice of explanatory variables to use in the quantile regression method. This highlighted k, k^2, $k^{1/2}$, $k^{-1/2}$, and $k^{3/2}$ as potentially useful variables.

Trending series were used in the simulation analysis. We only considered processes for which the smoothing parameters generally were not at their bounds of zero or one. We had difficulty finding processes which satisfied this condition, apart from ARIMA(1, 1, 2) processes for which damped Holt's is optimal. We only used normal error terms as the simple smoothing study had investigated the case of nonnormal forecast errors.

In comparing the results for the three sets of regressors used in the quantile regression method, we found a slight improvement in the estimation for the short term when using k and k^2 together. In Figure 2 we present the graphs corresponding to estimation of the 75th quantile with quantile regression using k and k^2 as regressors. Interestingly, quantile regression and the empirical approach matched the theoretical method for the ARIMA(1, 1, 2) processes for which damped Holt's is optimal. Quantile regression was superior for the processes with deterministic trend. For the 95th quantile, in the medium to long term, quantile regression was outperformed by the theoretical formula, and to a lesser extent by the empirical approach, except for processes with a deterministic trend for which the methods fared similarly.

The results for Holt's method were broadly similar to those for damped Holt's. One difference was that using both k and k^2 as regressors was a radical improvement over using k and $k^{1/2}$ alone. Since Holt's method is optimal for ARIMA(0, 2, 2) processes, we were surprised to find that the new approach, using k and k^2, and the empirical method were more successful than the theoretical method for one of two ARIMA(0, 2, 2) processes considered.

5.4. Summary of the Simulation Study

Using k and k^2 together as regressors proved to be more successful than k or $k^{1/2}$ alone. Quantile regression produced better results for the 75th quantile than for the 95th. The new method was able to accommodate nonnormal forecast error whilst the traditional methods ran into trouble. Success was recorded for a range of simple data generating processes, including those for which the exponential smoothing methods are optimal. The method was particularly efficient when slight deterministic trend was present. The simulation work indicates that the quantile regression method tends to underestimate in the medium and long term, particularly for the quantiles in the extremes of the distribution.

One potential weakness of the quantile regression method lies in the inability of the fit errors to act as proxy for the true post-sample forecast errors. In §2.2, we described the empirical method of Williams and Goodman (1971) for constructing prediction intervals from post-sample forecast errors. Surprisingly, using these errors in the quantile regression approach produced very similar results to quantile regression applied to fit errors.

6. Real Time Series Study

In order to further investigate the usefulness of the new approach, we carried out comparative analysis using real time series. The data used was a subset of

Figure 2 **Proportion of Post-sample Forecast Errors Below 75th Quantile Estimates for Damped Holt's Exponential Smoothing Applied to Simulated Data**
□ Theoretical ∗ Empirical • Quantile Regression Using k and k^2

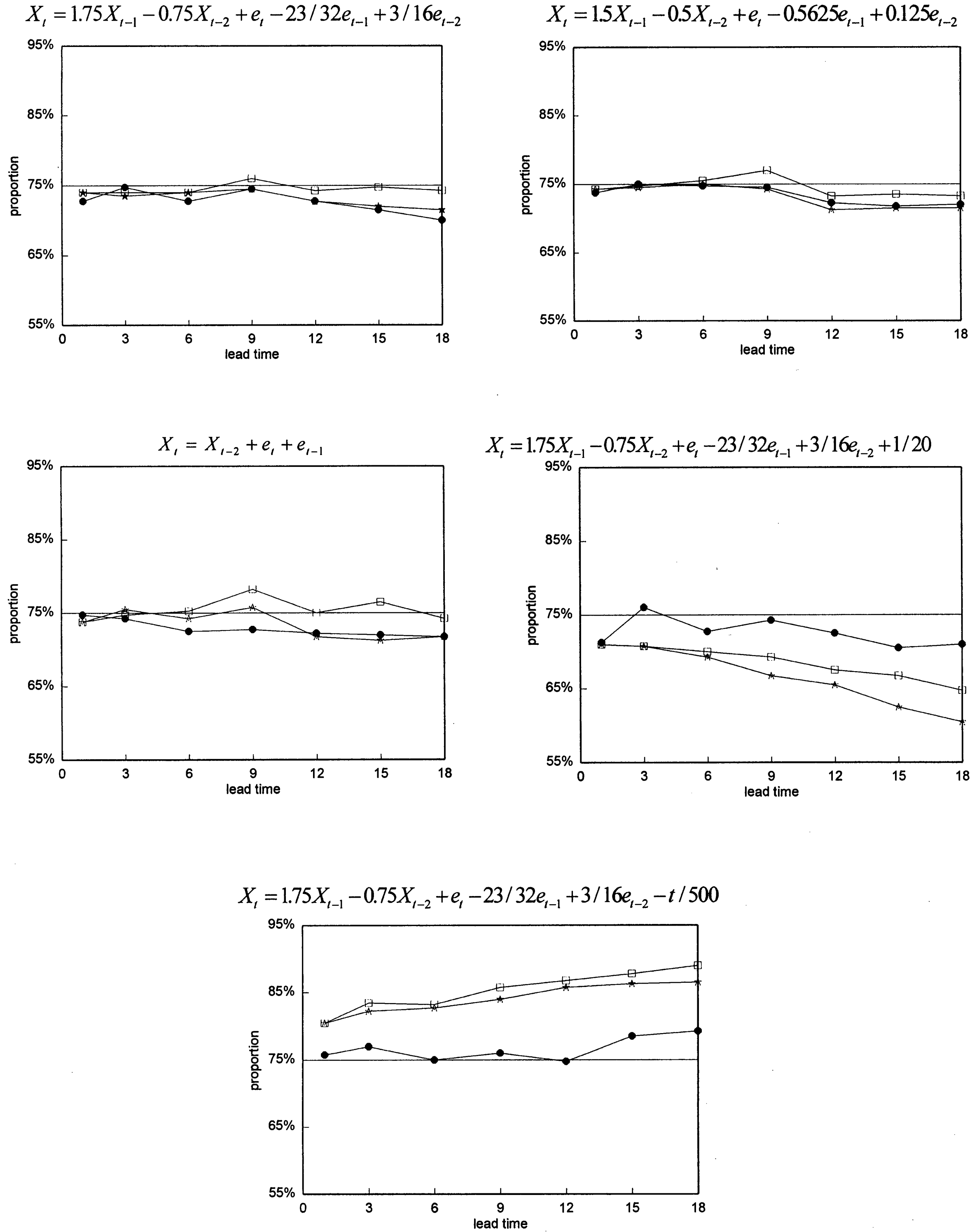
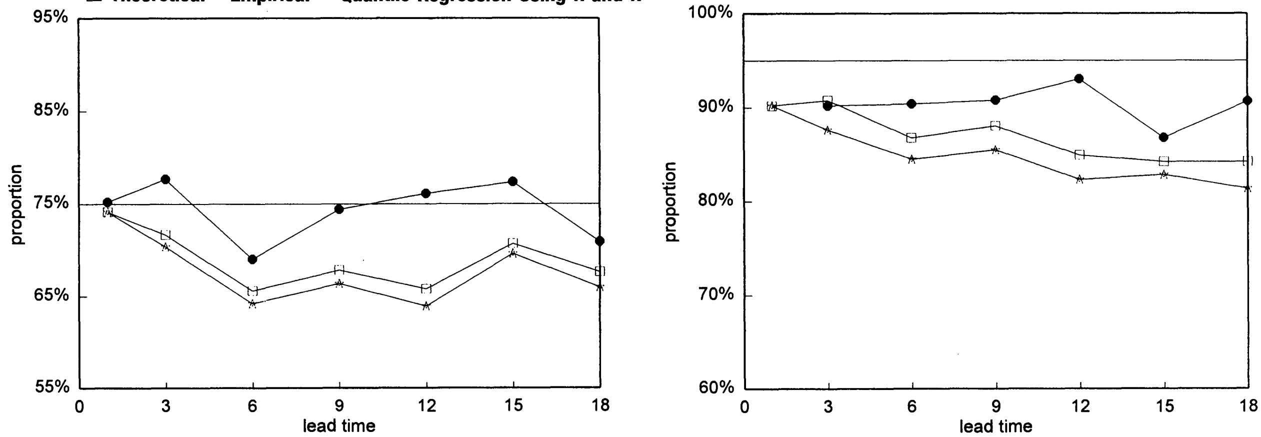


Figure 3 **Proportion of Post-sample Forecast Errors Below 75th and 95th Quantile Estimates for Simple Exponential Smoothing Applied to a Subset of the M3-Competition Data**

□ Theoretical * Empirical • Quantile Regression Using *k* and k^2



the 3,003 time series from the M3-Competition.[2] The subset contained the 1,020 series which consisted of at least 98 observations. Restriction to these series enabled us to use 80 observations to fit the exponential smoothing parameters and 18 to assess the quality of the estimated forecast error quantiles (as in the simulation study). The 1,020 were a mixture of industry, demographic, meteorological, financial, and economic series. All consisted of monthly observations, so prior to forecasting and quantile estimation, we deseasonalised the data, using the decomposition method, based on ratio-to-moving averages, which was used in the M-Competition (Makridakis et al. 1982).

As in the simulation study, we compared theoretical, empirical, and quantile regression estimation of the 75th and 95th quantiles for simple, Holt's, and damped Holt's exponential smoothing. We assessed estimation quality by recording the proportion of times (out of 1,020) that the post-sample forecast error fell below the respective quantile estimate. We found that using k and k^2 together as regressors in quantile regression was more successful than k or $k^{1/2}$ alone. Figures 3 to 5 show graphs comparing theoretical and empirical quantile estimation with quantile regression using k and k^2 as regressors.

Quantile regression was better for the 75th quantile than for the 95th, as shown in Figures 3 and 4 for the simple and damped Holt's methods. Figure 3 shows that quantile regression was the best approach for the simple method. The theoretical approach was the most successful for the 95th quantile for Holt's and damped Holt's, as illustrated in Figure 4 for damped Holt's.

We also investigated estimation of the lower tail of the predictive distribution. The results of Figure 5 for the damped Holt's method were typical, indicating impressive performance by the theoretical approach for the 5th quantile. Disappointingly, the quantile regression method underestimated the lower half of the distribution, particularly for 15- and 18-steps-ahead.

Comparison of Figures 4 and 5 highlights the apparent consistency of the quantile regression procedure which underestimated the 5th and 95th quantiles to a similar extent. In contrast, the empirical and theoretical approaches estimated the lower tail with greater accuracy than the upper tail. The nonnegativity of all the series, and the fact that they all have an upward trend, suggests that the predictive distribution for many of the series may be skewed. The prediction intervals estimated by the theoretical and empirical methods are symmetric, which perhaps

[2] The M3-Competition is a similar forecasting competition to the M-Competition (Makridakis et al. 1982). It is being organised by Spyros Makridakis and Michel Hibon of INSEAD, France. A summary of the main findings will be published in the *International Journal of Forecasting*.

Figure 4 **Proportion of Post-sample Forecast Errors Below 75th and 95th Quantile Estimates for Damped Holt's Exponential Smoothing Applied to a Subset of the M3-Competition Data**
□ Theoretical ∗ Empirical • Quantile Regression Using k and k^2

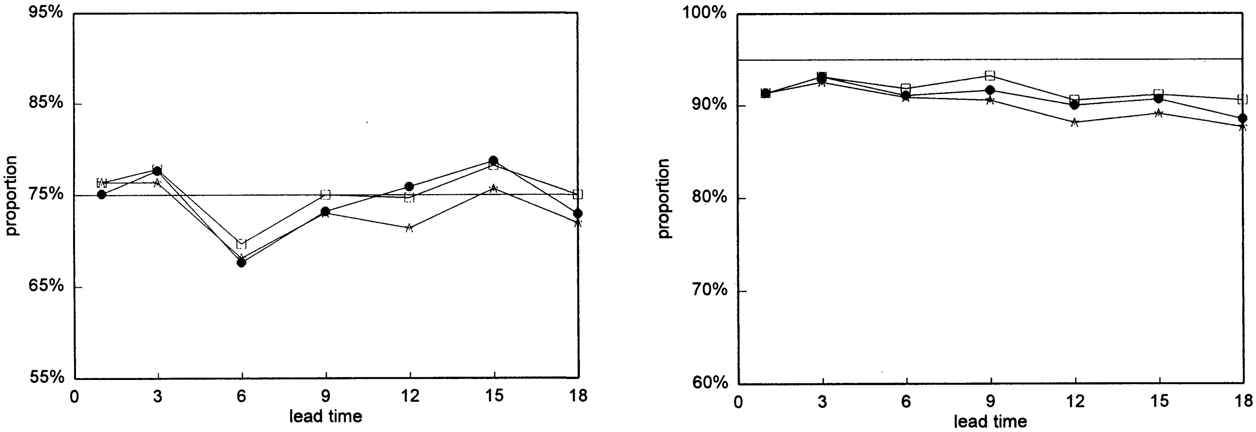
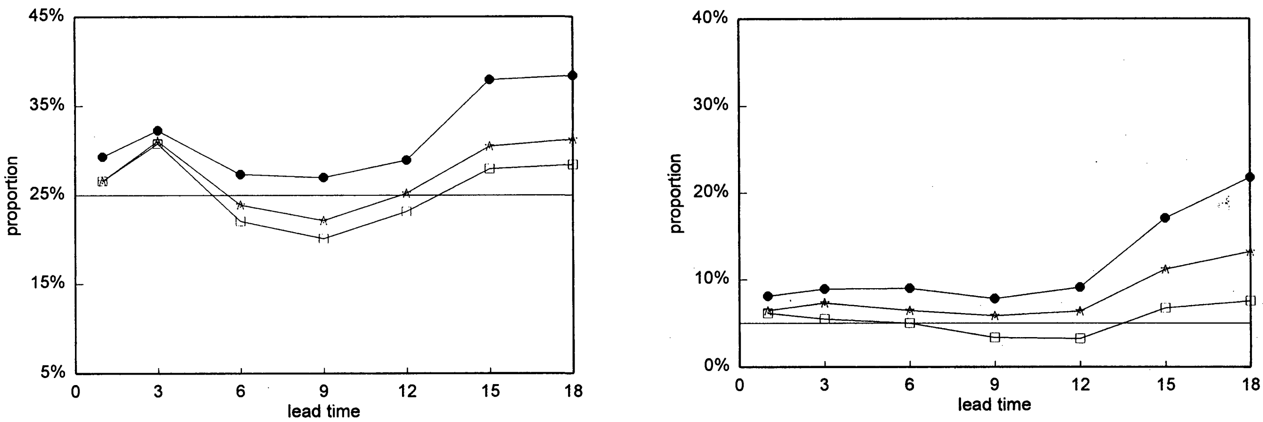


Figure 5 **Proportion of Post-sample Forecast Errors Below 25th and 5th Quantile Estimates for Damped Holt's Exponential Smoothing Applied to a Subset of the M3-Competition Data**
□ Theoretical ∗ Empirical • Quantile Regression Using k and k^2



explains why estimation quality was uneven. It therefore seems that the option is between using an approach which produces consistent (under-) estimation of the tails or using a method which produces more accurate estimation *on average* but fails to model asymmetry.

This raises the issue that we should be cautious when inferring from studies such as this which look at average performance for many different series. The ability of one method to outperform another on average is an admirable feature, but it tells us little about the variability in performance, and thus robustness is not addressed. Unfortunately, it is not obvious how one might assess the variability of quantile estimation performance across different series. Although similar problems exist for simulation studies, the controlled environment does enable one to generate series with known characteristics and hence investigate issues such as asymmetric distributions more carefully. It is also worth noting that in the simulation and real data studies, the quantile regression procedure was applied in a simplistic way. In practice, one might seek to specify regressors for the method which are more appropriate to an individual series.

In summary, the analysis of real data showed that the theoretical method was generally the most accurate for the tails of the distribution beyond the short term. The study also highlighted the potential advantage of the quantile regression approach for nonnormal distributions. However, overall, the work did not provide conclusive support for any one method.

7. Concluding Comments

In this paper we have presented a new approach to assessing the predictive distribution for exponential smoothing methods. The method is empirically based but produces models for the forecast error quantiles in terms of k, the forecast lead time. Simulation and real data studies gave promising results, although beyond the short term, quantile regression appears to underestimate quantiles in the tail of the distribution.

This approach should be useful for various other forecasting applications. Since exponential smoothing is often applied with parameters not estimated by a formal analysis (e.g., minimising 1-step-ahead errors), the appropriateness of the theoretical formulae would be in doubt in this situation. Thus, another application would be to exponential smoothing-based forecasts of aggregate or cumulative demand, as the theoretical formulae of Yar and Chatfield (1990) and Johnston and Harrison (1986) again rely on the assumption that the data-generating process is the ARIMA model for which the particular exponential-smoothing method is optimal. Empirically based approaches seem particularly suited to forecasting methods which do not attempt to model the underlying data-generating process. However, even when the process has been modelled, theoretical quantiles and intervals are often difficult to calculate. This is the case for many nonlinear methods and also for much simpler ARIMA models, if one wishes to include uncertainty due to parameter estimation, or one doubts the wisdom of a normality assumption.[3]

[3] The authors would like to acknowledge the helpful comments of two anonymous referees.

References

Bookbinder, J. H., A. Lordahl. 1989. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Trans.* **21** 302–312.

Bowerman, B. L., A. B. Koehler. 1989. The appropriateness of Gardner's simple approach and Chebyshev prediction intervals. Unpublished paper presented at the 9th International Symposium on Forecasting in Vancouver, B.C. June 18–20.

Business Forecast Systems. 1994. Forecast Pro for Windows, Business Forecast Systems Incorporated, Belmont, MA.

Chatfield, C. 1993. Calculating interval forecasts. *J. Business Econom. Statist.* **11** 121–135.

——, M. Yar. 1991. Prediction intervals for multiplicative Holt-Winters. *Internat. J. Forecasting* **7** 31–37.

Gardner, E. S., Jr. 1985. Exponential smoothing: The state of the art. *J. Forecasting* **4** 1–28.

——. 1988. A simple method of computing prediction intervals for time-series forecasts. *Management Sci.* **34** 541–546.

——, E. McKenzie. 1985. Forecasting trends in time series. *Management Sci.* **31** 1237–1246.

Gélinas, R., P. Lefrançois. 1993. On the estimation of time-series quantiles using smoothed order statistics. *Internat. J. Forecasting* **9** 227–243.

Gorr, W. L., C. Hsu. 1985. An adaptive filtering procedure for estimating regression quantiles. *Management Sci.* **31** 1019–1029.

Gould, W. W. 1992. Quantile regression with bootstrapped standard errors. *Stata Tech. Bull.* **9** 19–21.

Johnston, F. R., P. J. Harrison. 1986. The variance of lead-time demand. *J. Oper. Res. Soc.* **3** 303–308.

Kinderman, A. J., J. G. Ramage. 1976. Computer generation of normal random numbers. *J. Amer. Statist. Assoc.* **71** 893–896.

Koenker, R. W., G. W. Bassett. 1978. Regression quantiles. *Econometrica* **46** 33–50.

——, ——. 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50** 43–62.

Ledolter, J., B. Abraham. 1984. Some comments on the initialization of exponential smoothing. *J. Forecasting* **3** 79–84.

Makridakis, S., A. Anderson, R. Carbone, R., Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, R. Winkler. 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecasting* **1** 111–153.

——, R. L. Winkler. 1989. Sampling distributions of post-sample forecast errors. *Appl. Statist.* **38** 331–342.

Meade, N., T. Islam. 1995. Prediction intervals for growth curve forecasts. *J. Forecasting* **14** 413–430.

Roberts, S. A. 1982. A general class of Holt-Winters type forecasting models. *Management Sci.* **28** 808–820.

Rogers, W. H. 1992. Quantile regression standard errors. *Stata Technical Bulletin* **9** 16–19.

Rosas, A. L., V. M. Guerrero. 1994. Restricted forecasts using exponential smoothing techniques. *Internat. J. Forecasting* **10** 515–527.

Stata. 1993. STATA 3.1, Stata Corporation, College Station, TX.

Williams, W. H., M. L. Goodman. 1971. A simple method for the construction of empirical confidence limits for economic forecasts. *J. Amer. Statist. Assoc.* **66** 752–754.

Yar, M., C. Chatfield. 1990. Prediction intervals for the Holt-Winters forecasting procedure. *Internat. J. Forecasting* **6** 127–137.

Accepted by Gabriel Bitran; received September 1996. *This paper has been with the authors* 10 *months for* 2 *revisions.*