Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Econometrics 135 (2006) 377–398

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Interval forecasts and parameter uncertainty

Bruce E. Hansen*

Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706, USA

Available online 30 August 2005

Abstract

Forecast intervals generalize point forecasts to represent and incorporate uncertainty. Forecast intervals calculated from dynamic models typically sidestep the issue of parameter estimation. This paper shows how to construct asymptotic forecast intervals which incorporate the uncertainty due to parameter estimation. Our proposed solution is a simple proportional adjustment to the interval endpoints, the adjustment factor depending on the asymptotic variance of the interval estimates. Our analysis is in the context of a forecasting equation with an error independent of the forecasting variables but with unknown distribution. The methods are illustrated with a simulation experiment and an application to the US monthly unemployment rate.

JEL classification: C53

Keywords: Forecast intervals; Nonparametric; Quantile; Estimation

1. Introduction

Econometric time-series models are routinely used for generating forecasts. Often these take the form of point forecasts,[1] but there has been recent increased attention given to interval forecasts and the closely related density forecast. Interval forecasts

*Tel.: +1 608 263 3880; fax: +1 608 263 3876.
E-mail address: bhansen@ssc.wisc.edu.

[1]See, for example, the classical work by Granger and Newbold (1986).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.07.030

are typically computed from point estimates of the conditional distribution function, without an adjustment for parameter estimation error. This is in contrast to the treatment of prediction intervals found in standard econometrics textbooks, which derives an exact prediction interval for the case of independent data with normal errors. The contribution of this paper is to show how to adjust forecast intervals to account for estimation error. Our recommendation is that forecast interval endpoints should be proportionately adjusted, with the adjustment factor depending on the sample size and the asymptotic variance of the point estimate of the forecast interval endpoints. Adjusted intervals endpoints can also be calculated as the quantiles from the convolution of the estimated conditional density function with the density of the estimation error.

A clean and explicit description of forecast interval construction can be found in Granger et al. (1989). A thorough review of a variety of methods can be found in Chatfield (1993). Zarnowitz and Braun (1993) provide a detailed description of the history of the NBER-ASA economic forecasts. Granger (1996) discusses issues regarding the desired nominal coverage level of forecast intervals. Christoffersen (1998) discusses the evaluation of interval forecasts. Diebold et al. (1998) introduce methods for evaluation of density forecasts, and Diebold et al. (1999a) extend the method to bi-variate data. Tay and Wallis (2000) survey density forecasts, and Wallis (2003a) proposes tests of interval forecasts. Related discussions can be found in Granger and Pesaran (1999), Diebold et al. (1999b), and Wallis (2003b). Corradi and Swanson (2004) introduce bootstrap methods for selection between misspecified predictive density models. Bayesian forecast intervals are proposed in Otrok and Whiteman (1997, 1998), Robertson et al. (2003) and Cogley et al. (2003).

As a practical example of interval forecasts, the Bank of England publishes a quarterly *Inflation Report* which includes a *fan chart* for the future path of inflation. These are graphs of the deciles of the forecast distribution, and are equivalent to forecast intervals. For details see Britton et al. (1998) and Bean and Jenkinson (2001).

With the exception of the Bayesian methods, none of the papers listed in the prior two paragraphs discuss the impact of parameter estimation error upon forecast interval construction. Our paper is a first step in that direction in a classical statistical framework.

We follow the recent forecasting literature (e.g. Stock and Watson (1999, 2003)) and take a direct approach to k-step ahead prediction by specifying a k-step ahead conditional distribution function. This is in contrast to the plug-in method which iterates a one-step-ahead prediction k times. The direct approach is easist for implementation, estimation, and inference, and is more robust against specification error. However, the plug-in approach may be more efficient when the one-step-ahead forecast model is correctly specified. For discussions see Tiao and Tsay (1994), Fan and Yao (2003, Section 10.1.4), and Ing (2003).

Our framework is semiparametric. We specify a parametric forecasting equation which reduces the time-series to an independent error whose distribution is unknown and nonparametric. This is contrast to much of the density forecasting literature which is fully parametric.

An important limitation of our analysis is that it is confined to the case where the forecasting error is independent of the conditioning variables. This precludes misspecification. In the context of linear forecasting models, for example, the independence assumption excludes ARCH effects. Our analysis is also confined to simple asymptotic approximations to the static converage probability of forecast intervals. Issues regarding dynamic coverage properties are not investigated. We also do not consider the issue of forecast interval comparisons or evaluation, as this is treated elsewhere (e.g. Christoffersen, 1998).

Section 2 presents our linear k-step-ahead univariate forecasting model. Section 3 discusses our two-step quantile estimator, and presents the asymptotic distribution of the quantile estimates. Section 4 presents an estimate of the asymptotic variance. Section 5 examines rough (conventional) forecast intervals. Section 6 introduces our new corrected forecast intervals. We show that the coverage probability of the nonparametrically corrected intervals has an improved convergence rate relative to the uncorrected intervals. Section 7 presents an alternative correction based on a convolution argument. Section 8 presents a small simulation experiment. Section 9 illustrates the methods with an application to the US unemployment rate. Some possible extensions are discussed in Section 10. The Appendix contains the proofs of the main Theorems.

Following standard convention, let $\Phi(x)$ and $\phi(x)$ denote the standard normal cdf and pdf, respectively, and let $\Phi\sigma(x) = \Phi(x/\sigma)$ and $\phi_\sigma(x) = \sigma^{-1}\phi(x/\sigma)$ denote the $\mathrm{N}(0, \sigma^2)$ cdf and pdf.

2. Forecasting model

Consider a stationary univariate time series y_t and let x_t denote a conditioning set such as the recent history $x_t = \{y_t, y_{t-1}, \ldots, y_{t-l+1}\}$. Let $\{y_{t+k}, x_t : t = 1, \ldots, n\}$ denote the sample of observations. We are interested in developing a forecast interval for y_{t+k} given $x_t = x$. This will require estimation of the k-step-ahead conditional distribution function (CDF) of y_{t+k} given $x_t = x$

$$G(y|x) = \mathrm{P}(y_{t+k} \leqslant x | x_t = x). \tag{1}$$

For some $0 < \alpha < 1$ let the αth conditional quantile of y_{t+k} given $x_t = x$ be denoted ξ, that is, the solution to the equation $\alpha = G(\xi|x)$. Also, let $g(y|x) = (\mathrm{d}/\mathrm{d}y)\, G(y|x)$ denote the conditional density function.

Our approach is appropriate for semiparametric forecasting models where a parametric equation reduces the time-series to a nonparametric independent error. The parametric component is the equation

$$h(y_{t+k}, x_t, \beta) = e_t. \tag{2}$$

The $d \times 1$ parameter β is unknown while the function h is known. The latter is strictly increasing in the argument y and has the inverse relationship

$$y_{t+k} = \eta(x_t, \beta, e_t),$$

where η is strictly increasing in its argument e. The error e_t is assumed independent of x_t, although it is k-dependent since it is a k-step forecast error. Let

$$F(z) \equiv \mathrm{P}(e_t \leqslant z)$$

and $f(z) = (\mathrm{d}/\mathrm{d}z)F(z)$ denote the unknown marginal distribution function and density function of e_t. Let q denote the αth quantile of F.

An example of (2) is the linear forecasting model

$$y_{t+k} = x_t'\beta + e_t. \tag{3}$$

Here, $h(y, x, \beta) = y - x'\beta$ and $\eta(x, \beta, e) = x'\beta + e$. Another example is a linear model with heteroskedasticity

$$y_{t+k} = x_t'\beta_1 + \sigma(x_t, \beta_2)e_t, \tag{4}$$

in which case $h(y, x, \beta) = \left(y - x'\beta_1\right)/\sigma(x, \beta_2)$ and $\eta(x, \beta, e) = x'\beta_1 + \sigma(x, \beta_2)e$.

Under our assumptions, the k-step-ahead CDF for y_{t+k} is

$$G(y|x) = F(h(y, x, \beta)) \tag{5}$$

and the αth conditional quantile of y_{t+k} given $x_t = x$ is

$$\xi = \eta(x, \beta, q).$$

Observe that $q = h(\xi, x, \beta)$. Sometimes it will be convenient to write $\theta = (\beta, q)$ and $\eta(x, \theta) = \eta(x, \beta, q)$. We will let (θ, β, q, ξ) denote generic values of the parameters and let $(\theta_0, \beta_0, q_0, \xi_0)$ denote the unknown true values. The model is semiparametric since the Eq. (2) is parametric while the distribution F is nonparametric.

The assumption that e_t is independent of x_t is quite strong, and requires that the model (2) is correctly specified. In practice, this is unlikely to be accurate, and therefore, our methods should be viewed as an approximation to the realistic case of a misspecified forecasting model.

Our goal is to construct a forecast interval for y_{t+k} conditional on $x_t = x$. Observe that the one-sided forecast interval $(-\infty, \xi]$ has exact coverage α. Since ξ is unknown it must be estimated. We turn to this problem in the next section.

3. Two-step quantile estimation

Let $\hat{\beta}$ be an estimator of the parameter β in Eq. (2) which can be written as an approximate method of moments estimator. That is for some function $l_t(\beta)$, $\hat{\beta}$ satisfies

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} l_t(\hat{\beta}) = \mathrm{o}(1) \tag{6}$$

and the function $l_t(\beta)$ is continuously differentiable in β. This estimation framework includes least-squares, generalized least-squares, generalized method of moments, maximum likelihood, and quasimaximum likelihood. Let $\hat{e}_t = h(y_{t+k}, x_t, \hat{\beta})$ denote the estimation residuals. The distribution function F is estimated from the empirical distribution function (EDF) of the residuals $\hat{e}_t$:

$$\hat{F}(z) = \frac{1}{n}\sum_{t=0}^{n-1} 1(\hat{e}_t \leqslant z).$$

The αth quantile q is estimated by $\hat{q}$, the αth empirical quantile of the residuals, which solves $\alpha = \hat{F}(\hat{q})$. The estimator of the CDF is

$$\hat{G}(y|x) = \hat{F}(h(y, x, \hat{\beta})).$$

and that of the αth conditional quantile is

$$\hat{\xi} = \eta(x, \hat{\beta}, \hat{q}). \tag{7}$$

Note that the conditional distribution $\hat{G}(y|x)$ is discrete and puts probability mass n^{-1} at each of the n points $\eta(x, \hat{\beta}, \hat{e}_t)$.

To derive the asymptotic distribution of $\hat{\xi}$, we use the following set of regularity conditions.

Assumption 1. Let $r > 1$.

1. The sequence $\{y_t, x_t\}$ is strictly stationary and absolutely regular with mixing coefficients ψ_j such that $\sum_{j=1}^{\infty} j^{1/(r-1)}\psi_j < \infty$.
2. The forecast error e_t defined in (2) is independent of x_t, k-dependent, with marginal distribution function $F(z)$ and continuously differentiable density $f(z)$ with $f(q) > 0$. The conditional density of y_{t+k} given $x_t = x$ is bounded: $g(y|x) \leqslant \overline{g} < \infty$.
3. For all β, $\mathrm{E}|l_t(\beta)|^{2r} < \infty$.
4. $\mathrm{E}l_t(\beta) = 0$ only if $\beta = \beta_0$.
5. $\mathrm{E}l_t l_t' = L > 0$ and $\mathrm{rank}(l_\beta) = d$ where $l_t = l_t(\beta_0)$ and $l_\beta = (\partial/\partial\beta')\,\mathrm{E}l_t(\beta_0)$.
6. For some $C < \infty$ and all β

$$\mathrm{E} \sup_{\beta_1 : |\beta - \beta_1| < \delta} |l_t(\beta) - l_t(\beta_1)|^{2r} \leqslant C\delta.$$

7. The function $\eta(x, \theta) = \eta(x, \beta, q)$ satisfies

$$\sup_{|\theta - \theta_1| \leqslant \delta} |\eta(x, \theta_1) - \eta(x, \theta)| \leqslant a(x)\delta$$

with $\mathrm{E}a(x_t) < \infty$.

Theorem 1. *Under Assumption* 1,

$$\sqrt{n}(\hat{\xi} - \xi) \rightarrow^d \mathrm{N}(0, \sigma_\xi^2), \tag{8}$$

where

$$\sigma_\xi^2 = \sum_{j=-k}^{k} Eu_t u_{t+j}, \tag{9}$$

$$u_t = \eta_q(x)\left(\frac{1(e_t \leqslant q_0) - \alpha}{f(q_0)}\right) + (\eta_\beta(x)' - \eta_q(x)\gamma')I_\beta^{-1} l_t,$$

$$\eta_\beta(x) = \frac{\partial}{\partial \beta}\eta(x, \beta_0, q_0),$$

$$\eta_q(x) = \frac{\partial}{\partial q}\eta(x, \beta_0, q_0),$$

$$\gamma = \mathrm{E}\left(\frac{\partial}{\partial y} h(\eta(x_t, \theta_0), x_t, \beta_0)\eta_\beta(x_t)\right).$$

For the linear model (3), $\eta_q = 1$, $\eta_\beta(x) = x$, and $\gamma = \mathrm{E}x_t$, so

$$u_t = \left(\frac{1(e_t \leqslant q_0) - \alpha}{f(q_0)}\right) - (x - \mathrm{E}x_t)'(\mathrm{E}x_t x_t')^{-1} x_t e_t. \tag{10}$$

The covariance takes the form (9) since the forecast errors are k-dependent. For convenience, define $s_\xi^2 = n^{-1}\sigma_\xi^2$ so that we have the approximation

$$\hat{\xi} \sim \mathrm{N}(\xi_0, s_\xi^2). \tag{11}$$

Notice that the asymptotic variance is written in terms of the random variable u_t (see (10)) which has two components. The first is due to the nonparametric estimation of F. The second reflects the estimation of β. Notice, however, that this latter term disappears in the linear model when $x = \mathrm{E}x_t$, the conditioning variables equal their unconditional mean.

Finally, notice that the asymptotic variance σ_ξ^2 generally will vary with the quantile α and x, although the nature of this dependence is difficult to quantify.

4. Variance estimation

We will need an estimate of the asymptotic variance (9). Given an estimate $\hat{u}_t$ of u_t (discussed shortly) a natural choice is

$$\hat{\sigma}_\xi^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2 + 2\sum_{j=1}^{k}\frac{1}{n}\sum_{t=1}^{n-j} \hat{u}_t \hat{u}_{t+j}. \tag{12}$$

Alternatively, a HAC estimator with data-dependent bandwidth may be used for $\hat{\sigma}_\xi^2$, such as those discussed in Andrews (1991), Newey and West (1994), West (1997) and Hirukawa (2005). This could be viewed as a robust estimator, and might have better properties when Eq. (2) is misspecified (in which case u_t is not k-dependent).

Our estimate $\hat{u}_t$ is

$$\hat{u}_t = \hat{\eta}_q(x)\left(\frac{1(\hat{e}_t \leqslant \hat{q}) - \alpha}{\hat{f}(\hat{q})}\right) + (\hat{\eta}_\beta(x)' - \hat{\eta}_q(x)\hat{\gamma}')\hat{I}_\beta^{-1}\hat{l}_t,$$

$$\hat{\eta}_q(x) = \frac{\partial}{\partial q}\,\eta(x, \hat{\beta}, \hat{q}),$$

$$\hat{\eta}_\beta(x) = \frac{\partial}{\partial \beta}\,\eta(x, \hat{\beta}, \hat{q}),$$

$$\hat{l}_\beta = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \beta'}\, l_t(\hat{\beta}),$$

$$\hat{l}_t = l_t(\hat{\beta}).$$

Our estimate for $f(q)$ is

$$\hat{f}(\hat{q}) = \frac{1}{n}\sum_{t=1}^{n} \phi_{r_0}(\hat{q} - \hat{e}_t), \tag{13}$$

a Gaussian kernel estimator with bandwidth r_0. An estimate of the asymptotic MSE-minimizing choice for the bandwidth r_0 is

$$\hat{r}_0 = \left(\frac{\hat{f}_{s_0}(\hat{q})}{2\sqrt{\pi}(\hat{f}^{(2)}_{s_2}(\hat{q}))^2 n} \right)^{1/5}, \tag{14}$$

where

$$\hat{f}_{s_0}(\hat{q}) = \frac{1}{n}\sum_{t=1}^{n} \phi_{s_0}(\hat{q} - \hat{e}_t), \tag{15}$$

$$\begin{aligned} \hat{f}^{(2)}_{s_2}(\hat{q}) &= \frac{1}{n}\sum_{t=1}^{n} \phi^{(2)}_{s_2}(\hat{q} - \hat{e}_t), \\ &= \frac{1}{ns_2}\sum_{t=1}^{n} \left(\left(\frac{\hat{e}_t - \hat{q}}{s_2} \right)^2 - 1 \right) \phi_{s_2}(\hat{q} - \hat{e}_t) \end{aligned} \tag{16}$$

and s_0 and s_2 are preliminary bandwidths. Letting $\hat{\sigma}$ denote the sample standard deviation for $\hat{e}_t$, we suggest the Gaussian reference rules $s_0 = 1.06\hat{\sigma}n^{-1/5}$ and $s_2 = 0.94\hat{\sigma}n^{-1/9}$, which are MISE-optimal for estimation of the density and its second derivative when f is Gaussian.

Finally, given $\hat{\sigma}^2_\xi$, we can define the asymptotic standard error for $\hat{\xi}$

$$\hat{s}_\xi = n^{-1/2}\hat{\sigma}_\xi. \tag{17}$$

5. Forecast intervals

Recall, our goal is to construct a forecast interval for y_{t+k} conditional on $x_t = x$. As discussed before, an exact one-sided forecast interval is $(-\infty, \xi_0]$. Replacing the unknown endpoint ξ_0 by our point estimate $\hat{\xi}$, we obtain the *rough forecast interval*

$(-\infty, \hat{\xi}]$. We call this interval *rough* because it does not incorporate sampling uncertainty due to parameter estimation.

Theorem 2. *Under Assumption* 1,

$$P(y_{n+k} \leqslant \hat{\xi} | x_n = x) = F(q_0) + \tfrac{1}{2}(f(q_0)h_2 + f'(q_0)h_1^2)s_{n\xi}^2 + O(n^{-3/2}), \tag{18}$$

where

$$h_1 = \frac{\partial}{\partial \xi} h(\xi_0, x, \beta_0),$$

$$h_2 = \frac{\partial^2}{\partial \xi^2} h(\xi_0, x, \beta_0).$$

The bias in the conditional coverage probability is of order n^{-1} due to s_ξ^2, the variance of the estimate $\hat{\xi}$. In general, the bias depends on the derivatives h_1 and h_2. However, they take simpler forms in the regression case (either model (3) or (4)) as then $h_2 = 0$, and (18) simplifies to

$$P(y_{n+k} \leqslant \hat{\xi} | x_n = x) = F(q_0) + \tfrac{1}{2} f'(q_0) h_1^2 s_{n\xi}^2 + O(n^{-3/2}). \tag{19}$$

Furthermore, in the homoskedastic regression model (3), $h_1 = 1$ and in the heteroskedastic regression model (4), $h_1 = \sigma(x, \beta_2)^{-1}$.

In (19), the direction of bias depends on the derivative of the density f at q. For the one-sided interval $(-\infty, \hat{\xi}]$, q is in the right tail where typically $f'(q) < 0$ so the leading term in the bias is negative (undercoverage). Similarly, for intervals of the form $[\hat{\xi}, \infty)$, q is in the left tail where typically $f'(q) < 0$ and again the leading term in the bias is negative. Thus Corollary 1 shows that to a first order of approximation, parameter estimation tends to bias downward the coverage probabilites of rough forecast intervals.

Most typically, we are interested in two-sided forecast intervals. Let q_1 and q_2 be the α_1th and α_2th quantile of e_t, etc., where $0 < \alpha_1 < \alpha_2 < 1$ so that the rough forecast interval with nominal coverage $\alpha_2 - \alpha_1$ is $[\hat{\xi}_1, \hat{\xi}_2]$. From (19) this has coverage

$$P(\hat{\xi}_1 \leqslant y_{n+k} \leqslant \hat{\xi}_2 | x_n = x) = \alpha_2 - \alpha_1 + \frac{s_{n\xi}^2 h_1^2}{2}(f'(q_2) - f'(q_1)) + O(n^{-3/2}).$$

Typically, the bias term does not disappear. For example, in the leading case where f is symmetric and $\alpha_2 = 1 - \alpha_1$ then $f'(q_2) = -f'(q_1)$. Since two-sided intervals can always be written as the intersection of two one-sided intervals, we focus on one-sided intervals for the remainder of our formal discussion.

6. Nonparametric forecast intervals

We can correct the coverage probability to order $O(n^{-3/2})$ by adjusting the quantile q to set the right-hand-side of (18) equal to α. That is, let q^* be the value that satisfies

$$\alpha = F(q^*) + \tfrac{1}{2}(f(q^*)h_2 + f'(q^*)h_1^2)s_{n\xi}^2. \tag{20}$$

By a Taylor expansion,

$$q^* = q - \frac{1}{2}\left(h_2 + h_1^2 \frac{f'(q)}{f(q)}\right) s_{n\xi}^2 + O(n^{-3/2}). \tag{21}$$

A simple approximation to (21) can be made using the Gaussian reference density. Suppose that $f(x) = \phi_{\sigma_e}(x)$. Then

$$\frac{f'(q)}{f(q)} = -\frac{q}{\sigma_e^2}.$$

In this case we estimate (21) by

$$\hat{q}_s^* = \hat{q}\left(1 + \frac{h_1^2 \hat{s}_{n\xi}^2}{2\hat{\sigma}_e^2}\right) - \frac{h_2 \hat{s}_{n\xi}^2}{2}, \tag{22}$$

where $\hat{\sigma}_e^2 = n^{-1}\sum_{t=0}^{n-1} \hat{e}_t^2$. We call this the "simple reference adjustment". (While simple, it is not guaranteed to improve performance for error distributions distinct from the Gaussian.) In particular, for the homoskedastic linear regression (3), this takes the form of the proportional adjustment

$$\hat{q}_s^* = \hat{q}\left(1 + \frac{\hat{s}_{n\xi}^2}{2\hat{\sigma}_e^2}\right). \tag{23}$$

Given $\hat{q}_s^*$, we define $\hat{\xi}_s^* = \eta(x, \hat{\beta}, \hat{q}_s^*)$ and define the one-sided corrected forecast interval $(-\infty, \hat{\xi}_s^*]$, and similarly for two-sided intervals.

An alternative is to estimate (21) using the nonparametric estimator (13) and its derivative. For a bandwidth r_1 the latter estimate is

$$\hat{f}'(\hat{q}) = \frac{1}{n}\sum_{t=1}^{n} \phi_{r_1}^{(1)}(\hat{q} - \hat{e}_t) = -\frac{1}{nr_1}\sum_{t=1}^{n}\left(\frac{\hat{q} - \hat{e}_t}{r_1}\right)\phi_{r_1}(\hat{q} - \hat{e}_t).$$

An estimate of the asymptotic MSE-minimizing choice for r_1 is

$$\hat{r}_1 = \left(\frac{3\hat{f}_{s_0}(\hat{q})}{4\sqrt{\pi}(\hat{f}_{s_3}^{(3)}(\hat{q}))^2 n}\right)^{1/7},$$

where $\hat{f}_{s_0}(\hat{q})$ is defined in (15) with $s_0 = 1.06\hat{\sigma}n^{-1/5}$, and

$$\begin{aligned}\hat{f}_{s_3}^{(3)}(\hat{q}) &= \frac{1}{n}\sum_{t=1}^{n} \phi_{s_3}^{(3)}(\hat{q} - \hat{e}_t) \\ &= \frac{1}{ns_3}\sum_{t=1}^{n}\left(3\left(\frac{\hat{e}_t - \hat{q}}{s_2}\right) - \left(\frac{\hat{e}_t - \hat{q}}{s_2}\right)^3\right)\phi_{s_3}(\hat{q} - \hat{e}_t)\end{aligned}$$

with $s_3 = 0.93\hat{\sigma}n^{-1/11}$, the MISE-optimal bandwidth for the third derivative when f is Gaussian.

We then set

$$\hat{q}^*_{np} = \hat{q} - \frac{1}{2}\left(h_2 + h_1^2 \frac{\hat{f}'(\hat{q})}{\hat{f}(\hat{q})}\right) s^2_{n\xi} \tag{24}$$

which we call the "nonparametric adjustment".

Given $\hat{q}^*_{np}$, we define $\hat{\xi}^*_{np} = \eta(x, \hat{\beta}, \hat{q}^*_{np})$ and define the corrected forecast interval $(-\infty, \hat{\xi}^*_{np}]$.

Theorem 3. *Under Assumption* 1,

$$\mathrm{P}(y_{n+k} \leqslant \hat{\xi}^*_{np} | x_n = x) = \alpha + \mathrm{O}(n^{-3/2}), \tag{25}$$

Comparing Theorems 2 and 3, we find that the corrected forecast interval has improved coverage error relative to the rough forecast interval.

7. Convolution adjustment

For the case $h_2 = 0$ we also consider an adjusted interval which avoids the use of a bandwidth or reference density. Note that the accuracy of correct coverage is the probability of the event $\{y_{n+k} \leqslant \hat{\xi}\} = \{e_n \leqslant h(\hat{\xi}, x, \beta_0)\} \simeq \{e_n \leqslant h_1(\hat{\xi} - \xi_0)\}$. Consider the random variable $e_n - h_1(\hat{\xi} - \xi_0)$, let F^* denote its distribution function and q^* is αth quantile. (We use the same notation as in (20)–(21), since we will show that they are equivalent to order $\mathrm{O}(n^{-3/2})$.) By several Taylor expansions

$$\begin{aligned}
\alpha &= F^*(q^*) \\
&= \mathrm{EP}(e_n - h_1(\hat{\xi} - \xi_0) \leqslant q^* | \hat{\xi}) \\
&= \mathrm{E}F(q^* + h_1(\hat{\xi} - \xi_0)) \\
&= \mathrm{E}[F(q^*) + f(q^*)h_1(\hat{\xi} - \xi_0) + \tfrac{1}{2}f'(q^*)h_1^2(\hat{\xi} - \xi_0)^2] + \mathrm{O}(n^{-3/2}) \\
&= F(q^*) + \tfrac{1}{2}f'(q^*)h_1^2 s_\xi^2 + \mathrm{O}(n^{-3/2})
\end{aligned}$$

which is the same as (20) when $h_2 = 0$ as asserted.

Since two definitions for q^* are equivalent to order $\mathrm{O}_p(n^{-3/2})$, an alternative to (24) is to estimate the αth quantile of the distribution of the random variable $e_n - h_1(\hat{\xi} - \xi_0)$. Note this is the sum of two approximately independent components. Thus by the convolution formula and the approximation $h_1(\hat{\xi} - \xi_0) \sim \mathrm{N}(0, s_1^2)$, where $s_1^2 = h_1^2 s_\xi^2$ we have

$$\begin{aligned}
F^*(z) &= \mathrm{P}(e_n - h_1(\hat{\xi} - \xi_0) \leqslant z) \\
&= \int_{-\infty}^{z} \int_{-\infty}^{\infty} \phi_{s_1}(t - u) f(u)\, \mathrm{d}u\, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} \Phi_{s_1}(z - u) f(u)\, \mathrm{d}u.
\end{aligned}$$

Set $\hat{s}_1 = \hat{h}_1 \hat{s}_\xi$ and estimate F^* with

$$\hat{F}^*(z) = \int_{-\infty}^{\infty} \Phi_{\hat{s}_1}(z-u)\,\mathrm{d}\hat{F}(u) = \frac{1}{n}\sum_{t=0}^{n-1} \Phi_{\hat{s}_1}(z-\hat{e}_t) \tag{26}$$

and set $\hat{q}_c^*$ to be the αth quantile of $\hat{F}^*(z)$, the solution to the equation

$$\alpha = \frac{1}{n}\sum_{t=0}^{n-1} \Phi_{\hat{s}_1}(\hat{q}_c^* - \hat{e}_t). \tag{27}$$

Numerically, the solution to (27) is easy to find using Newton iteration. Set the starting value $\hat{q}_c^*(1) = \hat{q}$ and then iterate on the equation

$$\hat{q}_c^*(i+1) = \hat{q}_c^*(i) - \frac{\frac{1}{n}\sum_{t=0}^{n-1}\Phi_{\hat{s}_1}(\hat{q}_c^*(i) - \hat{e}_t) - \alpha}{\frac{1}{n}\sum_{t=0}^{n-1}\phi_{\hat{s}_1}(\hat{q}_c^*(i) - \hat{e}_t)}$$

which will typically converge in an small number of iterations.

Given $\hat{q}_c^*$, we define $\hat{\xi}_c^* = \eta(x, \hat{\beta}, \hat{q}_c^*)$ and define the corrected forecast interval $(-\infty, \hat{\xi}_c^*]$. When $h_2 = 0$ it has the same coverage as the nonparametrically adjusted interval to order $O_p(n^{-3/2})$.

8. Simulation

A small simulation is used to illustrate the accuracy of the proposed forecast intervals. Our forecasts are based on the linear forecasting model (3) with a fixed number of autoregressive lags. This is appropriate for the case of a linear autoregression with iid errors.

The data are generated by the autoregression

$$y_t = 0.8y_{t-1} + e_t.$$

For our first set of experiments, the error e_t is iid from one of the first six Marron and Wand (1992) mixture-normal distributions (Gaussian, skewed, strongly skewed, kurtotic, outlier, and bimodal).

50,000 samples of length $n = 40$, 100, and 200 were generated, and used to generate forecast intervals at the horizons of $k = 2, 6$, and 10. For each sample, a linear forecasting model was estimated with 2, 6, and 10 lags. The number of included autoregressive lags exceeds those in the "true" data generation process in order to assess the impact of model dimension (and associated parameter estimation) upon forecast interval accuracy. Rough and corrected 80% forecast intervals were calculated (10% and 90% forecast quantiles). Realized out-of-sample observations were generated. The intervals and the realizations were compared, with a "success" recorded if the realization was contained in the interval. The frequency of accurate coverage (the percentage of samples for which the realization landed in the interval) is reported in Table 1. Due to the large number of simulation replications,

Table 1
(a) and (b) Nominal 80% forecast interval coverage rates

Sample size	Forecast horizon	Lag order $l=2$				Lag order $l=6$				Lag order $l=10$			
		R	S	C	NP	R	S	C	NP	R	S	C	NP
(a) *Marron– Wand density # 1* (*Gaussian*)													
$n=40$	$k=2$	0.73	0.77	0.77	0.76	0.68	0.74	0.73	0.72	0.61	0.72	0.70	0.68
	$k=6$	0.66	0.71	0.70	0.70	0.60	0.68	0.67	0.66	0.54	0.64	0.62	0.61
	$k=10$	0.64	0.68	0.67	0.67	0.58	0.64	0.63	0.62	0.51	0.59	0.58	0.57
$n=100$	$k=2$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.73	0.77	0.77	0.76
	$k=6$	0.75	0.78	0.77	0.77	0.73	0.76	0.76	0.76	0.71	0.75	0.75	0.74
	$k=10$	0.74	0.76	0.76	0.76	0.71	0.75	0.75	0.74	0.69	0.74	0.73	0.73
$n=200$	$k=2$	0.79	0.80	0.80	0.79	0.78	0.79	0.79	0.79	0.77	0.79	0.79	0.78
	$k=6$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.75	0.78	0.78	0.78
	$k=10$	0.77	0.78	0.78	0.78	0.76	0.78	0.78	0.78	0.75	0.77	0.77	0.77
Marron– Wand density # 2 (*skewed*)													
$n=40$	$k=2$	0.73	0.77	0.77	0.76	0.67	0.74	0.74	0.72	0.61	0.71	0.70	0.68
	$k=6$	0.67	0.71	0.71	0.70	0.60	0.67	0.66	0.65	0.55	0.64	0.63	0.61
	$k=10$	0.64	0.68	0.68	0.67	0.58	0.64	0.63	0.62	0.51	0.59	0.58	0.57
$n=100$	$k=2$	0.77	0.79	0.79	0.78	0.75	0.78	0.78	0.78	0.73	0.77	0.77	0.76
	$k=6$	0.75	0.77	0.78	0.77	0.72	0.76	0.76	0.75	0.71	0.75	0.75	0.74
	$k=10$	0.74	0.77	0.76	0.76	0.72	0.75	0.75	0.75	0.69	0.74	0.73	0.73
$n=200$	$k=2$	0.79	0.80	0.80	0.80	0.78	0.79	0.79	0.79	0.77	0.78	0.78	0.78
	$k=6$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.75	0.78	0.78	0.77
	$k=10$	0.77	0.79	0.78	0.78	0.76	0.78	0.78	0.77	0.75	0.77	0.77	0.77
Marron– Wand density # 3 (*strongly skewed*)													
$n=40$	$k=2$	0.72	0.76	0.77	0.75	0.67	0.74	0.74	0.72	0.61	0.71	0.71	0.69
	$k=6$	0.67	0.72	0.71	0.70	0.61	0.68	0.67	0.66	0.55	0.64	0.62	0.61
	$k=10$	0.64	0.68	0.68	0.67	0.59	0.64	0.63	0.63	0.52	0.60	0.59	0.58
$n=100$	$k=2$	0.77	0.79	0.79	0.78	0.75	0.78	0.78	0.77	0.72	0.76	0.78	0.76
	$k=6$	0.75	0.77	0.77	0.77	0.73	0.76	0.76	0.76	0.70	0.75	0.75	0.74
	$k=10$	0.73	0.76	0.76	0.76	0.71	0.75	0.74	0.74	0.69	0.74	0.73	0.73
$n=200$	$k=2$	0.79	0.79	0.79	0.79	0.77	0.78	0.79	0.78	0.76	0.78	0.79	0.78
	$k=6$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.75	0.78	0.78	0.77
	$k=10$	0.76	0.78	0.78	0.78	0.75	0.78	0.78	0.77	0.75	0.77	0.77	0.77
(b) *Marron– Wand density # 4* (*kurtotic*)													
$n=40$	$k=2$	0.73	0.77	0.77	0.76	0.68	0.74	0.74	0.72	0.62	0.72	0.71	0.69
	$k=6$	0.66	0.71	0.71	0.70	0.60	0.68	0.67	0.65	0.54	0.64	0.63	0.61
	$k=10$	0.64	0.68	0.67	0.67	0.57	0.64	0.63	0.62	0.51	0.59	0.58	0.56
$n=100$	$k=2$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.74	0.77	0.77	0.76
	$k=6$	0.75	0.77	0.77	0.77	0.72	0.76	0.76	0.76	0.70	0.75	0.75	0.74
	$k=10$	0.74	0.76	0.76	0.76	0.71	0.75	0.75	0.74	0.69	0.74	0.74	0.73
$n=200$	$k=2$	0.79	0.80	0.80	0.79	0.78	0.79	0.79	0.79	0.77	0.79	0.79	0.79
	$k=6$	0.77	0.78	0.78	0.78	0.76	0.78	0.78	0.78	0.75	0.78	0.78	0.77
	$k=10$	0.77	0.78	0.78	0.78	0.76	0.78	0.78	0.77	0.75	0.77	0.77	0.77
Marron– Wand density # 5 (*outlier*)													
$n=40$	$k=2$	0.73	0.76	0.80	0.75	0.68	0.74	0.78	0.72	0.62	0.72	0.75	0.68
	$k=6$	0.65	0.70	0.71	0.68	0.60	0.67	0.68	0.64	0.53	0.64	0.64	0.59
	$k=10$	0.61	0.66	0.67	0.64	0.55	0.62	0.62	0.59	0.49	0.58	0.58	0.54
$n=100$	$k=2$	0.77	0.78	0.81	0.79	0.76	0.78	0.81	0.78	0.74	0.77	0.81	0.77
	$k=6$	0.74	0.76	0.78	0.76	0.72	0.75	0.77	0.75	0.71	0.75	0.76	0.74
	$k=10$	0.72	0.75	0.76	0.74	0.71	0.74	0.76	0.74	0.69	0.73	0.74	0.72
$n=200$	$k=2$	0.79	0.80	0.81	0.80	0.78	0.79	0.81	0.80	0.77	0.78	0.81	0.79
	$k=6$	0.77	0.78	0.79	0.78	0.76	0.78	0.79	0.78	0.76	0.78	0.79	0.78
	$k=10$	0.76	0.77	0.78	0.77	0.75	0.77	0.78	0.77	0.75	0.77	0.78	0.77

Table 1 (*continued*)

Sample size	Forecast horizon	Lag order $l=2$				Lag order $l=6$				Lag order $l=10$			
		R	S	C	NP	R	S	C	NP	R	S	C	NP
Marron– Wand density # 6 (bimodal)													
$n=40$	$k=2$	0.72	0.77	0.76	0.75	0.67	0.74	0.73	0.72	0.61	0.71	0.70	0.68
	$k=6$	0.66	0.72	0.71	0.70	0.60	0.68	0.67	0.66	0.54	0.64	0.62	0.61
	$k=10$	0.64	0.68	0.67	0.67	0.58	0.64	0.63	0.63	0.51	0.60	0.58	0.57
$n=100$	$k=2$	0.77	0.79	0.79	0.79	0.75	0.78	0.77	0.77	0.73	0.77	0.76	0.76
	$k=6$	0.75	0.77	0.77	0.77	0.73	0.76	0.76	0.76	0.70	0.75	0.75	0.74
	$k=10$	0.73	0.76	0.76	0.76	0.71	0.75	0.75	0.74	0.69	0.74	0.74	0.73
$n=200$	$k=2$	0.78	0.79	0.79	0.79	0.78	0.79	0.79	0.79	0.76	0.78	0.78	0.78
	$k=6$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.75	0.78	0.77	0.77
	$k=10$	0.77	0.78	0.78	0.78	0.76	0.78	0.78	0.78	0.74	0.77	0.77	0.77

these estimates have standard errors of only 0.002, so the reported frequencies are precise.

The cells in Table 1 report the frequency with which the forecast intervals contained the out-of-sample realization. Ideally, they should have 80% coverage. Deviations from 0.80 indicate inaccurate coverage.

As expected, the accuracy of the forecast intervals increases with sample size, decreases with the forecast horizon, and decreases with the model order. In particular, when $n=40$ the accuracy is quite low unless $k=2$. In general, the forecast intervals have coverage less than the desired 80%. This appears to be due to the large bias in autoregressive estimation in small samples with strong serial correlation.

In all cases our three corrected intervals always perform better than the rough intervals. While the coverage probabilities of the three corrected intervals are fairly similar, the simple reference adjustment nearly always has the best coverage, with the convolution adjustment second best. An exception to this rule arises for the Outlier error distribution, where we see that the convolution adjustment yields the best coverage.

For our second set of experiments, we introduce two forms of misspecification. First, we generated e_t from the MA(1)

$$e_t = \varepsilon_t + 0.8\varepsilon_{t-1}$$

with ε_t iid $N(0,1)$. Equivalently, y_t is an infinite-order autoregression so all finite-order autoregressions are misspecified and have omitted dynamics. The results from application of the same procedures are reported in the first panel of Table 2. The performance are very similar to the simple AR(1) case. This form of misspecification does not have a meaningful effect on coverage accuracy.

Finally, we generated e_t from the ARCH(1) model

$$e_t \sim N(0, 0.2 + 0.8e_{t-1}^2).$$

Conditional heteroskedasticity violates the assumption of iid errors, and is thus a form of misspecification. The results are presented in the second panel of Table 2.

Table 2
Nominal 80% forecast interval coverage rates misspecified cases

Sample size	Forecast horizon	Lag order $l = 2$				Lag order $l = 6$				Lag order $l = 10$			
		R	S	C	NP	R	S	C	NP	R	S	C	NP
MA(1)													
$n = 40$	$k = 2$	0.71	0.76	0.76	0.75	0.66	0.73	0.72	0.72	0.60	0.70	0.69	0.68
	$k = 6$	0.64	0.70	0.69	0.68	0.58	0.66	0.65	0.64	0.51	0.62	0.60	0.59
	$k = 10$	0.61	0.66	0.65	0.65	0.55	0.62	0.61	0.60	0.48	0.57	0.55	0.54
$n = 100$	$k = 2$	0.77	0.79	0.79	0.78	0.75	0.78	0.78	0.77	0.73	0.77	0.77	0.77
	$k = 6$	0.74	0.77	0.77	0.77	0.72	0.76	0.76	0.75	0.70	0.75	0.74	0.74
	$k = 10$	0.72	0.76	0.76	0.75	0.70	0.74	0.74	0.74	0.68	0.73	0.73	0.72
$n = 200$	$k = 2$	0.79	0.79	0.79	0.79	0.77	0.79	0.79	0.79	0.77	0.79	0.79	0.78
	$k = 6$	0.77	0.79	0.79	0.79	0.76	0.78	0.78	0.78	0.75	0.78	0.78	0.77
	$k = 10$	0.76	0.78	0.78	0.78	0.75	0.77	0.77	0.77	0.74	0.77	0.77	0.77
ARCH(1)													
$n = 40$	$k = 2$	0.58	0.63	0.64	0.62	0.53	0.62	0.62	0.60	0.49	0.60	0.60	0.57
	$k = 6$	0.55	0.61	0.61	0.60	0.50	0.59	0.58	0.57	0.45	0.57	0.55	0.54
	$k = 10$	0.54	0.60	0.59	0.59	0.49	0.57	0.56	0.55	0.43	0.54	0.52	0.51
$n = 100$	$k = 2$	0.59	0.62	0.63	0.62	0.58	0.62	0.63	0.62	0.57	0.61	0.63	0.62
	$k = 6$	0.61	0.64	0.65	0.64	0.59	0.63	0.64	0.63	0.57	0.63	0.63	0.63
	$k = 10$	0.61	0.64	0.65	0.64	0.59	0.63	0.64	0.63	0.57	0.63	0.64	0.63
$n = 200$	$k = 2$	0.60	0.61	0.62	0.62	0.59	0.61	0.62	0.62	0.59	0.61	0.62	0.62
	$k = 6$	0.62	0.65	0.64	0.65	0.61	0.63	0.64	0.64	0.61	0.63	0.64	0.62
	$k = 10$	0.62	0.64	0.65	0.65	0.61	0.64	0.65	0.65	0.60	0.64	0.65	0.64

R = rough forecast interval; S = simple reference adjustment;
C = convolution adjustment; NP = nonparametric adjustment.

We observe that the coverage probabilities dramatically fall, in all cases. The coverage rates do not even improve greatly with larger sample sizes, as the forecasting model mis-specifies the conditional variance. While the corrected intervals have slightly better coverage rates than the rough intervals, our assessment is that this experiment highlights the critical role played by the independence assumption.

In summary, one message from the simulation is that any of the corrections can provide a useful improvement over the standard forecast intervals. A surprising finding is that the simple reference adjustment (22) is generally the best performer. This is an easy adjustment, only requiring the long-run variance estimation (12), and is thus recommended for empirical practice.

A second message is that while the corrections make an improvement, the magnitude of the improvement is often small compared with the size of the original discrepancy. Methods which could yield further improvements would be quite desirable.

9. Unemployment rate forecasts

We examine a common forecasting target, the US monthly unemployment rate. Our measure is the percentage of unemployed in the civilian labor force, which has

Table 3
Empirical coverage rates US unemployment rate nominal 80% forecast intervals

	Rough intervals				Simple reference intervals			
	$n = 60$	$n = 120$	$n = 180$	$n = 240$	$n = 60$	$n = 120$	$n = 180$	$n = 240$
$k = 1$	0.73	0.80	0.83	0.83	0.77	0.81	0.83	0.84
$k = 3$	0.63	0.75	0.81	0.83	0.71	0.79	0.84	0.85
$k = 6$	0.54	0.72	0.79	0.83	0.65	0.79	0.82	0.85
$k = 9$	0.39	0.67	0.74	0.80	0.51	0.75	0.78	0.82
$k = 12$	0.32	0.59	0.70	0.78	0.42	0.67	0.77	0.81

an available sample from 1948:01 through 2004:07. We use the simple linear forecasting model (3), with the number of lags determined by minimizing the AIC (Calculated separately for each sample/horizon combination).

First, we calculate a series of 80% pseudo out-of-sample forecast intervals (10% and 90% forecast quantiles) using rolling samples. We vary the sample size n among 60, 120, 180, and 240, and the forecast horizon k among 1, 3, 6, 9, and 12. For example, for the $n = 240$, $k = 12$ combination, we start with the observations from 1948:01 to 1957:12 to generate a forecast interval for 1958:12. We then roll forward one month, using the sample from 1948:02 to 1958:1 to generate a forecast interval for 1959:1. We continue in this fashion, one month at a time, until a forecast interval for the final observation 2004:01 is generated. We then calculated the frequency, across forecasts, in which the forecast interval contained the actual unemployment rate. These empirical coverage probabilities are reported in Table 3 for the rough and simple reference intervals.

In all cases, the simple reference forecast intervals have higher coverage rates than the rough forecast intervals. In most cases the rough intervals have insufficient coverage and the corrected intervals have coverage rates closer to the desired 80%. In a few cases (large sample sizes and small forecast horizon) the rough intervals have approximately correct coverage rates and the corrected intervals are somewhat overcovering. For the small sample size of $n = 60$, even the corrected intervals have insufficient coverage except for one-step-ahead forecasts ($k = 1$). For $n = 180$ their coverage rates are good at all horizons. The coverage rates were also calculated for the convolution and nonparametric intervals and their performance was similar to the simple reference intervals so are not reported.

In Fig. 1 we plot the forecast quantiles for the case of $n = 180$ and $k = 3$. We plot the 10% and 90% simple reference forecast quantiles, as well as the actual series.

Fig. 2 plots a fan chart, displaying out-of-sample simple reference forecast intervals for August 2004 through January 2006. It is calculated using the most recent 180 observations. The lines represent the forecast deciles. The forecast intervals are asymmetric with a thick upper tail. The forecast distribution fans out as the horizon increases, reflecting uncertainty, while the median drifts downwards over the first 12 months and then slightly turns upwards.

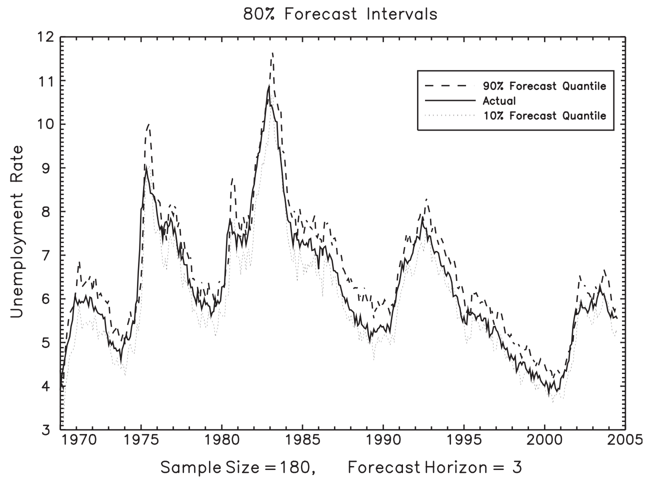

Fig. 1. 80% forecast intervals.

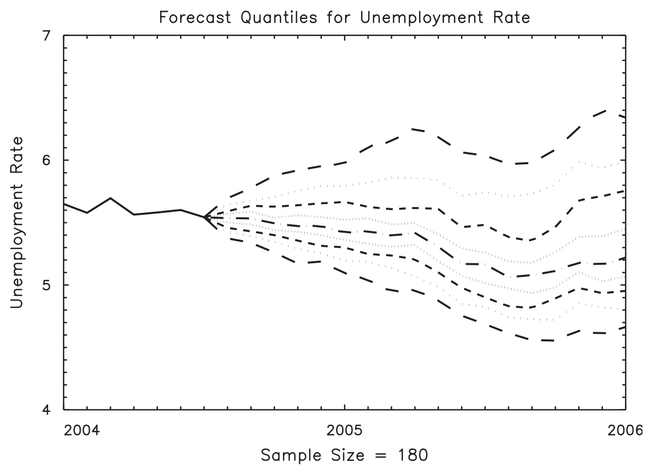

Fig. 2. Forecast quantiles for unemployment rate.

10. Extensions

10.1. Multiple-horizon multi-step forecasts

Our theoretical analysis has been confined to the case of a single model used for forecasting at a specific horizon. Yet in many contexts, including the construction of fan charts, the actual goal is to produce forecasts at multiple horizons. The methods described in this paper can be used to construct such forecast intervals, horizon by horizon, but we have not examined the properties of such joint forecasts.

Furthermore, how should we simultaneously construct forecasting models for multiple horizons? In this paper we have confined attention to direct estimation techniques, which require the construction of an explicit forecasting equation for each separate horizon. When multiple horizions are simultaneously considered, how should these models be related? This issue is not explored here, and the answer is uncertain.

Finally, we have confined attention to forecasts computed from direct estimation. An alternative is one-step plug-in forecasting. It would be useful to extend our analysis to this setting. However, while plug-in point forecasts are straightforward to generate, interval forecasts are quite difficult and would require more sophisticated calculations.

10.2. Bootstrap methods

The methods presented here are asymptotic. It would be fruitful to explore alternative distributional approximations such as the bootstrap. The simulation results of Section 8 suggest that there is considerable room for improvement in small samples. However, it is unclear how to use the bootstrap to construct accurate forecast intervals. There are several issues which require careful treatment, including the following.

First, forecasts are conditioned on the final values of the time series. Thus bootstrap samples must have the same final values. Construction of such bootstrap samples is tricky. Standard time-series bootstrap methods (e.g. the block bootstrap) do not satisfy this requirement.

Second, since the forecast error is non-Gaussian, conventional studentization techniques cannot be used as required for good bootstrap performance. Percentile methods may need to be employed.

10.3. Estimation bias

Our approximations utilize the asymptotic approximation that parameter estimates are approximately unbiased in large samples. Our methods and adjustments focus on estimation variance exclusively. However, it is well-known that autoregressive estimates are highly biased in moderate sample sizes. It is quite reasonable to expect that this is an important factor in the spotty performance of our forecast intervals. Investigation of the impact of parameter estimation bias on forecast interval accuracy may be a fruitful avenue for research. Bootstrap methods may be necessary to estimate and correct for this bias.

10.4. Smooth distribution estimators

This paper follows the conventional literature and estimates the distribution function of the equation error using the discrete empirical distribution function. Alternatively, a kernel estimator could be used, resulting in a smooth distribution estimate. See Hansen (2004) for a description of such methods for univariate data.

It should be possible to modify the methods discussed here to allow for a smooth distribution estimator, but that is left for future research.

Acknowledgements

Thanks to Ken West, Bill Brown, Norman Swanson, and three referees for helpful suggestions and comments. Research supported by the National Science Foundation.

Appendix

The proof of Theorem 1 is facilitated by a technical lemma. Define

$$m_t(\theta) = \begin{pmatrix} l_t(\beta) \\ 1(y_{t+k} \leqslant \eta(x_t, \theta)) - \alpha \end{pmatrix},$$

$\overline{m}_n(\theta) = \frac{1}{n}\sum_{t=0}^{n-1} m_t(\theta)$, and $m(\theta) = \mathrm{E}m_t(\theta)$. Let $\Theta \subset R^{d+1}$ be compact.

Lemma 1. $\sqrt{n}(\overline{m}_n(\theta) - m(\theta)) \Rightarrow S(\theta)$, *a Gaussian process over* $\theta \in \Theta$.

Proof of Lemma 1. By Doukhan et al. (1995) Theorem 1, Application 1, under our Assumption 1 it is sufficient to show $\int_0^1 H(u)^{1/2}\,\mathrm{d}u < \infty$, where $H(u)$ denotes the entropy with bracketing with respect to the L^{2r} norm. Andrews (1994) Theorem 5 shows that the latter is satisfied if for all θ

$$\mathrm{E} \sup_{\theta_1 : |\theta - \theta_1| < \delta} |m_t(\theta) - m_t(\theta_1)|^{2r} \leqslant C\delta. \tag{28}$$

for some $C < \infty$. We can ignore the $l_t(\beta)$ component of $m_t(\theta)$ as trivial under Assumption 1.6.

Under Assumption 1.7, for all $|\theta - \theta_1| < \delta$

$$\begin{aligned} |m_t(\theta) - m_t(\theta_1)|^{2r} &= |1(y_{t+k} \leqslant \eta(x_t, \theta)) - 1(y_{t+k} \leqslant \eta(x_t, \theta_1))|^{2r} \\ &\leqslant 1(|y_{t+k} - \eta(x_t, \theta)| \leqslant a(x_t)\delta). \end{aligned}$$

Thus,

$$\begin{aligned} \mathrm{E} \sup_{\theta_1 : |\theta - \theta_1| < \delta} |m_t(\theta) - m_t(\theta_1)|^{2r} &\leqslant \mathrm{P}(|y_{t+k} - \eta(x_t, \theta)| \leqslant a(x_t)\delta) \\ &= \mathrm{E}G(\eta(x_t, \theta) + a(x_t)\delta | x_t) - G(\eta(x_t, \theta) - a(x_t)\delta | x_t)] \\ &\leqslant 2\overline{g}\mathrm{E}a(x_t)\delta \end{aligned}$$

using the bound $g(y|x) \leqslant \overline{g}$.

Proof of Theorem 1. Assumption 1 is a standard set of conditions for $\hat{\theta} \rightarrow_p \theta_0$. By a Taylor expansion,

$$0 = m(\theta_0) = m(\hat{\theta}) + Q(\theta_0 - \hat{\theta}) + \mathrm{o}_p(n^{-1/2}), \tag{29}$$

where

$$Q = \frac{\partial}{\partial \theta'} \mathrm{E}m(\theta_0).$$

Set

$$\eta_\theta(x) = \begin{pmatrix} \eta_\beta(x) \\ \eta_q(x) \end{pmatrix}.$$

By (7), a second Taylor expansion, (29), the fact that $\sqrt{n}\overline{m}_n(\hat{\theta}) = \mathrm{o}_p(1)$, Lemma 1 and the consistency of $\hat{\theta}$ for θ, and the central limit theorem,

$$\begin{aligned}
\sqrt{n}(\hat{\xi} - \xi) &= \sqrt{n}(\eta(x, \hat{\theta}) - \eta(x, \theta_0)) \\
&= \eta_\theta(x)' \sqrt{n}(\hat{\theta} - \theta_0) + \mathrm{o}_p(1) \\
&= \eta_\theta(x)' Q^{-1} \sqrt{n} m(\hat{\theta})) + \mathrm{o}_p(1) \\
&= \eta_\theta(x)' Q^{-1} \sqrt{n}(m(\hat{\theta}) - \overline{m}_n(\hat{\theta})) + \mathrm{o}_p(1) \\
&= \eta_\theta(x)' Q^{-1} \sqrt{n}(m(\theta_0) - \overline{m}_n(\theta_0)) + \mathrm{o}_p(1) \\
&\xrightarrow{d} \mathrm{N}(0, \sigma_\xi^2),
\end{aligned}$$

where σ_ξ^2 is defined in (9) with $u_t = \eta_\theta(x)' Q^{-1} m_t(\theta_0)$. It remains to simplify u_t.

First, observe that

$$\begin{aligned}
\frac{\partial}{\partial q} \mathrm{E}1(y_{t+k} \leqslant \eta(x_t, \theta))\Big|_{\theta=\theta_0} &= \frac{\partial}{\partial q} \mathrm{E}1(h(y_{t+k}, x_t, \beta) \leqslant q)\Big|_{\theta=\theta_0} \\
&= \frac{\partial}{\partial q} \mathrm{E}1(e_t \leqslant q)\Big|_{q=q_0} \\
&= f(q).
\end{aligned}$$

Second, since $G(y|x) = F(h(y, x, \beta))$,

$$G(y|x) = \frac{\partial}{\partial y} G(y|x) = f(h(y, x, \beta)) \frac{\partial}{\partial y} h(y, x, \beta).$$

Thus,

$$\begin{aligned}
\frac{\partial}{\partial \beta} \mathrm{E}1(y_{t+k} \leqslant \eta(x_t, \theta))\Big|_{\theta=\theta_0} &= \mathrm{E}\frac{\partial}{\partial \beta} G(\eta(x_t, \theta)|x_t)\Big|_{\theta=\theta_0} \\
&= \mathrm{E}f(h(\eta(x_t, \theta), x_t, \beta)) \frac{\partial}{\partial y} h(\eta(x_t, \theta), x_t, \beta) \frac{\partial}{\partial \beta} \eta(x_t, \theta)\Big|_{\theta=\theta_0} \\
&= \mathrm{E}\left(f(q_0) \frac{\partial}{\partial y} h(\eta(x_t, \theta), x_t, \beta_0) \eta_\beta(x_t) \right) \\
&= f(q_0)\gamma.
\end{aligned}$$

Hence

$$Q = \left(\begin{array}{cc} \dfrac{\partial}{\partial \beta'} \mathrm{E} l_t(\beta_0) & 0 \\ \dfrac{\partial}{\partial \beta'} \mathrm{E}1(y_{t+k} \leqslant \eta(x_t, \theta)) & \dfrac{\partial}{\partial q} \mathrm{E}1(y_{t+k} \leqslant \eta(x_t, \theta)) \end{array} \right) \Bigg|_{\theta=\theta_0} = \begin{pmatrix} l_\beta & 0 \\ f(q_0)\gamma' & f(q_0) \end{pmatrix}$$

and thus

$$Q^{-1} = \begin{pmatrix} l_\beta^{-1} & 0 \\ -\gamma' l_\beta^{-1} & f(q_0)^{-1} \end{pmatrix}.$$

Together,

$$\begin{aligned} u_t &= (\, \eta_\beta(x)' \quad \eta_q(x)\,) \begin{pmatrix} l_\beta^{-1} & 0 \\ -\gamma' l_\beta^{-1} & f(q_0)^{-1} \end{pmatrix} \begin{pmatrix} l_t \\ 1(e_t \leqslant q_0) - \alpha \end{pmatrix} \\ &= (\eta_\beta(x)' - \eta_q(x)\gamma') l_\beta^{-1} l_t + \eta_q(x) \left(\frac{1\left(e_t \leqslant q_0\right) - \alpha}{f(q_0)} \right). \qquad \square \end{aligned}$$

Proof of Theorem 2. By sequential conditioning

$$\begin{aligned} \mathrm{P}(y_{n+k} \leqslant \hat{\xi} | \hat{\xi}, x_n = x) &= \mathrm{P}(\eta(x, \beta_0, e_n) \leqslant \hat{\xi} | \hat{\xi}, x_n = x) \\ &= \mathrm{P}(e_n \leqslant h(\hat{\xi}, x, \beta_0) | \hat{\xi}, x_n = x) \\ &= F(h(\hat{\xi}, x, \beta_0)) \\ &= F(q_0) + f(q_0) h_1 (\hat{\xi} - \xi_0) \\ &\quad + \tfrac{1}{2}[f(q_0) h_2 + f'(q_0) h_1^2](\hat{\xi} - \xi_0)^2 \\ &\quad + \mathrm{O}(\hat{\xi} - \xi_0)^3, \end{aligned}$$

where the final equality is a Taylor expansion of $\hat{\xi}$ about ξ_0 and we have used $q_0 = h(\xi_0, x, \beta_0)$. Taking expectations conditional on $x_n = x$ alone

$$\begin{aligned} \mathrm{P}(y_{n+k} \leqslant \hat{\xi} | x_n = x) &= \mathrm{E}[\mathrm{P}(y_{n+k} \leqslant \hat{\xi} | \hat{\xi}, x_n = x) | x_n = x] \\ &= F(q_0) + f(q_0) h_1 \mathrm{E}(\hat{\xi} - \xi) + \tfrac{1}{2}(f(q_0) h_2 + f'(q_0) h_1^2) \mathrm{E}(\hat{\xi} - \xi_0)^2 \\ &\quad + \mathrm{O}(n^{-3/2}) = F(q_0) + \tfrac{1}{2}(f(q_0) h_2 \\ &\quad + f'(q_0) h_1^2) s_{n\xi}^2 + \mathrm{O}(n^{-3/2}), \end{aligned}$$

where the final equality uses the asymptotic distribution for $\hat{\xi}$. $\square$

Proof of Theorem 3. First, using (24),

$$\mathrm{E}(\hat{q}^*_{np} - q_0) \approx -\frac{1}{2}\left(h_2 + h_1^2 \frac{f'(q_0)}{f(q_0)}\right) s^2_{n\xi}$$

and thus,

$$\begin{aligned}\mathrm{E}(\hat{\xi}^*_{np} - \xi_0) &= \mathrm{E}(\eta(x, \hat{\beta}, \hat{q}^*_{np}) - \xi_0) \\ &\approx \frac{\partial}{\partial q}\eta(x, \beta, q_0)\mathrm{E}(\hat{q}^*_{np} - q_0) \\ &\approx -\frac{1}{h_1}\frac{1}{2}\left(h_2 + h_1^2 \frac{f'(q_0)}{f(q_0)}\right) s^2_{n\xi},\end{aligned}$$

since $(\partial/\partial q)\,\eta(x, \beta, q_0) = h_1^{-1}$ by the implicit function theorem.

By an expansion,

$$\begin{aligned}\mathrm{P}\left(y_{n+k} \leqslant \hat{\xi}^*_{np} | \hat{\xi}^*_{np}, x_n = x\right) &= \mathrm{P}(e_n \leqslant h(\hat{\xi}^*_{np}, x, \beta_0) | \hat{\xi}^*_{np}, x_n = x) \\ &= F(h(\hat{\xi}^*_{np}, x, \beta_0)) \\ &= F(q_0) + f(q_0)h_1(\hat{\xi}^*_{np} - \xi_0) + \tfrac{1}{2}[f(q_0)h_2 \\ &\quad + f'(q_0)h_1^2](\hat{\xi}^*_{np} - \xi_0)^2 + \mathrm{O}_p(n^{-3/2}).\end{aligned}$$

Taking expectations conditional on $x_n = x$,

$$\begin{aligned}\mathrm{P}(y_{n+k} \leqslant \hat{\xi}^*_{np} | x_n = x) &= \alpha + f(q_0)h_1\left[-\frac{1}{h_1}\frac{1}{2}\left(h_2 + h_1^2 \frac{f'(q_0)}{f(q_0)}\right) s^2_{n\xi}\right] \\ &\quad + \tfrac{1}{2}[f(q_0)h_2 + f'(q_0)h_1^2] s^2_{n\xi} + \mathrm{O}_p(n^{-3/2}) \\ &= \alpha + \mathrm{O}_p(n^{-3/2})\end{aligned}$$

as stated. □

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D.W.K., 1994. Empirical process methods in econometrics. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV, Elsevier Science, Amsterdam, pp. 2248–2296.

Bean, C., Jenkinson, N., 2001. The formulation of monetary policy at the Bank of England. Quarterly Bulletinm, Bank of England, 434–441.

Britton, E., Fisher, P., Whitley, J., 1998. The Inflation Report projections: understanding the fan chart. Quarterly Bulletin, Bank of England, 30–37.

Chatfield, C., 1993. Calculating interval forecasts. Journal of Business and Economic Statistics 11, 121–135.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39, 840–841.

Cogley, T., Morozov, S., Sargent, T.J., 2003. Bayesian fan charts for UK inflation: forecasting and sources of uncertainty in an evolving monetary system, working paper.

Corradi, V., Swanson, N.R., 2004. Predictive Density Accuracy Tests. Rutgers University, working paper.
Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.
Diebold, F.X., Hahn, J., Tay, A.S., 1999a. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. The Review of Economics and Statistics 81, 661–673.
Diebold, F.X., Tay, A.S., Wallis, K.F., 1999b. Evaluating density forecasts of inflation: the survey of professional forecasters. In: Granger, W.J., Engle, R., White, H. (Eds.), Cointegration, Causality, and Forecasting: a Festschrift in Honour of Clive. Oxford University Press, Oxford, pp. 76–90.
Doukhan, P., Massart, P., Rio, E., 1995. Invariance principles for absolutely regular empirical processes. Annales de l'Institut H. Poincare, Probabiltes et Statistiques 31, 393–427.
Fan, J., Qiwei, Y., 2003. Nonlinear Time Series: nonparametric and Parametric Methods. Springer, New York.
Granger, C.W.J., 1996. Can we improve the perceived quality of economic forecasts? Journal of Applied Econometrics 11, 455–473.
Granger, C.W.J., Newbold, P., 1986. Forecasting Economic Time Series, second ed. Academic Press, New York.
Granger, C.W.J., Pesaran, M.H., 1999. A decision theoretic approach to forecast evaluation. In: Chan, W.S., Li, W.K., Tong, H. (Eds.), Statistics and Finance: An Interface. Imperial College Press.
Granger, C.W.J., White, H., Kamstra, M., 1989. Interval forecasting: an analysis based upon ARCH-quantile estimators. Journal of Econometrics 40, 87–96.
Hansen, B.E., 2004. Bandwidth Selection for Nonparametric Distribution Estimation. University of Wisconsin, manuscript.
Hirukawa, M., 2005. A two-stage plug-in bandwidth selection and its implementation for covariance estimation. Concordia University, manuscript.
Ing, C.K., 2003. Multistep prediction in autoregressive processes. Econometric Theory 19, 254–279.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Annals of Statistics 20, 712–736.
Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.
Otrok, C., Whiteman, C.H., 1997. What to do when the crystal ball is cloudy: conditional and unconditional forecasting in Iowa. Proceedings of the National Tax Association. pp. 326–334.
Otrok, C., Whiteman, C.H., 1998. Bayesian leading indicators: measuring and predicting economic conditions in Iowa. International Economic Review 39, 997–1014.
Robertson, J.C., Tallman, E.W., Whiteman, C.H., 2003. Forecasting using relative entropy, Federal Reserve Bank of Atlanta, Working paper no. 2002–2022.
Stock, J.H., Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293–335.
Stock, J.H., Watson, M.W., 2003. Combination forecasts of output growth in a seven-country data set. Harvard University.
Tay, A.S., Wallis, K.F., 2000. Density forecasting: a survey. Journal of Forecasting 19, 235–254.
Tiao, G.C., Tsay, R.S., 1994. Some advances in nonlinear and adaptive modeling in time series. Journal of Forecasting 13, 109–131.
Wallis, K.F., 2003a. Chi-square tests of interval and density forecasts, and the Bank of England's Fan Charts. International Journal of Forecasting 19, 165–175.
Wallis, K.F., 2003b. Forecast uncertainty, its representation and evaluation. Boletin Inflacion y Analisis Macroeconomico. pp. 89–98.
West, K.D., 1997. Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator. Journal of Econometrics 76, 171–191.
Zarnowitz, V., Braun, P., 1993. Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In: Stock, J.H., Watson, M.W. (Eds.), Business Cycles, Indicators, and Forecasting. University of Chicago Press, Chicago.