Empirical Simultaneous Prediction Regions for Path-Forecasts

Òscar Jordá
Federal Reserve Bank of San Francisco
and University of California, Davis

Malte Knuppel
Deutsche Bundesbank

Massimiliano Marcellino
European University Institute

May 2012

Empirical Simultaneous Prediction Regions for Path-Forecasts[*]

Abstract

This paper investigates the problem of constructing prediction regions for forecast trajectories 1 to H periods into the future – a path forecast. We take the more general view that the null model is only approximative and in some cases it may be altogether unavailable. As a consequence, one cannot derive the usual analytic expressions nor resample from the null model as is usually done when bootstrap methods are used. The paper derives methods to construct approximate rectangular regions for simultaneous probability coverage which correct for serial correlation. The techniques appear to work well in simulations and in an application to the Greenbook path-forecasts of growth and inflation.

JEL Classification Codes: C32, C52, C53

Keywords: path-forecast, forecast uncertainty, simultaneous prediction region, Scheffé's S-method, Mahalanobis distance.

Òscar Jordà
  Economic Research, MS 1130
  Federal Reserve Bank of San Francisco
  101 Market St.
  San Francisco, CA 94105
  e-mail: oscar.jorda@sf.frb.org

Department of Economics
University of California, Davis
One Shields Ave.
Davis, CA 95616
e-mail: ojorda@ucdavis.edu

Malte Knüppel
Deutsche Bundesbank
Wilhelm-Epstein-Str. 14
D-60431 Frankfurt am Main, Germany
e-mail: malte.knueppel@bundesbank.de

Massimiliano Marcellino
Department of Economics
European University Institute
Via della Piazzuola 43
50133 Firenze, Italy
e-mail: massimiliano.marcellino@eui.eu

---

1 Introduction

What is the proper way to report the uncertainty associated with the forecast of a variable over several periods – a path-forecast in Jordà and Marcellino (2010) parlance? This is a deceptively simple question yet intricate to answer. The region ensuring that each period a given proportion of realizations is contained within will not, in general, contain that proportion of the path-forecast's realizations. Conversely, the region designed to simultaneously contain a proportion of all future trajectories will, in general, lack the desired per-period coverage. Part of this tension resides in making multiple probability statements, another part is explained by the inherent serial correlation in the forecast errors of typical time-series prediction models, and yet another part is due to differences in the prediction error rate one may wish to control for. These are some of the central themes in our paper.

Imagine a central bank considering the optimal path for the policy rate based on inflation forecasts for the next eight quarters. The implicit control problem underlying the policy choices likely emphasizes the overall trajectory of inflation relative to a pre-determined target over an occasional oscillation in prices. The former is more likely to form the basis for a revision in policy than the latter. And in that case, simultaneous prediction error control may be preferable to family-wise prediction error control. A good discussion on optimal policy projections and their role for monetary policy is available in Svensson (2009) and references therein. Moreover, many central banks now release inflation path-forecasts and policy path-projections such as the Swedish Riksbank, the Bank of England or more recently, the U.S. Federal Reserve, to cite a few.

Socioeconomic models lack the precision and stability of models of physical phenomena

so that their approximative nature makes them more vulnerable to traditional assumptions commonly used to derive closed-form expressions. In the extreme, the decision-maker may in fact be ignorant of the model used to generate the forecasts themselves. This is often the case when the forecasts are not produced under the direct supervision of the decision-maker. For example, several forecasting firms (Blue Chip Forecasting, Macroeconomic Advisors, Standard and Poor, etc.) release U.S. inflation forecasts but not sufficient details about the forecasting null model. How should prediction intervals be calculated when the forecasting model may be misspecified, or altogether unavailable?

Providing appropriate prediction intervals for path-forecasts in these conditions is complicated. Prediction bands for the path-forecast should appropriately summarize the information in high-dimensional densities. For these reasons, we investigate a simple method to construct empirical simultaneous prediction regions based on the methods by Williams and Goodman (1971), whose ideas precede but are related to the resampling methods in Efron's (1979) well-known bootstrap procedures, and subsequent literature (see, e.g. Politis, Romano and Wolf 1999). In particular, the methods we propose are based on the evaluation sample of forecast errors at different horizons. Given typical predictive sample sizes, non-parametric estimates of the joint forecast error distribution and its relevant quantile contours (which one would need to construct non-parametric simultaneous prediction regions) are impractical. Instead, we begin by constructing the empirical Mahalanobis (1936) distance from which we induce a simultaneous prediction region. Critical values can then be directly obtained from the evaluation sample or using distributional assumptions. Next, we provide a rectangular approximation to the induced simultaneous prediction region to obtain appropriate,

2

per-period, prediction intervals using a version of Scheffé (1953) projections.

The procedures proposed rely on the assumption that the empirical distribution of the forecast errors remains constant over time. While this may appear restrictive, it is a weaker assumption than making conjectures about the data generating process and hence its stability over time – that is, a changing DGP approximated by an appropriately changing null model may nevertheless produce stable forecast errors. Such a way of thinking differs from common applications of bootstrap procedures such as forecasting applications with ARIMA models in Masarotto (1990), Thombs and Schucany (1990), Kim (1999), Pascual, Romo and Ruiz (2000), and Clements and Taylor (2001). This literature typically relies on residual resampling given the null model. In contrast, we make no assumption about whether the null model is or is not correct.

Using our assumptions on the null forecasting model means that typical asymptotic arguments are unavailable. Even with infinite evaluation samples, large sample approximations speak of the approximate distribution of the moments of the empirical forecast error distribution, but not what that distribution may look like. Instead, we provide ample simulation evidence about how our procedures may work in practice. In addition, we provide a realistic application of our procedures by assessing the growth and inflation forecasts generated by the U.S. Federal Reserve, whose null model is not revealed to the public.

The simulations show that the use of prediction intervals based on forecast errors is strongly advisable if the null model is misspecified, particularly if the simultaneous probability coverage of the intervals is to be evaluated. The effects of incorrect distributional assumptions turn out to be relatively well contained, at least in the settings considered in

this paper. As further illustration, we construct prediction regions using Greenbook[1] fore-
casts produced by the staff of the U.S. Federal Reserve. The model used to generate the
forecasts is not made available to the public. Nevertheless, we are able to compute empirical
prediction regions that quantify practical measures of uncertainty surrounding these fore-
casts. Such measures are important to evaluate forecasting uncertainty in general, and to
policymakers in particular when entertaining forecasts from other sources.

2 Statistical Discussion

The typical forecasting environment often contains the following elements: (1) a training
sample $\{y_t\}_{t=1}^R$ and an evaluation sample $\{y_\tau\}_{\tau=R+1}^T$; (2) an assumption about the stochastic
process for y_t; (3) model estimates based on this assumption using the training sample; (4)
an implicit or explicit forecast loss-function whose associated optimal forecast often is an
estimate of the conditional mean of y_t based on the assumed model and its estimates; and
(5) ancillary attributes that derive from the interaction between the assumed model and the
properties of y_t. These attributes determine forecast bias, accuracy, construction of prediction
intervals, and so on.

Instead, we proceed from a more agnostic but more general starting point that will allow
us to deal, for example, with forecasts whose provenance is unknown. Specifically, let $y_\tau(h)$
be a random variable and let $\widehat{y}_\tau(h)$ denote a realization of this random variable – an h-step
ahead prediction of $y_{\tau+h}$. The observed realization of $y_{\tau+h}$ is denoted as $\widehat{y}_{\tau+h}$. Hence the
forecast error $\widehat{u}_\tau^h \equiv \widehat{y}_{\tau+h} - \widehat{y}_\tau(h)$ is a realization of the random variable $u_\tau^h \equiv y_{\tau+h} - y_\tau(h)$,
whose distribution is assumed to have mean zero (i.e., without loss of generality we assume

---

[1] Greenbook forecasts are now refered to as Tealbook forecasts.

that the forecasts are unbiased) with constant variance σ_h^2 and independent of τ. Further, we will assume that the u_τ^h are weakly stationary. Notice that this assumption is less restrictive than making any stationarity assumptions directly on $y_\tau(h)$ or $y_{\tau+h}$.

In what follows, we are interested in the 1-to-H step-ahead predictions with $\tau = R + 1, ..., T - H$ observations available for forecast evaluation with $N = T - H - R$. We also find it convenient to define the $H \times 1$ random vectors $Y_\tau(H) \equiv [y_\tau(1), ..., y_\tau(H)]'$; $Y_{\tau,H} \equiv [y_{\tau+1}, ..., y_{\tau+H}]'$ and $U_\tau(H) \equiv [u_\tau^1, ..., u_\tau^H]'$ with $\widehat{Y}_\tau(H)$, $\widehat{Y}_{\tau,H}$ and $\widehat{U}_\tau(H)$ denoting the actual forecasts, realizations and forecast errors, respectively. We assume that the distribution of the forecast error path $U_\tau(H)$ is elliptically contoured, that is, it has the form:

$$f_U(U_\tau(H)) = \det{(\Omega_H)}^{-1/2} g[(U_\tau(H) - \mu_U)' \Omega_H^{-1}(U_\tau(H) - \mu_U)] \tag{1}$$

where $g(.)$ is a non-negative monotonic function and does not depend on τ with Ω_H an $H \times H$ covariance matrix. This assumption is restrictive but encompasses the majority of forecasting environments encountered in practice and makes the justification for the calculations that follow straight-forward. We call (1) the path-error generating distribution.

This distribution reflects a variety of sources of uncertainty associated with the predictions in $\widehat{Y}_\tau(H)$, among these the most common are: (1) uncertainty about the data generating distribution of $Y_{\tau,H}$ from which innovations are drawn; (2) uncertainty about the null model used to describe the data generating distribution; and (3) uncertainty about the parameter estimates of the null model in finite samples with respect to their population values. The reader is referred to Clements and Hendry (1998) for a more exhaustive list. For our purposes, it is unnecessary to be explicit about this breakdown. Our framework also extends to forecasting the paths of a vector of random variables and we do this in the simulations,

but here we prefer to keep the discussion simple. It will be important for our purposes that the path-error generating distribution be stable over time. This assumption is less restrictive than assuming that the data generating distribution is itself stable. In the next sections we discuss some of the particular statistical issues on forecasting uncertainty with particular effort in distinguishing the results that do not require distributional assumptions from those that do. The availability of N observations for evaluation will allow us to characterize empirically the path-error generating distribution.

2.1 Period-Specific Prediction Intervals

Consider first the simpler problem of constructing a prediction interval for the individual elements of the path-forecast. That is, for each $y_{\tau+h}$, call this interval S_h, such that $P(y_{\tau+h} \in S_h) \geq 1 - \alpha$ for some pre-specified error-level α chosen by the researcher and for $h = 1, ..., H$. In the simpler case of unimodally distributed u_τ^h (to avoid having to discuss disjoint intervals here),

$$S_h = \left\{ y_{\tau+h} : \underline{c} \leq w_h(u_\tau^h) \leq \overline{c} \right\} \tag{2}$$

where $w_h(.) : \mathbb{R} \to \mathbb{R}$. In the simplest case $w_h(u_\tau^h) = u_\tau^h = y_{\tau+h} - y_\tau(h)$ and an empirical prediction interval could be calculated from the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\left\{ \widehat{u}_\tau^h \right\}_{\tau=R+1}^{T-h} = \left\{ w_h(\widehat{u}_\tau^h) \right\}_{\tau=R+1}^{T-h}$ when N is sufficiently large. If the $\left\{ \widehat{u}_\tau^h \right\}_{\tau=R+1}^{T-h}$ are thought to be roughly Gaussian and using

$$\widehat{\sigma}_h^2 = \frac{1}{N-1} \sum_{\tau=R+1}^{T-h} (\widehat{u}_\tau^h - \overline{u}^h)^2$$

with $\overline{u}^h$ denoting the sample mean, then another example is

$$w_h(\widehat{u}_\tau^h) \equiv \frac{\widehat{u}_\tau^h - \overline{u}^h}{\widehat{\sigma}_h} \sim t_{N-1}$$

6

from which symmetric and approximately $100(1 - \alpha)\%$ coverage level predictive intervals can be easily constructed as is customary from the appropriate quantiles of the t-distribution.

Distribution-free methods for any population with a continuous density function can be used to construct prediction intervals using the theory of ranks. Suppose a sample of $\{y_t\}_{t=1}^{R}$ is available and weakly stationary (stationarity of the y_t could be trivially replaced by stationarity of the u_τ^h and rank the $\hat{u}_\tau^h$ instead but we prefer to keep the presentation simple) and rank in ascending order the elements in the sample such that $y_{[1]} \leq ... \leq y_{[R]}$ where $y_{[i]}$ denotes the i-th ranked observation. Danziger and Davis (1964) show that the probability that $y_{\tau+1}$ will fall between $y_{[r_1]}$ and $y_{[r_2]}$ for $1 \leq r_1 < r_2 \leq R$ is

$$P(y_{[r_1]} \leq y_{\tau+1} \leq y_{[r_2]}) = \frac{r_2 - r_1}{R + 1}. \tag{3}$$

Notice that in this case we would likely set $R = T$ since there is no advantage to reserving some of the sample for out-of-sample forecast evaluation. Intervals based on the data itself are disarmingly simple to construct, but do not take advantage of the conditioning information set. Rather, their derivation is based on the unconditional distribution of the data over the sample and hence may be too wide relative to the more traditional approach. In addition, small sample sizes may make these intervals particularly imprecise.

In order to assess the errors of coverage from different methods we introduce the auxiliary variable $d_\tau^h \in \{0, 1\}$. This variable records whether a given realization $y_{\tau+h}$ falls outside the pre-determined prediction region S_h, $d_\tau^h = 1$, or not, $d_\tau^h = 0$. If we choose S_h such that $P(y_{\tau+h} \in S_h) \geq 1 - \alpha$ then $E(d_\tau^h) = P(d_\tau^h = 1) = \alpha$, and the auxiliary variable d_τ^h provides a simple way to calculate the empirical value of α through averaging either in applications or in simulations.

2.2 Simultaneous Prediction Intervals

Constructing prediction regions for $Y_{\tau,H}$ is considerably more complicated. Suppose the $\{S_h\}_{h=1}^{H}$ individual prediction regions are independent across horizons to begin with (an assumption that usually fails in time series data, but temporarily set aside for expositional purposes here). Using the binomial distribution, it is easy to see that the probability that h out of the next H realizations of y_τ will fall within the individual per-period prediction intervals is given by

$$B(h, H, 1 - \alpha) = \sum_{i=h}^{H} \binom{H}{i} (1 - \alpha)^i \alpha^{H-i}. \tag{4}$$

For example, if $\alpha = 0.05$, $h = 5$ and $H = 6$ then $B(5, 6, 0.95) = 0.9672$ but this probability drops down to 0.7351 if all six realizations are required to fall inside the intervals defined by the individual prediction regions, a number that is quite different from the intended point-wise 0.95 level.

Distribution-free arguments similar to those used in expression (3) can be used to construct intervals that include at least h out of the next H forecasts. This provides a natural link between expressions (3) and (4).[2]   Using the same rank ordering of the data for $1 \leq r_1 < r_2 \leq R$, then the probability that h out of the next H forecasts will fall between

---

[2] Further explanations as well as the following two formulas can be found in Chew (1968).

$y_{[r_1]}$ and $y_{[r_2]}$ is

$$P\left\{ y_{[r1]} \leq y_{\tau+i} \leq y_{[r2]}\right\} = \frac{\sum_{j=h}^{H} \left(\begin{array}{c} (H-j) + (r_1 + R - r_2) \\ \\ H-j \end{array} \right) \left(\begin{array}{c} j + r_2 - r_1 - 1 \\ \\ j \end{array} \right)}{\left(\begin{array}{c} R+H \\ \\ H \end{array} \right)}$$

for all i in the subset of indexes of $\{1,...,H\}$ with cardinality at least equal to h. As an example, if we choose $1 = r_1 < r_2 = R$ and require that all H future realizations $\{y_{\tau+h}\}_{h=1}^{H}$ be contained within the interval $[y_{[1]}, y_{[R]}]$ then the probability associated with this interval is

$$P\left\{ y_{[1]} \leq y_{\tau+1},...,y_{\tau+H} \leq y_{[R]}\right\} = \frac{R(R-1)}{(R+H)(R+H-1)}.$$

At this point it is useful to define another auxiliary variable for the purposes of evaluating path-forecast error rates. Let, $k_{\tau}^{H} = \sum_{h=1}^{H} d_{\tau}^{h}$, that is, the number of realizations in the path that fall outside the union of per-period prediction regions S_h. In multiple comparison situations, it is often desired to control the probability error rate

$$P(k_{\tau}^{H} > 0) = \gamma$$

which we will call here the *family-wise prediction error rate* (FWPER) to use a nomenclature that is similar to that used in hypothesis testing. In the example just described, notice that by choosing individual error rates with $P(d_{\tau}^{h} = 1) = \alpha_h$ for $h = 1, ...H$, we have that

$$P(k_{\tau}^{H} > 0) = B(H, H, 1 - \alpha_h)$$

so that $\gamma = 0.7351$ if $H = 6$ and $\alpha_h = 0.05$.

9

A crude bound relating γ with the individual α_h is provided by Bonferroni's inequality, using the fact that

$$1 - P(k_\tau^H > 0) = 1 - \alpha_1 - ... - \alpha_H$$

that is

$$P(\cap_{h=1}^H d_\tau^h = 0) \geq 1 - \sum_{h=1}^H P(d_\tau^h = 1).$$

Hence, by choosing $\alpha_h = \alpha/H$ one ensures that $\gamma = \alpha$. There are, of course, several alternatives to providing a bound for γ using the α_h under more restrictive assumptions and there are step-wise methods that can be used, such as Holm's (1979) procedure. However, Bonferroni's procedure is by far the most commonly used adjustment when an adjustment is made at all. For extensions of the Bonferroni procedure the reader is referred to Lehmann and Romano (2005).

An alternative criterion to FWPER is, using analogous notation to the hypothesis testing literature, k-FWPER. Here the goal is to control the probability error rate

$$P(k_\tau^H > k) = \gamma$$

with $k > 0$. Returning to our previous example and using expression (4), we have

$$P(k_\tau > k) = B(H - k, H, 1 - \alpha) = \gamma.$$

As mentioned above, under independence with $\alpha = 0.05$, $k = 1$ and $H = 6$ results in $\gamma = 0.9672$ which is close to $1 - \alpha$. It is easy to see that by making the original FWPER criterion more flexible, k-FWPER can be used to account for the length of the path-forecast where one can use k as a function of H. Although there is no specific function that would

10

provide a tight connection between k and H, the researcher can choose slightly more flexibility on an ad-hoc basis.

Common practice in economics is to construct per-period (point-wise) prediction intervals with the desired coverage and ignore any issues related to simultaneity and control of the prediction error rate of the family. This practice is not entirely without justification. Implicit to this common practice is perhaps the desire to control the expected prediction error rate (EPER)

$$E\left(\frac{k_\tau^H}{H}\right) = \gamma, \tag{5}$$

which has the nice property of being directly related to the marginal performances of each individual prediction interval. The EPER is related to α_h by

$$E\left(\frac{k_\tau^H}{H}\right) = \frac{\alpha_1 + ... + \alpha_H}{H} = \gamma$$

and although there is no requirement that all $\alpha_h = \alpha$, it is immediately clear that such a choice would guarantee that $\gamma = \alpha$. In the special case were the family consists of a single prediction interval,

$$\gamma = E\left(\frac{k_\tau^1}{1}\right) = E\left(d_\tau^1 = 1\right) = P(d_\tau^1 = 1) = \alpha$$

and EPER and FWPER coincide but in general

$$E\left(\frac{k_\tau^H}{H}\right) \le P\left(k_\tau^H > 0\right) \le H \times E\left(\frac{k_\tau^H}{H}\right).$$

Independence across horizons will often fail in a forecasting setting – positive serial correlation in y_t means that if $y_{\tau+1}$ is above $\widehat{y}_\tau(1)$ then $y_{\tau+2}$ is more likely to be above $\widehat{y}_\tau(2)$ than otherwise. In terms of the path-error generating distribution of $U_\tau(H)$, this simply means that Ω_H will not be spherical.

Consider then the problem of constructing a prediction region S such that $P(Y_{\tau,H} \in S) \geq 1 - \alpha$. In particular, it helps to transform the problem as follows. Let

$$S = \left\{ Y_{\tau,H} : \underline{\delta} \leq W(U_\tau(H)) \leq \overline{\delta} \right\}$$

where $W(.) : \mathbb{R}^H \to \mathbb{R}$ is a one-dimensional distance function.

In statistics, a popular distance function $W(.)$ related to the class of elliptical distributions described in (1) is the Mahalanobis distance[3]

$$W_\tau = \sqrt{(Y_\tau(H) - Y_{\tau,H})' \Omega_H^{-1} (Y_\tau(H) - Y_{\tau,H})} \in [0, \infty) \tag{6}$$

where Ω_H can be estimated from the evaluation sample using

$$\widehat{\Omega}_H = \frac{1}{N-1} \sum_{\tau=R+1}^{T-H} \left(\widehat{U}_\tau(H) - \overline{U}(H) \right) \left(\widehat{U}_\tau(H) - \overline{U}(H) \right)' \tag{7}$$

where

$$\overline{U}(H) = \frac{1}{N} \sum_{\tau=R+1}^{T-H} \widehat{U}_\tau(H)$$

and will have an expectation of 0_H if forecasts are unbiased.

The Mahalanobis distance evaluates the distance of a point from the center of the ellipsoid accounting for correlation. To help draw the connection between a prediction region based on (6) with the per-period prediction regions based on (2), notice that Ω_H is symmetric and positive definite and hence admits a unique Cholesky decomposition $\Omega_H = QQ'$ where Q is a lower triangular matrix. Another way to think of this decomposition is as the projection of the h-step ahead forecast error into the 1 to $h - 1$ previous forecast errors.

Define the path of orthogonalized forecast errors as

$$V_\tau(H) = Q^{-1}(U_\tau(H) - \overline{U}(H)).$$

---

[3] Note that an elliptical distribution is not necessarily normal.

Rewriting (6) in terms of these orthogonalized forecast errors, we have

$$
\begin{aligned}
W_\tau &= \sqrt{(U_\tau(H) - \overline{U}(H))'\Omega_H^{-1}(U_\tau(H) - \overline{U}(H))} = \sqrt{V_\tau(H)'V_\tau(H)} \qquad (8) \\
&= \sqrt{\sum_{h=1}^{H} v_\tau(h)^2}.
\end{aligned}
$$

Summarizing, the simultaneous prediction region S with $P(Y_{\tau,H} \in S) = 1 - \alpha$ based on the Mahalanobis distance W is an ellipsoid given by the condition

$$
\Pr(W_\tau^2 \leq \delta^2) = 1 - \alpha,
$$

which can be transformed into the sphere given by the condition

$$
P\left(\sum_{h=1}^{H} v_\tau(h)^2 \leq \delta^2\right) = 1 - \alpha. \qquad (9)
$$

A sphere such as that in the previous expression, is still a complicated object to represent in two-dimensional space – a rectangular region is far more convenient. The question we examine next is how to choose the vertices of this rectangular region.

2.3   Per-period Simultaneous Prediction Intervals: A *Balanced* Approach

Expression (9), is a correct probabilistic statement about the proportion of realizations of $Y_{\tau,H}$ expected to fall within the multi-dimensional sphere described in that expression for an appropriately chosen value of δ. However, this is not something we can represent in two-dimensional space. One solution is to use a version of Scheffé's (1953) projections to our problem.

Consider the easier case where $H = 2$ first, with the additional simplification implied by the assumption

$$
Y_{\tau,2} \sim N(Y_\tau(2), I).
$$

Then the region such that $P(Y_{\tau,2} \in S) = 1 - \alpha$ is a circumference that results from noticing that

$$(Y_{\tau,2} - Y_\tau(2))'(Y_{\tau,2} - Y_\tau(2)) \sim \chi_2^2$$

and hence

$$P\left[W_\tau^2 \leq \chi_2^2(1 - \alpha)\right] = 1 - \alpha \tag{10}$$

where $\chi_2^2(1 - \alpha)$ refers to the critical value of a chi-squared random variable with two degrees of freedom evaluated at confidence level $1 - \alpha$. This prediction region is displayed in figure 1 for $\alpha = 0.05$. Scheffé (1953) introduced a rectangular projection of this region using an inversion of the usual F-statistic in a simultaneous hypothesis test. The goal was to search for all the directions where the null hypothesis could be rejected. Such a search is implicitly a data-snooping exercise and the virtue of the Scheffé (1953) procedure is to correct the statistical statements to account for this search process.

Geometrically and in the context of the illustrative example in figure 1, the Scheffé (1953) intervals can be calculated without difficulty. These are $y_{\tau+h} \in y_\tau(h) \pm \chi_2^2(1 - \alpha)^{1/2}$ for $h = 1, 2$ (when the covariance matrix has to be estimated one can use instead the F-distribution critical values, which is the original form that appears in Scheffé, 1953). The Scheffé (1953) *bounds* are displayed in figure 1 as the box that contains the circumference corresponding to expression (10). As Miller (1981) explains, Scheffé (1953) intervals in practical applications will generally be wider than Bonferroni intervals for small H, but narrower when H is large. There is, however, another important lesson that figure 1 contains in regard to the allocation of uncertainty across horizons.

The bounds of the Scheffé (1953) projections are, in our opinion, unreasonably conser-

vative for most path-forecasting problems. The reason is that they bound the worst case scenario in which the total uncertainty of the pair $y_{\tau+1}$, $y_{\tau+2}$ is allocated to one realization but not the other. Protection against such an extreme event may make sense when the objective is to quantify the worst possible risks. But in most forecasting applications the end-user probably favors a more balanced allocation of uncertainty. The strictly balanced allocation of uncertainty for the pair $y_{\tau+1}$, $y_{\tau+2}$ is displayed in figure 1 as the inner rectangular region encased by the circumference. That region intersects the original circumference at the point where $y_{\tau+1}$ and $y_{\tau+2}$ simultaneously achieve maximum variability. The diagonal represented in figure 1 has slope equal to the ratio of the variances of $y_{\tau+1}$ and $y_{\tau+2}$ (which in this case is 1). In what follows we provide a more formal derivation of how this region is constructed and the implicit forecast-error control that would justify such a calculation.

The following lemma from Turner and Bowden (1977), which is a version of the Cauchy-Schwarz inequality, will be useful in the derivations that follow.

Lemma 1 *Let a and V be $H \times 1$ real vectors. Let δ be a real constant. Then $V'V \leq \delta^2$ if and only if $|a'V| \leq \delta(a'a)^{1/2}$ for all real vectors a.*

Returning to expression (8) and applying the lemma, we have

$$P(W_\tau^2 \leq \delta^2) = P(V_\tau(H)'V_\tau(H) \leq \delta^2) = P\left(\left|a'V_\tau(H)\right| \leq \delta(a'a)^{1/2} \text{ for all } a\right)$$

where a is any $H \times 1$ real valued vector. As explained in Turner and Bowden (1977), if at least one element of a is restricted to a specific value, one obtains the inequality

$$P(V_\tau(H)'V_\tau(H) \leq \delta^2) \leq P\left(\left|a'V_\tau(H)\right| \leq \delta(a'a)^{1/2}\right).$$

For reasons that will be apparent momentarily, consider what happens when $a = (\frac{1}{H}, ..., \frac{1}{H})'$.

Notice that $a'a = \frac{1}{H}$ and hence

$$P\left(|a'V_\tau(H)| \leq \sqrt{\frac{\delta^2}{H}}\right) = P\left(\left|\frac{1}{H}\sum_{h=1}^{H} v_\tau(h)\right| \leq \sqrt{\frac{\delta^2}{H}}\right) \geq P\left(V_\tau(H)'V_\tau(H) \leq \delta^2\right) = 1 - \alpha.$$

(11)

The middle term of expression (11) makes a probabilistic statement about the interval associated with the average value of the orthogonalized components of the path-forecast.

Now suppose that we consider H vectors of identical length, the vector $a_1 = a = (\frac{1}{H}, ..., \frac{1}{H})'$ and all $H - 1$ vectors $a_2, a_3, \ldots, a_H$ orthogonal to it. In many cases, these can be vectors with typical element $\{\pm\frac{1}{H}\}$.[4] We then have that

$$P\left(|a'V_\tau(H)| \leq \sqrt{\frac{\delta^2}{H}}\right) \geq P\left(\bigcap_{h=1}^{H} |a'_h V_\tau(H)| \leq \sqrt{\frac{\delta^2}{H}}\right) \geq 1 - \alpha,$$

(12)

where the middle term describes an H-cube. Figure 1 displays the H-cube associated with this probabilistic statement for $H = 2$, which can be seen as the rhombus encasing the circumference given by (10). Rotating this rhombus gives the box corresponding to the Scheffé bounds mentioned above.

As argued above, the Scheffé bounds appear rather conservative for the purpose of path-forecasting. We therefore propose a smaller H-cube which intersects the original circumference at the point where $v_\tau(1)$ and $v_\tau(2)$ simultaneously achieve maximum variability. In H-dimensional space, this region is defined by the H conditions

$$|V_\tau(H)| \leq \sqrt{\frac{\delta^2}{H}}\mathbf{i}_H,$$

where $\mathbf{i}_H$ is an $H \times 1$ vector of ones, and δ will depend on the desired α. Note that this

---

[4] These cases are given by $H = 2$ and $H = 4k$ with $k \in \mathbb{N}$, where the orthogonal vectors can be obtained as the columns of an $H \times H$ Hadamard matrix multiplied by $1/H$. In the other cases, some elements of a_2, a_3, $\ldots$, a_H differ from $\pm\frac{1}{H}$.

region guarantees that the allocation of prediction-error control is balanced in the sense of Beran (1988).

Recall that $V_\tau(H) = Q^{-1}(U_\tau(H) - \overline{U}(H))$ so that to express the intervals in terms of the original forecast errors we have

$$\left| Q^{-1}(U_\tau(H) - \overline{U}(H)) \right| \leq \sqrt{\frac{\delta^2}{H} \mathbf{i}_H}. \tag{13}$$

Premultiplying expression (13) by $|Q|$ and then applying the matrix inequality $|AB| \leq |A||B|$ to the first expression, we get[5]

$$\left| U_\tau(H) - \overline{U}(H) \right| \leq |Q| \sqrt{\frac{\delta^2}{H} \mathbf{i}_H},$$

and if the forecasts are unbiased such that $\overline{U}(H) = 0$ so that $\overline{Y}_{\tau,H} = \overline{Y}_\tau(H)$, then

$$Y_{\tau,H} \in Y_\tau(H) \pm |Q| \sqrt{\frac{\delta^2}{H} \mathbf{i}_H}. \tag{14}$$

Expression (14) forms the basis for the per-period prediction intervals that we propose in this paper and which we will call *Scheffé bands* or *intervals* to use the Jordà and Marcellino (2010) nomenclature and to stand in opposition to the Scheffé bounds discussed earlier.

Several remarks deserve mention. First and returning to figure 1, the intervals proposed in expression (14) correspond to the inner square mentioned earlier. That square is tangent to the circumference given by (10) and the rhombus given by the middle term of expression (11). The point-wise intervals implied by that square are of equal length to achieve balance as desired. Second, notice that the Scheffé intervals that we propose in (14) are narrower

---

[5] Jordà and Marcellino (2010) neglected to invert Q using the absolute value. This is usually not a problem in practice because path-forecasts tend to be positively correlated. However, the general result requires the absolute value transformation. This correction was first introduced in a manuscript by Staszewska-Bystrova (2012). The matrix inequality used can be found in Lütkepohl (1996, p.40).

than the intervals based on the marginal distribution of each element (which in this example would correspond to the $[-1.96, 1.96] \times [-1.96, 1.96]$ square) and much narrower than the Scheffé bounds. This is explained by our desire to control for general variability along the path rather than controlling for worst-case scenario type variability. Third, in practice we have $\widehat{Y}_\tau(H)$ rather than $Y_\tau(H)$ and Q has to be estimated. That is, the elements of prediction interval itself are random. When evaluated at their average (or estimates), we obtain the traditional prediction interval. In this paper space considerations prevent us from discussing tolerance intervals, a probabilistic statement related to the quantiles of the distribution of the prediction interval. Tolerance intervals would be natural in any discussion associated with measures of Value at Risk (VaR), for example.

In practice and when the null distribution is unknown and one wants to make as few assumptions as possible, the evaluation sample or N-sample for short (R-sample will refer to the sample used to estimate the model and $T = R + N + H$ comprises all the available data), can be used to construct the Scheffé intervals. Suppose one wants to provide prediction intervals for $\widehat{Y}_T(H)$. The N-sample can then be used to estimate $\widehat{\Omega}_H$ and the Cholesky decomposition to calculate $\widehat{Q}$. Finally, using N-sample estimates of $\overline{Y}_\tau(H)$ together with $\widehat{\Omega}_H$, one can generate $\{\widehat{W}_\tau^2\}_{\tau=R+1}^{T-H}$ and use $\widehat{W}_{[1-\alpha]}^2$ as an estimate of δ^2. Alternatively and if the data are reasonably well approximated by the Gaussian distribution, δ^2 could be calculated as the critical value of a chi-square distributed random variable at $1 - \alpha$ confidence level or, if a small sample correction is desired, as H times the critical value of an F-distributed random variable with H, N degrees of freedom at $1 - \alpha$ confidence level.

Another practical matter has to do with the fact that the critical value chosen depends

on the length of the interval considered. As a result, the width of the intervals for the earlier forecasts depends on the researcher's choice of H. This is a natural consequence of the simultaneous nature of the evaluation problem. However, one could construct stable per-period intervals with some additional coverage replacing the intervals in expression (14) with

$$Y_{\tau,H} \in Y_\tau(H) \pm |Q| \left[\sqrt{\frac{\delta_h^2}{h}} \right]_{h=1}^{H} \tag{15}$$

where δ_h^2 denotes the critical value calculated for period h (either empirically or using the chi-square or F-distribution assumption) and where the $H \times 1$ vector in brackets has typical element $\sqrt{\frac{\delta_h^2}{h}}$ for $h = 1, ..., H$. Below we use this version of the Scheffé bands for the Monte Carlo simulations and the application since we feel researchers may feel more comfortable choosing a method of prediction interval construction that is independent of H.[6]

We conclude this section with a brief discussion of prediction-error rate control. As we mentioned earlier, there does not appear to be a universally agreed upon control metric. FWPER seems appropriate if one is concerned about controlling for even the occasional stray realization in an otherwise properly predicted path-forecast. We have instead argued in favor of a metric that considers the entire path-forecast variability simultaneously while generating per-period prediction intervals that allocate uncertainty in a balanced manner. Such an interval is designed to capture stray *trajectories* as opposed to stray *observations*. In what follows we will call control of the probability error rate based on the Mahalanobis distance induced by the path-forecast constrained to have equal per-period variability as

---

[6] The difference between expressions (14) and (15) depends on the choice of α. For common values of α such as 0.05 or 0.32, (15) tends to yield wider intervals.

Wald control.[7]   That is, the Scheffé bands that we propose generate an implied value

for W^2 in expressions such as (10). Call this value W_s^2. In Monte Carlo simulations we

then compute the proportion of the simulated path-forecasts whose $\widehat{W_\tau}^2$ exceeds W_s^2. Proper

calibration of the bands produces Wald control that corresponds to the desired α error rate.

It should be clear that neither version of the Scheffé bands proposed is designed to

guarantee correct FWPER control. And yet in many applications and depending on the

serial correlation patterns in the data, it will turn out that FWPER is close to the equivalent

α control level. In contrast, Bonferroni bands control the FWPER (by design), but their

Wald coverage can deviate substantially from the pre-specified nominal level. We examine

these and many other features in the simulation exercises discussed in the next section.

3 Probability Coverage of Alternative Prediction Bands: Simulation Evidence

We now present simulation evidence on the probability coverage of different types of prediction bands for a path-forecast. The next section describes the experimental design, and the following section reports the results. Concerning FWPER control, if the actual coverage of the bands exceeds (falls short of) the nominal level, the bands are commonly said to be too wide (narrow). For convenience, we adopt the same terminology when evaluating the Wald coverage.

---

[7] Wald control seems appropriate since the Mahalanobis distance is the main ingredient in constructing the celebrated Wald statistic.

3.1 Experimental design

We simulate data from the AR(1)-model

$$y_t = \rho y_{t-1} + \epsilon_t, \tag{16}$$

with coefficient ρ equal to 0.5 or 0.9, $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = 1$. Three distributions are considered for the errors ϵ_t. The first is the standard normal distribution, the second is the t-distribution with 4 degrees of freedom, scaled to have a variance equal to 1. In both cases, the errors are i.i.d. For the third distribution, ϵ_t is generated by the ARCH process

$$\epsilon_t = \sqrt{z_t} v_t$$

$$z_t = a_0 + a_1 \epsilon_{t-1}^2$$

where v_t is i.i.d. standard normal, and $a_0 = 1 - a_1$, so that $E[\epsilon_t^2] = 1$. The parameter a_1 is set to $1/\sqrt{3}$. Thus, in the case of the t-distributed errors and the ARCH errors, the kurtosis of ϵ_t does not exist for the parameter values chosen here, but it does exist in their local neighborhoods. Choosing these parameter values yields similar unconditional non-normal distributions of ϵ_t. For example, using the critical value $c = 0.99$ (1.96, 2.58), which yields a coverage of $P(-c < \epsilon_t < c) = 0.68$ (0.95, 0.99) in the case of the standard normal distribution, the coverage equals 0.77 (0.95, 0.98) in the case of the t-distribution and 0.75 (0.95, 0.98) in the case where ϵ_t is generated by the ARCH process. Note that the (unconditional) covariance matrices of the forecast errors are identical for all three error distributions considered.

Let R denote the estimation sample size, N denote the predictive sample size, and H denote the length of the path-forecast considered. We consider the values $R = 25, 100$;

$N = 40, 80$; and $H = 1, 4, 8$ and 12. For each combination R, N and H we generate 10,000 Monte Carlo samples.

At each replication, we use that sample to generate path-forecasts of length H over a predictive sample N using OLS estimates from an AR(1) model without constant fitted on R observations. Then, we construct three types of prediction bands: (1) traditional marginal error bands; (2) Bonferroni bands; and (3) Scheffé bands (using expression (15)). We consider both the analytic formulas derived in Jordà and Marcellino (2010), based on the null model being the DGP, as well as the empirical formulas described in the previous section based on the forecast errors only.[8]   The empirical covariance matrices of the forecast errors are estimated assuming unbiased forecasts. We examine bands for 68% and 95% probability coverage rates.

As mentioned above, coverage is evaluated in terms of FWPER and Wald control. FWPER control means that any path with one or more elements outside the prescribed bands is considered to fall outside the bands. Wald control is calculated using the Mahalanobis distance squared and then recording as being outside the bands all those paths with distances above W_s^2, defined in the previous section. We do not use the empirical critical values because the empirical predictive sample sizes usually do not allow for accurate estimates of their α-quantiles. In practice, a normal distribution for error terms is frequently assumed and the normality hypothesis, probably, could often not be rejected due to the small sample size even if the errors are not normal. Therefore, we consider situations where the errors are non-normal, but the bands are constructed using the normality assumption, and evaluations

---

[8] In the case of using the analytic formulas, the size N of the predictive sample is obviously inconsequential.

of the Wald coverage also rest on the normal distribution. This is probably a more realistic situation in practice.

In addition to the design described above, we also investigate a simple case of misspecification, where we assume that the forecaster uses the misspecified model

$$y_t = \epsilon_t$$

instead of the true model given by (16), so that $\widehat{Y}_T(H)$ is an $H \times 1$ vector of zeros. The only parameter that has to be estimated then in the case of model-based bands is the second moment of y_t. Finally, the true covariance matrix needed for evaluating the Wald coverage is unknown and has to be estimated from the N-sample.[9]

3.2 Simulation Results

Tables 1 and 2 summarize the results of our experiments. Table 1 contains the results obtained with forecasts from the null model, whereas the results for the case of misspecification can be found in Table 2.

The top panel in Table 1 refers to the case where all parameters of the model are known. The left half of these results contains information about the deviations of coverage with respect to the nominal level. For example, a value of 10 means that the coverage is 10 percentage points higher than the nominal level. The results displayed are averages over the deviations observed for both values of ρ, both values of α, and the four values of H considered. In the right half of the table, the averages over the absolute deviations can be found. If the values for the average deviations and the absolute deviations coincide,

---

[9] Additional simulations and results for a VAR model based on Stock and Watson (2001) with several types of misspecifications can be found in Jordà, Knüppel, and Marcellino (2010).

23

this means that the sign of the deviations is identical for all combinations of ρ, α, and H considered. The individual coverage results can be found in Table A.1.

The remaining results in Table 1 are concerned with cases where the bands are based on estimates. We report the deviations with respect to the case where all parameters are known. Thus, the results contain information about the effects of estimation uncertainty, not about nominal coverage per se. That is, it might be possible that, say, the bands based on estimated parameters are narrower than the bands based on known parameters, leading to smaller coverage. Yet, if the bands in the case of known parameters are too wide, it could be that both effects cancel out, so that the coverage of the bands based on estimated parameters is actually equal to the nominal level. However, here we are primarily interested in the effects of estimation uncertainty, so that smaller deviations from the case of known parameters are preferable.

The left half of the results for each case of estimated bands considered contains information about the average deviations of coverage with respect to the case of known parameters, where the results are averages over the deviations observed for both values of ρ, both values of α, and the four values of H. In the right half, the averages over the absolute deviations with respect to the the case of known parameters can be found. The individual coverage results can be found in Tables A.1 - A.7.

Turning to the case of known parameters, for normal errors the results of Jordà and Marcellino (2010) are reproduced. Marginal bands are far too narrow with respect to FWPER and Wald control, Bonferroni bands tend to be moderately too wide for FWPER control and have problems with Wald control, being too narrow in some cases and too wide in

others. Scheffé bands are too narrow with respect to FWPER control as expected, but only marginally too wide with respect to Wald control. Considering the individual results, one finds that the FWPER control of Scheffé bands improves with the persistence of the process.

Concerning the non-normal errors, the results for the t-distribution and the ARCH process are fairly similar. In general, the coverage of the bands becomes larger. This might be surprising at first sight, but it is directly related to the coverage for certain critical values reported above. For a nominal coverage of 68%, the bands based on a normal distribution are actually too wide if the errors are t-distributed or follow an ARCH process. For a nominal coverage of 95%, the normal bands have approximately the correct coverage.[10]

If the bands are constructed using estimates of ρ and $E\left[\epsilon_t^2\right]$, the bands tend to be narrower than with known parameters. While there are exceptions to this for the marginal bands, it holds for virtually all specifications considered in the case of the Bonferroni and the Scheffé bands. For all bands, the average effects are of similar magnitudes. The effects of estimation uncertainty increase with ρ and H. The effects can be large for $R = 25$, but they are fairly contained for $R = 100$. For example, with $\rho = 0.9$, $\alpha = 0.32$, $H = 12$, and $R = 25$, the FWPER coverage of the Bonferroni bands is 15 percentage points smaller than with known parameters. For $R = 100$, the difference drops to 5 percentage points. The deviations from the case with known parameters tend to be smaller for the non-normal errors.

If forecast errors are used to construct the bands, the estimation uncertainty also tends to narrow the bands. This effect is most pronounced for the Scheffé bands, which, in contrast to the other bands, require the additional estimation of covariances. The non-normal errors

---

[10] Had we considered 99% bands, we would, of course, find, that the bands based on the normal distribution tend to have a smaller coverage in the cases of the non-normal errors.

cause stronger deviations than the normal errors, with the ARCH errors having the largest impact. This result appears plausible, because the conditional heteroskedasticity makes inference about the unconditional second moments of the forecast errors difficult in small predictive samples. Comparing the results for $R = 25$ to those for $R = 100$, it turns out that the estimation uncertainty with respect to ρ has negligible effects on the coverage. In contrast to that, the predictive sample size N is decisive. With $N = 40$, the average deviations attain up to -9 percentage points for FWPER control and up to -12 percentage points for Wald control. These numbers drop to -6 and -7 percentage points, respectively, with $N = 80$. Looking at the individual coverage results in Tables A.4 - A.7, it turns out that deviations larger than 10 percentage points mainly occur for $H = 8$ and $H = 12$ if $N = 40$, whereas with $N = 80$, such deviations are mostly restricted to $H = 12$ only.[11]

Table 2 reports the results for the misspecified forecasting model and has the same structure as Table 1. The individual coverage results can be found in Tables A.8 to A.12. However, the results which now serve as the reference concerning the deviations in coverage are obtained by simulations with $N = 1000$. In this case, the covariance matrix of the forecast errors is estimated in a sufficiently accurate manner, so that the coverage obtained can be regarded as the actual coverage given the misspecification considered. The results are actually not too different from the results for the well-specified model with known parameters.[12]

---

[11] It should be noted that, concerning the results for bands based on forecast errors, the estimation uncertainty affecting the bands is only one reason for the deviations with respect to Wald coverage. A second reason is given by the effect of the estimation uncertainty on the covariance matrix which is used to determine the Wald coverage of the forecast path. Additional simulations not reported here show that the second effect accounts for -1 to -3 percentage points of the average deviations with respect to the Wald coverage if $N = 40$. With $N = 80$, the effect becomes negligible for the average deviations. Concerning the individual deviations, the effect tends to be larger in the case of longer paths.

Comparing the Scheffé bands and the Bonferroni bands for the misspecified model, both bands have almost identical FWPER control on average, but the Scheffé bands perform much better with respect to Wald control.

If the bands are constructed based on the misspecified model instead, the FWPER control changes strongly for the Scheffé bands. For the marginal and the Bonferroni bands, the changes are only moderate for $R = 25$ and small for $R = 100$. This, however, is due to the special case analyzed here. Both methods for the construction of bands give identical results, asymptotically, for the marginal and the Bonferroni method here, because the variance of the actual forecast errors is equal to $\left(1 - \rho^2\right)^{-1}$ for each horizon, and the same holds for the variance of the forecast errors derived from the misspecified model. The Scheffé bands, however, depend on the method of construction. Only if the forecast errors are used for the construction, the Scheffé bands widen with the forecast horizon. This example illustrates that the Scheffé bands contain more information than the other bands, since the positive correlation of the forecast errors only becomes visible in the (forecast error-based) Scheffé bands.

Since both methods for the construction of bands, asymptotically, yield identical marginal and Bonferroni bands, respectively, their coverage does not depend on the method used. Yet, Table 2 suggests that the Wald control of model-based bands is very different from the Wald control of bands generated from forecast errors. The reason for this apparent difference is given by the misspecification of the covariance matrix used to evaluate the Wald coverage

---

[12] The reason for the similarity is given by the similarity of the correlations of the forecast errors. Consider, for example, the correlation between the 1- and the 2-step-ahead forecast error. In the case of correct specification, this correlation equals $\rho/\sqrt{1 - \rho^2}$, whereas it simply equals ρ in the case of misspecification. Both correlations are similar if $|\rho|$ is not too large, as, for example, in the case $\rho = 0.5$. For larger forecast horizons h and $h + 1$ with $h > 1$, the differences between the correlations become even smaller.

in the case of model-based bands. The evaluation of the Wald coverage, however, only makes sense if the estimator of the covariance matrix of the forecast errors is consistent. If the forecasting model is misspecified, this can only be achieved using the covariance matrix based on empirical forecast errors.

As in the case of correct specification and bands based on forecast errors, the coverage of the Scheffé bands is more affected by the estimation uncertainty than the coverage of the bands which do not make use of covariances. If ϵ_t follows an ARCH-process, the estimation uncertainty for the covariance matrix becomes larger than with homoskedastic errors, leading to stronger deviations of the coverage with respect to the reference case. However, the differences between the average coverages of the three error distributions are again small.

To summarize, if the parameters that the bands are based on have to be estimated, this tends to lead to lower coverage. The effects of estimation uncertainty are of similar magnitudes for all bands if the bands are based on a model. If they are based on forecast errors, the effects are strongest for the Scheffé bands, because the additional estimation of covariances is required. In the case of model misspecification, in contrast to the forecast-error-based construction of bands, the model-based construction can lead to pronounced coverage problems, and the evaluation of the Wald coverage can give very misleading results. If the errors are assumed to be normal although they are not, the effects on the coverage depend on the nominal level of α and the type of non-normality, but they are relatively small in most settings considered here.

4 An Application to the Greenbook Forecasts

The staff at the Board of Governors of the Federal Reserve routinely produces, among many other variables, inflation and output forecasts. These forecasts, colloquially known as the Greenbook forecasts (and more recently as Tealbook forecasts), are made available to all the members of the Federal Open Market Committee (FOMC) prior to each FOMC meeting. Because the forecasting model is not made available nor forecasting uncertainty routinely reported, we find it a natural environment where our methods can be applied. The sample of Greenbook forecasts is relatively large so that it can be used to evaluate our simultaneous prediction regions by splitting the predictive sample and saving the second subsample for evaluation. Notice that while the forecasts and the realizations are available, the null model is not and hence one has to rely entirely on the empirical approach we have proposed.

4.1 Data and Construction of Bands

We have collected quarterly Greenbook forecasts starting in 1974:Q2, for which $H = 5$ is available. The data for $h = 1$ represents a nowcast, so that we have path-forecasts up to 1-year ahead. The sample ends in 2003:Q4, since the Greenbook data are released with a delay of 5 years. Forecasts are for output growth (where output is measured by GNP until 1991, GDP afterwards) and inflation (measured as growth in the GNP deflator until 1991, in the GDP deflator afterwards) measured as $400 * ln(y_t/y_{t-1})$. For comparison with actual data, we use the chain-weighted GNP and GDP, and their deflators.[13]

   Using forecasts and corresponding actual data, we have constructed the sequences of

---

[13] Here, we use the data as published in 2009. In Jordà, Knüppel, and Marcellino (2010) we also consider the role of real time data in the context of this application.

path-forecast errors and tested them for normality using the test of Bai and Ng (2005). Having obtained non-rejections of normality in all cases, we proceed using the normality assumption.[14] We have then used the error sequences to estimate the Ω_H matrix empirically on the predictive sample. Of course, the analytic approach is unavailable since the null model that produces the forecasts is unknown.

Next, we construct marginal, Bonferroni and Scheffé bands for the growth and inflation forecasts for different nominal coverage rates. In figure 2 we plot the (one-sided) 95% bands, centered at zero for convenience.[15] The Bonferroni bands are in general wider than the Scheffé bands for low values of h, but sometimes narrower for higher values of h. The commonly used marginal bands are clearly the narrowest except for $h = 1$. A similar picture (but with larger values for each type of band) emerges for a 68% (one standard deviation) coverage rate.[16]

We now assess the relative performance of the bands in the application at hand in terms of actual coverage rates. For the evaluation, we split the sample at $T^* = 1985$:Q1. With the sample of forecast paths starting in 1974:Q2 and ending in T^*, we get 40 forecast paths and corresponding errors from which we construct the empirical estimate $\widehat{\Omega}_H$. This estimate is used to construct uncertainty bands for the forecast path from $T^* + 1$, i.e. from 1985:Q2. Then we roll the window of forecast paths to calculate $\widehat{\Omega}_H$, one observation at a time by deleting the first path and adding one path at the end, and we also roll the forecast path for which we construct uncertainty bands. We do so until the uncertainty bands are constructed

---

[14] Detailed results are available upon request.

[15] This means that for any given path-forecast, adding and subtracting the values we provide in the figure would generate the appropriate 95% prediction interval.

[16] Available upon request.

for the final path starting in $T = 2003$:Q4, for which the final window of forecast paths starting from 1992:Q4 to 2002:Q3 is used to determine $\widehat{\Omega}_H$. In this way, we have produced uncertainty bands for 75 forecast paths. This procedure yields a set of measures of path-forecast uncertainty that can be compared with the actual realizations in order to compute the actual coverage rate of each type of prediction band. Table 3 reports actual coverage rates under FWPER and Wald control for U.S. inflation and output growth for nominal levels of 68% and 95%.

As expected, the marginal bands perform very poorly, in particular for 68% nominal coverage rates, but also for 95% they are dominated by the Bonferroni and Scheffé bands. The ranking of the latter is less clear-cut. In terms of FWPER control, the coverage of the Bonferroni bands is, in general, closer to the nominal levels. For Wald control, the Scheffé bands tend to yield the best results. Thus, the empirical results are overall in line with those found in the Monte Carlo simulations, but they also confirm a certain robustness in the coverage of the bands even in the likely presence of structural changes in distribution of the forecast errors due to the Great Moderation described by McConnell and Pérez Quirós (2000).

5   Conclusions

This paper introduces new methods to compute prediction intervals with simultaneous probability coverage where variability is distributed evenly across the elements of the path-forecast depending on their variance and serial correlation. We believe such bands reflect the prediction uncertainty that is most likely to be relevant for the end-user of such forecasts. We make few assumptions to derive our results, including situations where the model generating the

31

forecasts may be unavailable. Using our methods, one can still construct prediction intervals using the empirical distribution of past forecast errors. Moreover, the empirical distribution appears to offer reasonable protection against possible misspecification of the forecasting model as our simulation evidence demonstrates.

There is no agreement as to the best metric of prediction error control for path-forecasts and this leads to considerable confusion in the literature as to how to determine and evaluate the probability coverage of any proposed prediction intervals. However, simulation evidence suggests that the bands that we discuss in this paper appear to work best under a variety of metrics and also under a variety of misspecification alternatives. Traditional methods of prediction interval calculation often work well under one metric but not another and suffer greatly under misspecification. In an empirical application to the Greenbook forecasts for inflation and output growth, we find that the use of traditional point-wise intervals based on the marginal distribution of forecast errors is not recommended. The Bonferroni bands yield the best results with respect to FWPER control but often provide poor Wald coverage. Scheffé bands have the best Wald coverage and perform reasonably well when evaluated with alternative metrics.

References

Bai, Jushan and Serena Ng (2005) "Tests for Skewness, Kurtosis and Normality for Time Series Data", *Journal of Business and Economic Statistics*, 23: 49-60.

Beran, Rudolph (1988) "Balanced Simultaneous Confidence Sets", *Journal of the American Statistical Association,* 83: 679-686.

Chew, Victor (1968) "Simultaneous Prediction Intervals", *Technometrics*, 10(2): 323-330.

Clements, Michael P. and David F. Hendry (1999) **Forecasting Economic Time Series**. Cambridge University Press: Cambridge, U.K.

Clements, Michael P. and Nick Taylor (2001) "Bootstrapping Prediction Intervals for Autoregressive Models", *International Journal of Forecasting,* 17: 247-267.

Danziger, L. and S. A. Davis (1964) "Tables of Distribution-free Tolerance Limits", *Annals of Mathematical Statistics*, 35: 1361-1365.

Efron, Bradley (1979) "Bootstrap Methods: Another Look at the Jackknife", *The Annals of Statistics*, 7(1): 1-26.

Holm, Sture (1979) "A simple sequentially rejective multiple test procedure", *Scandinavian Journal of Statistics,* 6: 65-70.

Jordà, Òscar (2009) "Simultaneous Confidence Regions for Impulse Responses", *The Review of Economics and Statistics*, 91(3): 629–647.

Jordà, Òscar and Massimiliano G. Marcellino (2010) "Path-Forecast Evaluation", *Journal of Applied Econometrics*, 25: 635–662.

Kim, Jae H. (1999) "Asymptotic and Bootstrap Prediction Regions for Vector Autoregression", *International Journal of Forecasting*, 15: 393-403.

Lehmann, Erich L. and Joseph P. Romano (2005) **Testing Statistical Hypotheses.** Springer: New York, NY.

Lütkepohl, Helmut, (1996) **Handbook of Matrices.** John Wiley & Sons: Chichester, U.K.

Mahalanobis, Prasanta C. (1936) "On the Generalized Distance in Statistics", *Procedures of the National Institute of Sciences of India*, 2(1): 49-55.

Masarotto, Guido (1990) "Bootstrap Prediction Intervals for Autoregressions", *International Journal of Forecasting,* 6: 229-239.

McConnell, Margaret M. and Gabriel Pérez Quirós (2000) "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?", *American Economic Review,* 90(5): 1464-1476.

Miller, Rupert G. (1981) **Simultaneous Statistical Inference, 2nd Edition.** Springer-Verlag: New York, NY.

Pascual, Lorenzo, Juan Romo and Esther Ruiz (2004) "Bootstrap Predictive Inference for ARIMA Processes", *Journal of Time Series Analysis,* 4(7): 449-465.

Politis, Dimitris N., Joseph P. Romano and Michael Wolf (1999) **Subsampling**. Springer Series in Statistics, Springer: New York, NY

Scheffé, Henry (1953) "A Method for Judging All Contrasts in the Analysis of Variance", *Biometrika*, 40(1-2): 87-110.

Staszewska-Bystrova, A. (2012) "Modified Scheffé's Prediction Bands", *mimeo.*

Stock, James H. and Mark W. Watson (2001) "Vector Autoregressions", *Journal of Economic Perspectives,* 15(4): 101-115.

Svensson, Lars E. O. (2009) "Evaluating Monetary Policy", National Bureau of Economic Research working paper 15385.

Thombs, Lori A. and William R. Schucany (1990) "Bootstrap Prediction Intervals for Autoregression", *Journal of the American Statistical Association*, 85: 486-492.

Turner, D.L. and Bowden, D.C. (1977), "Simultaneous Confidence Bands for Percentile Lines in the General Linear Model", *Journal of the American Statistical Association*, 72(360), 886-889.

Williams, W. H. and M. L. Goodman (1971) "A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts", *Journal of the American Statistical Association,* 66(336): 752-754.

Figure 1 – The Geometry of Two-Dimensional Prediction Regions
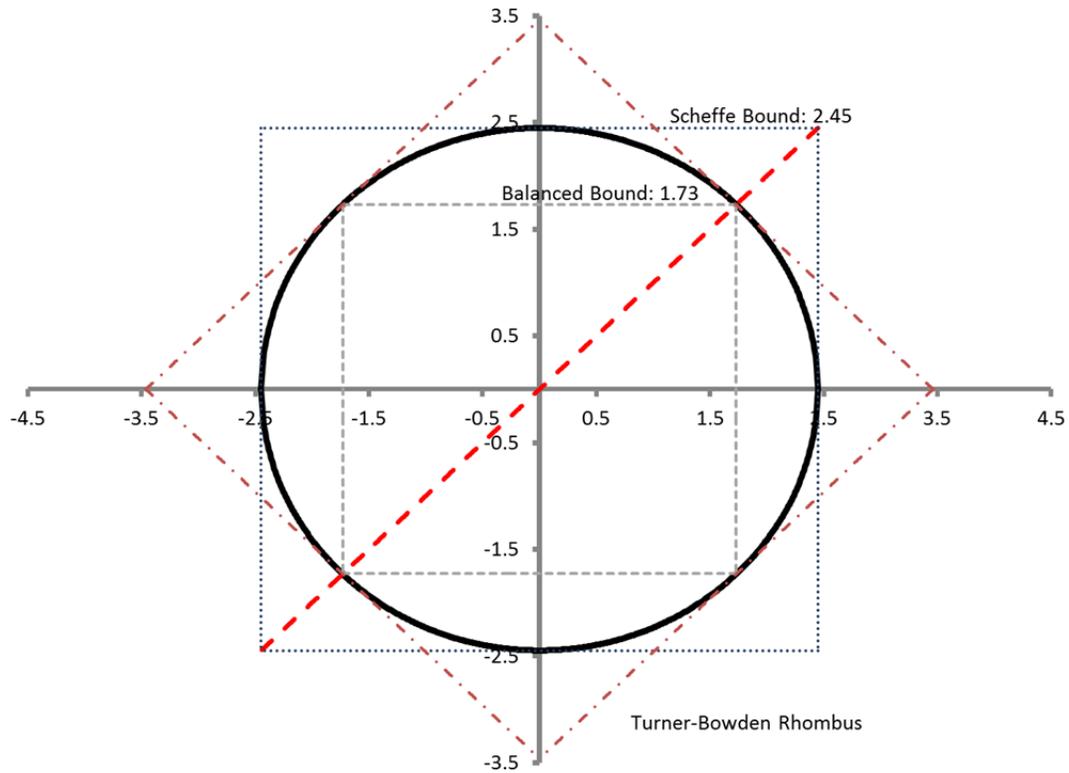
Notes: 95% chi-square circumference. Scheffé (1953) bounds are the square encasing this circumference. The Turner-Bowden Rhombus results from applying the Cauchy-Schwarz inequality. The bands that we propose are the square encircled by the 95% chi-square circumference, the result of rotating the Turner-Bowden rhombus.

Figure 2 – Width of Empirical Prediction Bands based on Greenbook Forecasts of Output and Inflation. Sample: 1974Q2 – 2003Q4.

Width of 95% Level Empirical Prediction Bands
U.S. Greenbook Output Forecasts. Sample 1974Q2 - 2003Q4

Width of 95% Level Empirical Prediction Bands
U.S. Greenbook Inflation Forecasts. Sample: 1974Q2 - 2003Q4
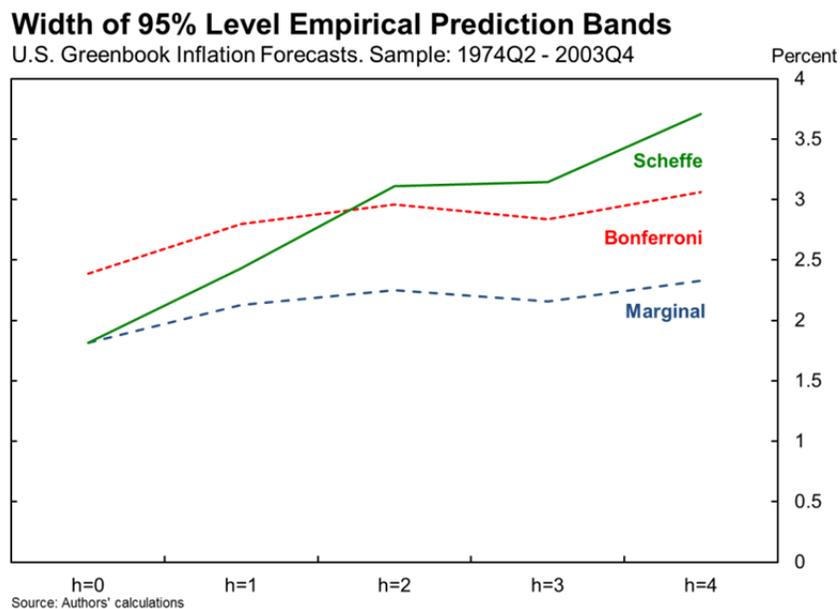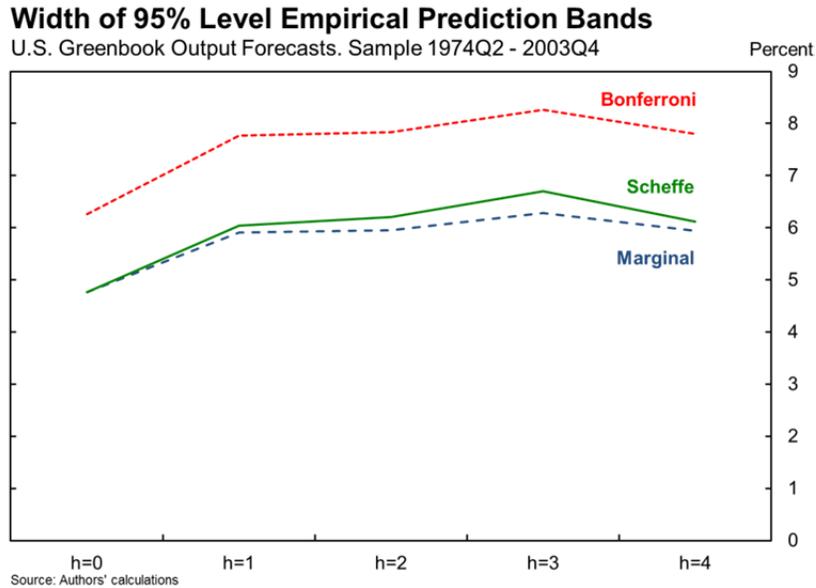
Notes: these bands show how much should be added/subtracted to a path-forecast to construct prediction bands with 95% coverage as defined in the paper (i.e., they have been re-centered at zero. The bands are constructed from the empirical sample of forecast errors using a rolling window that begins with the sample 1974Q2 to 1985Q1. The data comes from the Federal Reserve's Greenbook forecasts (now called Tealbook). Output is measured as GNP until 1991 and GDP afterwards. We report the forecasts based on the annualized quarterly growth rate. The inflation forecast is based on the GNP/GDP deflator and calculated as the annualized quarterly growth rate. For $h = 0$, it refers to a nowcast. For $h = 1, 2, 3, 4$, it refers to the h-period ahead forecast.

Table 1: Correct null model

Reference case: All parameters known
Deviations from nominal coverage levels

Metric	Null Distn.	average deviations			average absolute deviations		
		Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-26	4	-8	26	5	8
FWPER	**t**	-21	4	-5	24	6	7
	ARCH	-18	5	-3	20	7	5
	norm	-32	0	1	32	10	1
Wald	**t**	-26	3	3	28	8	4
	ARCH	-23	4	4	25	7	5

Notes: Average deviations from coverage levels over H, ρ and α. In percentage points
H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage

Bands based on model
Deviations from coverage levels obtained with all parameters known

Metric	Null D.	Estimation sample: R = 25						Estimation sample: R = 100					
		average deviations			average absolute deviations			average deviations			average absolute deviations		
		Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-4	-6	-6	4	6	6	-1	-2	-2	1	2	2
FWPER	**t**	-2	-2	-3	2	2	3	-1	0	-1	1	1	1
	ARCH	-2	-2	-3	3	2	3	-1	-1	-1	1	1	1
	norm	-1	-3	-5	4	4	5	0	-1	-1	1	1	1
Wald	**t**	2	0	-2	3	1	2	1	0	0	1	1	0
	ARCH	2	0	-2	3	1	2	1	0	0	1	0	0

Notes: Average deviations from coverage levels over H, ρ and α. In percentage points. H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage.

Bands based on forecast errors
Deviations from coverage levels obtained with all parameters known

Metric	Null D.	Estimation sample: R = 25. Evaluation sample: N = 40						Estimation sample: R = 25. Evaluation sample: N = 80					
		average deviations			average absolute deviations			average deviations			average absolute deviations		
		Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-1	-3	-5	2	3	5	-1	-2	-3	1	2	3
FWPER	**t**	-4	-4	-6	4	4	6	-2	-3	-4	2	3	4
	ARCH	-6	-6	-9	6	6	9	-4	-4	-5	4	4	5
	norm	-3	-6	-9	4	6	9	-2	-4	-5	3	4	5
Wald	**t**	-5	-8	-10	5	8	10	-3	-5	-5	3	5	5
	ARCH	-7	-10	-12	7	10	12	-5	-7	-7	5	7	7

Notes: Average deviations from coverage levels over H, ρ and α. In percentage points. H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage.

Metric	Null D.	Estimation sample: R = 100. Evaluation sample: N = 40						Estimation sample: R = 100. Evaluation sample: N = 80					
		average deviations			average absolute deviations			average deviations			average absolute deviations		
		Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-2	-3	-5	2	3	5	-1	-2	-3	1	2	3
FWPER	**t**	-4	-4	-6	4	4	6	-2	-3	-4	2	3	4
	ARCH	-6	-6	-9	6	6	9	-4	-4	-6	4	4	6
	norm	-2	-6	-9	4	6	9	-2	-3	-4	2	3	5
Wald	**t**	-4	-7	-10	5	7	10	-3	-4	-5	3	4	5
	ARCH	-7	-9	-12	7	9	12	-4	-6	-7	4	6	7

Notes: Average deviations from coverage levels over H, ρ and α. In percentage points. H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage.

Table 2: Misspecified null model

Reference case: Bands based on forecast errors. Evaluation sample: N = 1000
Deviations from nominal coverage levels

Metric	Null Distn.	*average deviations*			*average absolute deviations*		
		Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-24	5	-7	24	5	7
FWPER	**t**	-21	4	-6	22	6	7
	ARCH	-18	5	-4	19	7	6
	norm	-35	-5	0	35	15	1
Wald	**t**	-30	-2	2	31	11	4
	ARCH	-27	-1	2	28	10	5

Notes: *Average deviations from coverage levels over H, ρ and α. In percentage points*
H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage

Bands based on model
Deviations from coverage levels obtained with bands based on forecast errors and N = 1000

		Estimation sample: R = 25						Estimation sample: R = 100					
		average deviations			*average absolute deviations*			*average deviations*			*average absolute deviations*		
Metric	Null D.	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-7	-8	-30	7	8	30	-2	-2	-26	2	2	26
FWPER	**t**	-8	-9	-28	9	9	28	-3	-3	-22	3	3	22
	ARCH	-11	-11	-29	11	11	29	-4	-4	-22	4	4	22
	norm	28	15	-8	30	17	8	33	17	-2	33	18	2
Wald	**t**	23	11	-10	25	14	10	28	14	-3	29	15	3
	ARCH	19	9	-11	23	12	11	26	12	-4	27	13	4

Notes: *Average deviations from coverage levels over H, ρ and α. In percentage points. H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage.*

Bands based on forecast errors
Deviations from coverage levels obtained with bands based on forecast errors and N = 1000

		Evaluation sample: N = 40						Evaluation sample: N = 80					
		average deviations			*average absolute deviations*			*average deviations*			*average absolute deviations*		
Metric	Null D.	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé	Marginal	Bonf.	Scheffé
	norm	-4	-5	-8	4	5	8	-2	-3	-4	2	3	4
FWPER	**t**	-6	-6	-9	6	6	9	-3	-3	-5	3	3	5
	ARCH	-8	-8	-11	8	8	11	-5	-5	-6	5	5	6
	norm	-2	-4	-11	4	5	11	-1	-2	-5	2	3	5
Wald	**t**	-3	-6	-11	4	6	11	-2	-3	-5	3	3	5
	ARCH	-5	-8	-13	6	8	13	-3	-5	-7	4	5	7

Notes: *Average deviations from coverage levels over H, ρ and α. In percentage points. H is the forecast horizon, ρ is the AR parameter in the DGP and 1-α the coverage.*

Table 3: Coverage of alternative bands for US inflation and growth

nominal	Inflation						Growth					
	Coverage: 68%			Coverage: 95%			Coverage: 68%			Coverage: 95%		
	Marg	Bonf	Schef	Marg	Bonf	Schef	Marg	Bonf	Schef	Marg	Bonf	Schef
FWPER control												
actual	31	79	60	88	92	93	27	72	51	73	100	91
Wald control												
actual	16	73	75	75	88	91	40	83	73	85	99	99

Note: The Table reports the coverage rates of alternative bands for forecast paths in percent. Sample 1974:Q2-1985:Q1 is used for estimation of first variance covariance matrix of forecast errors. First forecast path used starts in 1985:Q2. Exercise is repeated in a rolling fashion. Last available forecast path comes from 2003:Q4, so that the coverage of bands for 75 paths is evaluated.

Table A.1: Actual coverage in percent with correct null model, model-based bands and known parameters

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
	norm	1	0.5	68	68	68	68	1	0.9	68	69	69	69	4	0.5	68	26	75	52	4	0.9	68	35	80	62
FWPER	t	1	0.5	68	77	77	77	1	0.9	68	76	76	76	4	0.5	68	39	78	62	4	0.9	68	48	83	70
	ARCH	1	0.5	68	75	75	75	1	0.9	68	75	75	75	4	0.5	68	45	82	66	4	0.9	68	51	84	71
	norm	1	0.5	68	68	68	68	1	0.9	68	69	69	69	4	0.5	68	27	83	67	4	0.9	68	16	64	68
Wald	t	1	0.5	68	77	77	77	1	0.9	68	76	76	76	4	0.5	68	43	85	75	4	0.9	68	30	74	76
	ARCH	1	0.5	68	75	75	75	1	0.9	68	75	75	75	4	0.5	68	49	86	78	4	0.9	68	33	76	78
	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	84	96	92	4	0.9	95	87	96	94
FWPER	t	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	83	92	89	4	0.9	95	87	95	93
	ARCH	1	0.5	95	96	96	96	1	0.9	95	95	95	95	4	0.5	95	87	94	92	4	0.9	95	88	94	93
	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	91	99	97	4	0.9	95	76	94	96
Wald	t	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	89	96	94	4	0.9	95	81	91	93
	ARCH	1	0.5	95	96	96	96	1	0.9	95	95	95	95	4	0.5	95	90	95	94	4	0.9	95	83	92	93
	norm	8	0.5	68	7	75	42	8	0.9	68	20	83	61	12	0.5	68	2	75	33	12	0.9	68	11	85	60
FWPER	t	8	0.5	68	16	73	50	8	0.9	68	31	83	69	12	0.5	68	6	70	41	12	0.9	68	20	85	69
	ARCH	8	0.5	68	22	79	56	8	0.9	68	37	85	70	12	0.5	68	11	76	49	12	0.9	68	27	85	69
	norm	8	0.5	68	9	93	67	8	0.9	68	1	50	68	12	0.5	68	3	97	68	12	0.9	68	0	35	68
Wald	t	8	0.5	68	22	89	74	8	0.9	68	5	63	74	12	0.5	68	12	92	74	12	0.9	68	0	53	74
	ARCH	8	0.5	68	29	89	77	8	0.9	68	7	68	77	12	0.5	68	19	91	77	12	0.9	68	1	60	77
	norm	8	0.5	95	71	96	87	8	0.9	95	80	97	94	12	0.5	95	60	95	81	12	0.9	95	74	97	94
FWPER	t	8	0.5	95	71	89	82	8	0.9	95	82	94	92	12	0.5	95	60	87	75	12	0.9	95	76	94	92
	ARCH	8	0.5	95	77	91	86	8	0.9	95	83	94	93	12	0.5	95	68	89	79	12	0.9	95	78	93	93
	norm	8	0.5	95	90	100	98	8	0.9	95	44	88	97	12	0.5	95	89	100	99	12	0.9	95	19	81	97
Wald	t	8	0.5	95	86	97	94	8	0.9	95	58	87	93	12	0.5	95	86	98	94	12	0.9	95	37	81	92
	ARCH	8	0.5	95	87	96	93	8	0.9	95	64	87	92	12	0.5	95	86	97	93	12	0.9	95	45	82	91

Notes: H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."

Table A.2: Actual coverage in percent with correct null model, model-based bands, R = 25

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
	norm	1	0.5	68	66	66	66	1	0.9	68	65	65	65	4	0.5	68	25	72	47	4	0.9	68	31	72	55
FWPER	t	1	0.5	68	76	76	76	1	0.9	68	76	76	76	4	0.5	68	40	78	60	4	0.9	68	45	79	67
	ARCH	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	45	82	64	4	0.9	68	48	81	68
	norm	1	0.5	68	66	66	66	1	0.9	68	65	65	65	4	0.5	68	30	79	64	4	0.9	68	17	63	63
Wald	t	1	0.5	68	76	76	76	1	0.9	68	76	76	76	4	0.5	68	45	84	74	4	0.9	68	32	73	73
	ARCH	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	52	85	77	4	0.9	68	37	76	77
	norm	1	0.5	95	93	93	93	1	0.9	95	93	93	93	4	0.5	95	79	92	85	4	0.9	95	79	91	89
FWPER	t	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	83	92	87	4	0.9	95	84	92	91
	ARCH	1	0.5	95	94	94	94	1	0.9	95	95	95	95	4	0.5	95	86	93	88	4	0.9	95	86	92	91
	norm	1	0.5	95	93	93	93	1	0.9	95	93	93	93	4	0.5	95	87	96	94	4	0.9	95	72	90	92
Wald	t	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	89	95	94	4	0.9	95	81	91	92
	ARCH	1	0.5	95	94	94	94	1	0.9	95	95	95	95	4	0.5	95	90	95	93	4	0.9	95	82	91	92
	norm	8	0.5	68	8	71	37	8	0.9	68	16	71	52	12	0.5	68	3	70	30	12	0.9	68	9	69	50
FWPER	t	8	0.5	68	18	73	46	8	0.9	68	26	78	64	12	0.5	68	8	70	39	12	0.9	68	18	76	61
	ARCH	8	0.5	68	25	78	52	8	0.9	68	30	80	65	12	0.5	68	14	77	44	12	0.9	68	22	77	63
	norm	8	0.5	68	14	86	62	8	0.9	68	2	50	60	12	0.5	68	8	88	60	12	0.9	68	0	44	59
Wald	t	8	0.5	68	28	87	72	8	0.9	68	9	65	72	12	0.5	68	19	89	71	12	0.9	68	4	57	70
	ARCH	8	0.5	68	36	87	75	8	0.9	68	14	70	75	12	0.5	68	29	89	75	12	0.9	68	6	64	74
	norm	8	0.5	95	66	91	75	8	0.9	95	68	89	86	12	0.5	95	56	91	68	12	0.9	95	59	87	84
FWPER	t	8	0.5	95	70	88	78	8	0.9	95	74	90	89	12	0.5	95	60	86	69	12	0.9	95	67	89	88
	ARCH	8	0.5	95	76	90	80	8	0.9	95	77	90	89	12	0.5	95	68	88	72	12	0.9	95	69	88	86
	norm	8	0.5	95	81	97	94	8	0.9	95	45	84	90	12	0.5	95	79	98	93	12	0.9	95	27	77	89
Wald	t	8	0.5	95	84	96	93	8	0.9	95	60	86	91	12	0.5	95	82	96	92	12	0.9	95	45	81	89
	ARCH	8	0.5	95	85	95	92	8	0.9	95	66	86	90	12	0.5	95	84	96	91	12	0.9	95	52	82	88

Notes: H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1- α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."

Table A.3: Actual coverage in percent with correct null model, model-based bands, R = 100

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	67	67	67	1	0.9	68	67	67	67	4	0.5	68	25	73	49	4	0.9	68	34	77	60
	t	1	0.5	68	76	76	76	1	0.9	68	77	77	77	4	0.5	68	39	77	60	4	0.9	68	47	82	69
	ARCH	1	0.5	68	74	74	74	1	0.9	68	75	75	75	4	0.5	68	44	82	64	4	0.9	68	52	84	71
Wald	norm	1	0.5	68	67	67	67	1	0.9	68	67	67	67	4	0.5	68	28	81	65	4	0.9	68	16	63	66
	t	1	0.5	68	76	76	76	1	0.9	68	77	77	77	4	0.5	68	44	84	75	4	0.9	68	29	74	75
	ARCH	1	0.5	68	74	74	74	1	0.9	68	75	75	75	4	0.5	68	49	86	78	4	0.9	68	35	77	78
FWPER	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	82	95	89	4	0.9	95	85	96	93
	t	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	83	92	88	4	0.9	95	87	94	93
	ARCH	1	0.5	95	96	96	96	1	0.9	95	95	95	95	4	0.5	95	86	93	90	4	0.9	95	88	94	93
Wald	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	89	98	97	4	0.9	95	76	93	95
	t	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	89	95	94	4	0.9	95	81	92	93
	ARCH	1	0.5	95	96	96	96	1	0.9	95	95	95	95	4	0.5	95	90	95	94	4	0.9	95	83	92	93
FWPER	norm	8	0.5	68	7	74	39	8	0.9	68	19	80	59	12	0.5	68	2	73	32	12	0.9	68	11	80	58
	t	8	0.5	68	16	74	50	8	0.9	68	29	82	67	12	0.5	68	8	72	42	12	0.9	68	19	82	67
	ARCH	8	0.5	68	23	80	55	8	0.9	68	34	82	68	12	0.5	68	13	77	48	12	0.9	68	23	82	68
Wald	norm	8	0.5	68	10	91	65	8	0.9	68	1	50	65	12	0.5	68	4	95	66	12	0.9	68	0	39	65
	t	8	0.5	68	24	89	75	8	0.9	68	6	64	75	12	0.5	68	16	91	74	12	0.9	68	1	55	74
	ARCH	8	0.5	68	33	89	77	8	0.9	68	10	68	76	12	0.5	68	24	90	77	12	0.9	68	2	62	76
FWPER	norm	8	0.5	95	68	94	83	8	0.9	95	77	95	92	12	0.5	95	58	94	76	12	0.9	95	69	95	92
	t	8	0.5	95	71	89	81	8	0.9	95	79	93	92	12	0.5	95	60	86	72	12	0.9	95	73	93	91
	ARCH	8	0.5	95	77	91	84	8	0.9	95	81	93	92	12	0.5	95	68	89	76	12	0.9	95	75	92	91
Wald	norm	8	0.5	95	86	100	98	8	0.9	95	44	87	95	12	0.5	95	85	100	98	12	0.9	95	22	81	96
	t	8	0.5	95	87	97	94	8	0.9	95	59	87	92	12	0.5	95	84	97	93	12	0.9	95	39	82	92
	ARCH	8	0.5	95	87	96	93	8	0.9	95	65	87	91	12	0.5	95	85	96	92	12	0.9	95	49	83	91

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*

Table A.4: Actual coverage in percent with correct null model, forecast error-based bands, R = 25, N = 40

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	67	67	67	1	0.9	68	68	68	68	4	0.5	68	25	73	49	4	0.9	68	34	77	59
	t	1	0.5	68	75	75	75	1	0.9	68	73	73	73	4	0.5	68	36	74	56	4	0.9	68	44	78	65
	ARCH	1	0.5	68	70	70	70	1	0.9	68	71	71	71	4	0.5	68	39	75	57	4	0.9	68	45	77	64
Wald	norm	1	0.5	68	67	67	67	1	0.9	68	68	68	68	4	0.5	68	26	79	62	4	0.9	68	14	58	62
	t	1	0.5	68	75	75	75	1	0.9	68	73	73	73	4	0.5	68	39	79	68	4	0.9	68	25	66	69
	ARCH	1	0.5	68	70	70	70	1	0.9	68	71	71	71	4	0.5	68	40	78	67	4	0.9	68	26	64	67
FWPER	norm	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	81	93	88	4	0.9	95	84	94	92
	t	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	79	89	84	4	0.9	95	83	92	90
	ARCH	1	0.5	95	93	93	93	1	0.9	95	93	93	93	4	0.5	95	81	90	85	4	0.9	95	82	91	89
Wald	norm	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	86	96	94	4	0.9	95	70	89	92
	t	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	84	93	90	4	0.9	95	73	86	89
	ARCH	1	0.5	95	93	93	93	1	0.9	95	93	93	93	4	0.5	95	83	91	88	4	0.9	95	71	84	86
FWPER	norm	8	0.5	68	8	72	36	8	0.9	68	21	79	58	12	0.5	68	3	71	24	12	0.9	68	13	77	55
	t	8	0.5	68	15	68	41	8	0.9	68	27	77	62	12	0.5	68	7	64	30	12	0.9	68	19	76	60
	ARCH	8	0.5	68	20	70	43	8	0.9	68	30	76	61	12	0.5	68	12	67	33	12	0.9	68	22	74	60
Wald	norm	8	0.5	68	13	83	54	8	0.9	68	1	40	54	12	0.5	68	9	83	44	12	0.9	68	0	27	43
	t	8	0.5	68	22	80	60	8	0.9	68	5	49	59	12	0.5	68	16	80	51	12	0.9	68	1	37	51
	ARCH	8	0.5	68	25	78	59	8	0.9	68	6	50	59	12	0.5	68	20	77	51	12	0.9	68	2	39	51
FWPER	norm	8	0.5	95	67	92	74	8	0.9	95	75	93	91	12	0.5	95	56	92	57	12	0.9	95	68	92	89
	t	8	0.5	95	65	85	70	8	0.9	95	74	90	88	12	0.5	95	53	82	55	12	0.9	95	67	88	87
	ARCH	8	0.5	95	67	85	70	8	0.9	95	73	88	86	12	0.5	95	57	82	55	12	0.9	95	66	86	85
Wald	norm	8	0.5	95	78	96	91	8	0.9	95	35	76	86	12	0.5	95	72	96	85	12	0.9	95	15	61	78
	t	8	0.5	95	77	93	86	8	0.9	95	44	74	81	12	0.5	95	71	92	80	12	0.9	95	25	62	74
	ARCH	8	0.5	95	75	90	83	8	0.9	95	45	72	79	12	0.5	95	69	89	77	12	0.9	95	28	61	72

Notes: H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."

Table A.5: Actual coverage in percent with correct null model, forecast error-based bands, R = 25, N = 80

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	67	67	67	1	0.9	68	67	67	67	4	0.5	68	26	73	50	4	0.9	68	36	78	61
	t	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	37	76	59	4	0.9	68	46	80	67
	ARCH	1	0.5	68	72	72	72	1	0.9	68	71	71	71	4	0.5	68	41	77	60	4	0.9	68	47	80	67
Wald	norm	1	0.5	68	67	67	67	1	0.9	68	67	67	67	4	0.5	68	26	80	64	4	0.9	68	15	60	65
	t	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	41	81	71	4	0.9	68	26	68	71
	ARCH	1	0.5	68	72	72	72	1	0.9	68	71	71	71	4	0.5	68	43	80	71	4	0.9	68	28	68	71
FWPER	norm	1	0.5	95	94	94	94	1	0.9	95	95	95	95	4	0.5	95	82	94	89	4	0.9	95	85	95	93
	t	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	81	91	87	4	0.9	95	85	92	91
	ARCH	1	0.5	95	93	93	93	1	0.9	95	94	94	94	4	0.5	95	84	91	88	4	0.9	95	85	93	91
Wald	norm	1	0.5	95	94	94	94	1	0.9	95	95	95	95	4	0.5	95	88	98	96	4	0.9	95	71	91	94
	t	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	86	94	92	4	0.9	95	76	88	90
	ARCH	1	0.5	95	93	93	93	1	0.9	95	94	94	94	4	0.5	95	86	93	91	4	0.9	95	76	88	90
FWPER	norm	8	0.5	68	7	73	39	8	0.9	68	21	82	60	12	0.5	68	2	73	28	12	0.9	68	13	81	59
	t	8	0.5	68	16	71	45	8	0.9	68	29	79	65	12	0.5	68	7	67	34	12	0.9	68	20	79	63
	ARCH	8	0.5	68	20	74	48	8	0.9	68	32	80	65	12	0.5	68	11	69	37	12	0.9	68	23	78	63
Wald	norm	8	0.5	68	10	86	61	8	0.9	68	1	43	61	12	0.5	68	6	89	54	12	0.9	68	0	29	55
	t	8	0.5	68	22	83	66	8	0.9	68	4	53	65	12	0.5	68	16	85	61	12	0.9	68	1	42	61
	ARCH	8	0.5	68	26	82	67	8	0.9	68	7	56	67	12	0.5	68	19	82	60	12	0.9	68	2	45	62
FWPER	norm	8	0.5	95	68	94	79	8	0.9	95	77	95	92	12	0.5	95	57	94	66	12	0.9	95	72	95	92
	t	8	0.5	95	66	87	76	8	0.9	95	77	92	90	12	0.5	95	56	83	63	12	0.9	95	71	91	90
	ARCH	8	0.5	95	71	88	77	8	0.9	95	77	90	89	12	0.5	95	60	84	65	12	0.9	95	72	89	88
Wald	norm	8	0.5	95	83	99	95	8	0.9	95	36	80	92	12	0.5	95	79	99	94	12	0.9	95	16	68	89
	t	8	0.5	95	81	95	90	8	0.9	95	49	79	87	12	0.5	95	77	95	87	12	0.9	95	29	69	83
	ARCH	8	0.5	95	80	93	88	8	0.9	95	51	77	85	12	0.5	95	75	93	85	12	0.9	95	32	69	81

Notes: H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."

Table A.6: Actual coverage in percent with correct null model, forecast error-based bands, R = 100, N = 40

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
	norm	1	0.5	68	68	68	68	1	0.9	68	68	68	68	4	0.5	68	26	72	49	4	0.9	68	35	76	59
FWPER	t	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	36	73	56	4	0.9	68	44	79	66
	ARCH	1	0.5	68	71	71	71	1	0.9	68	70	70	70	4	0.5	68	39	75	57	4	0.9	68	45	77	64
	norm	1	0.5	68	68	68	68	1	0.9	68	68	68	68	4	0.5	68	27	78	62	4	0.9	68	15	60	63
Wald	t	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	40	79	68	4	0.9	68	26	67	69
	ARCH	1	0.5	68	71	71	71	1	0.9	68	70	70	70	4	0.5	68	41	78	67	4	0.9	68	27	66	68
	norm	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	81	94	87	4	0.9	95	83	94	92
FWPER	t	1	0.5	95	93	93	93	1	0.9	95	94	94	94	4	0.5	95	80	89	84	4	0.9	95	83	92	90
	ARCH	1	0.5	95	93	93	93	1	0.9	95	93	93	93	4	0.5	95	80	89	84	4	0.9	95	83	91	89
	norm	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	87	97	94	4	0.9	95	71	90	92
Wald	t	1	0.5	95	93	93	93	1	0.9	95	94	94	94	4	0.5	95	85	93	91	4	0.9	95	74	87	89
	ARCH	1	0.5	95	93	93	93	1	0.9	95	93	93	93	4	0.5	95	83	91	88	4	0.9	95	73	85	87
	norm	8	0.5	68	8	72	34	8	0.9	68	19	78	57	12	0.5	68	3	71	25	12	0.9	68	12	77	56
FWPER	t	8	0.5	68	15	68	40	8	0.9	68	27	76	62	12	0.5	68	7	64	29	12	0.9	68	18	76	61
	ARCH	8	0.5	68	20	70	42	8	0.9	68	30	76	61	12	0.5	68	11	66	32	12	0.9	68	21	73	59
	norm	8	0.5	68	13	82	53	8	0.9	68	1	41	53	12	0.5	68	10	83	43	12	0.9	68	0	28	44
Wald	t	8	0.5	68	23	80	59	8	0.9	68	5	51	59	12	0.5	68	17	80	51	12	0.9	68	1	38	51
	ARCH	8	0.5	68	26	78	59	8	0.9	68	7	52	60	12	0.5	68	20	78	51	12	0.9	68	2	40	50
	norm	8	0.5	95	66	93	73	8	0.9	95	74	93	91	12	0.5	95	55	92	57	12	0.9	95	67	92	89
FWPER	t	8	0.5	95	64	85	69	8	0.9	95	74	89	89	12	0.5	95	53	82	54	12	0.9	95	67	88	87
	ARCH	8	0.5	95	67	86	70	8	0.9	95	73	87	86	12	0.5	95	57	81	54	12	0.9	95	66	86	84
	norm	8	0.5	95	79	97	91	8	0.9	95	37	77	87	12	0.5	95	72	96	85	12	0.9	95	16	64	79
Wald	t	8	0.5	95	77	93	86	8	0.9	95	46	75	82	12	0.5	95	71	92	80	12	0.9	95	27	65	75
	ARCH	8	0.5	95	75	91	83	8	0.9	95	47	74	80	12	0.5	95	70	90	77	12	0.9	95	29	63	72

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*

Table A.7: Actual coverage in percent with correct null model, forecast error-based bands, R = 100, N = 80

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	68	68	68	1	0.9	68	68	68	68	4	0.5	68	26	73	50	4	0.9	68	35	78	61
	t	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	37	75	59	4	0.9	68	45	79	67
	ARCH	1	0.5	68	72	72	72	1	0.9	68	72	72	72	4	0.5	68	41	78	60	4	0.9	68	47	80	67
Wald	norm	1	0.5	68	68	68	68	1	0.9	68	68	68	68	4	0.5	68	27	79	64	4	0.9	68	15	62	66
	t	1	0.5	68	75	75	75	1	0.9	68	74	74	74	4	0.5	68	40	82	71	4	0.9	68	26	69	71
	ARCH	1	0.5	68	72	72	72	1	0.9	68	72	72	72	4	0.5	68	44	81	71	4	0.9	68	29	69	71
FWPER	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	82	94	89	4	0.9	95	86	96	93
	t	1	0.5	95	94	94	94	1	0.9	95	95	95	95	4	0.5	95	80	90	86	4	0.9	95	85	93	91
	ARCH	1	0.5	95	93	93	93	1	0.9	95	94	94	94	4	0.5	95	82	92	88	4	0.9	95	85	92	91
Wald	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	88	98	96	4	0.9	95	73	92	95
	t	1	0.5	95	94	94	94	1	0.9	95	95	95	95	4	0.5	95	87	94	92	4	0.9	95	77	89	91
	ARCH	1	0.5	95	93	93	93	1	0.9	95	94	94	94	4	0.5	95	86	93	91	4	0.9	95	77	88	90
FWPER	norm	8	0.5	68	8	74	37	8	0.9	68	20	80	59	12	0.5	68	2	72	28	12	0.9	68	12	81	57
	t	8	0.5	68	16	71	45	8	0.9	68	29	80	65	12	0.5	68	7	67	33	12	0.9	68	18	78	64
	ARCH	8	0.5	68	21	74	48	8	0.9	68	31	79	64	12	0.5	68	11	70	37	12	0.9	68	22	78	63
Wald	norm	8	0.5	68	11	88	60	8	0.9	68	1	45	60	12	0.5	68	6	89	54	12	0.9	68	0	31	55
	t	8	0.5	68	23	84	67	8	0.9	68	5	56	66	12	0.5	68	15	85	62	12	0.9	68	1	43	61
	ARCH	8	0.5	68	27	82	67	8	0.9	68	8	57	67	12	0.5	68	20	83	62	12	0.9	68	2	46	61
FWPER	norm	8	0.5	95	68	94	80	8	0.9	95	77	95	93	12	0.5	95	58	94	67	12	0.9	95	70	94	92
	t	8	0.5	95	66	86	75	8	0.9	95	77	92	90	12	0.5	95	56	84	63	12	0.9	95	70	90	89
	ARCH	8	0.5	95	71	88	77	8	0.9	95	77	90	89	12	0.5	95	60	84	65	12	0.9	95	70	89	88
Wald	norm	8	0.5	95	83	99	95	8	0.9	95	39	82	92	12	0.5	95	79	99	94	12	0.9	95	17	71	90
	t	8	0.5	95	81	95	90	8	0.9	95	51	80	88	12	0.5	95	78	95	88	12	0.9	95	30	71	84
	ARCH	8	0.5	95	81	93	88	8	0.9	95	53	79	85	12	0.5	95	77	94	86	12	0.9	95	35	70	82

Notes: H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."

Table A.8: Actual coverage in percent with misspecified null model, bands based on forecast errors, N = 1000

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	68	68	68	1	0.9	68	67	67	67	4	0.5	68	26	74	53	4	0.9	68	47	84	65
	t	1	0.5	68	74	74	74	1	0.9	68	70	70	70	4	0.5	68	38	77	60	4	0.9	68	51	85	66
	ARCH	1	0.5	68	76	76	76	1	0.9	68	73	73	73	4	0.5	68	45	81	66	4	0.9	68	55	85	70
Wald	norm	1	0.5	68	68	68	68	1	0.9	68	67	67	67	4	0.5	68	26	81	66	4	0.9	68	12	54	71
	t	1	0.5	68	74	74	74	1	0.9	68	70	70	70	4	0.5	68	41	83	74	4	0.9	68	19	63	75
	ARCH	1	0.5	68	76	76	76	1	0.9	68	73	73	73	4	0.5	68	47	84	77	4	0.9	68	24	67	78
FWPER	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	83	95	91	4	0.9	95	89	97	94
	t	1	0.5	95	95	95	95	1	0.9	95	94	94	94	4	0.5	95	84	93	90	4	0.9	95	89	96	93
	ARCH	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	86	93	91	4	0.9	95	90	96	93
Wald	norm	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	89	98	97	4	0.9	95	65	88	93
	t	1	0.5	95	95	95	95	1	0.9	95	94	94	94	4	0.5	95	89	95	94	4	0.9	95	72	87	91
	ARCH	1	0.5	95	95	95	95	1	0.9	95	95	95	95	4	0.5	95	89	95	94	4	0.9	95	75	88	91
FWPER	norm	8	0.5	68	7	75	42	8	0.9	68	30	86	65	12	0.5	68	2	75	34	12	0.9	68	19	87	64
	t	8	0.5	68	15	74	49	8	0.9	68	35	86	66	12	0.5	68	6	71	40	12	0.9	68	24	86	66
	ARCH	8	0.5	68	21	78	55	8	0.9	68	40	86	69	12	0.5	68	11	75	48	12	0.9	68	29	85	69
Wald	norm	8	0.5	68	8	91	67	8	0.9	68	0	33	69	12	0.5	68	3	96	67	12	0.9	68	0	20	68
	t	8	0.5	68	21	88	73	8	0.9	68	2	48	74	12	0.5	68	12	91	73	12	0.9	68	0	37	73
	ARCH	8	0.5	68	29	88	76	8	0.9	68	4	54	76	12	0.5	68	19	90	74	12	0.9	68	1	44	74
FWPER	norm	8	0.5	95	70	95	86	8	0.9	95	83	97	94	12	0.5	95	59	96	80	12	0.9	95	78	97	94
	t	8	0.5	95	71	89	82	8	0.9	95	84	96	93	12	0.5	95	59	87	74	12	0.9	95	79	95	93
	ARCH	8	0.5	95	75	91	85	8	0.9	95	85	95	93	12	0.5	95	67	88	78	12	0.9	95	79	94	92
Wald	norm	8	0.5	95	87	100	98	8	0.9	95	27	75	94	12	0.5	95	87	100	99	12	0.9	95	9	62	95
	t	8	0.5	95	86	97	94	8	0.9	95	43	78	90	12	0.5	95	84	98	94	12	0.9	95	22	70	90
	ARCH	8	0.5	95	86	96	93	8	0.9	95	49	79	90	12	0.5	95	84	96	92	12	0.9	95	30	73	89

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1- α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*

Table A.9: Actual coverage in percent with misspecified null model, bands based on model, R = 25

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	67	67	67	1	0.9	68	58	58	58	4	0.5	68	25	70	29	4	0.9	68	33	70	36
	t	1	0.5	68	70	70	70	1	0.9	68	59	59	59	4	0.5	68	34	71	37	4	0.9	68	36	71	39
	ARCH	1	0.5	68	70	70	70	1	0.9	68	59	59	59	4	0.5	68	37	72	41	4	0.9	68	37	72	40
Wald	norm	1	0.5	68	67	67	67	1	0.9	68	58	58	58	4	0.5	68	58	94	63	4	0.9	68	55	86	58
	t	1	0.5	68	70	70	70	1	0.9	68	59	59	59	4	0.5	68	63	91	67	4	0.9	68	55	85	59
	ARCH	1	0.5	68	70	70	70	1	0.9	68	59	59	59	4	0.5	68	63	90	67	4	0.9	68	56	85	59
FWPER	norm	1	0.5	95	94	94	94	1	0.9	95	90	90	90	4	0.5	95	79	92	67	4	0.9	95	78	90	68
	t	1	0.5	95	93	93	93	1	0.9	95	91	91	91	4	0.5	95	78	88	70	4	0.9	95	77	89	69
	ARCH	1	0.5	95	93	93	93	1	0.9	95	89	89	89	4	0.5	95	77	87	69	4	0.9	95	77	88	69
Wald	norm	1	0.5	95	94	94	94	1	0.9	95	90	90	90	4	0.5	95	97	99	94	4	0.9	95	90	96	85
	t	1	0.5	95	93	93	93	1	0.9	95	91	91	91	4	0.5	95	94	98	90	4	0.9	95	89	95	84
	ARCH	1	0.5	95	93	93	93	1	0.9	95	89	89	89	4	0.5	95	92	96	89	4	0.9	95	89	95	84
FWPER	norm	8	0.5	68	8	69	11	8	0.9	68	19	69	22	12	0.5	68	3	69	5	12	0.9	68	12	68	15
	t	8	0.5	68	15	65	18	8	0.9	68	22	70	25	12	0.5	68	7	60	9	12	0.9	68	15	66	17
	ARCH	8	0.5	68	18	67	22	8	0.9	68	23	68	26	12	0.5	68	10	62	13	12	0.9	68	16	65	19
Wald	norm	8	0.5	68	54	99	62	8	0.9	68	50	91	55	12	0.5	68	52	100	62	12	0.9	68	48	93	54
	t	8	0.5	68	57	96	63	8	0.9	68	52	90	56	12	0.5	68	54	97	60	12	0.9	68	48	91	52
	ARCH	8	0.5	68	58	94	64	8	0.9	68	52	89	56	12	0.5	68	55	96	61	12	0.9	68	49	89	54
FWPER	norm	8	0.5	95	64	91	39	8	0.9	95	66	87	48	12	0.5	95	54	90	22	12	0.9	95	58	84	35
	t	8	0.5	95	62	83	43	8	0.9	95	66	85	50	12	0.5	95	50	79	27	12	0.9	95	59	82	38
	ARCH	8	0.5	95	63	83	46	8	0.9	95	66	84	50	12	0.5	95	54	79	33	12	0.9	95	58	80	39
Wald	norm	8	0.5	95	98	100	93	8	0.9	95	90	98	81	12	0.5	95	99	100	92	12	0.9	95	89	98	77
	t	8	0.5	95	95	99	88	8	0.9	95	88	96	79	12	0.5	95	95	99	86	12	0.9	95	87	97	75
	ARCH	8	0.5	95	93	98	86	8	0.9	95	88	95	79	12	0.5	95	93	99	84	12	0.9	95	86	96	75

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1- α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*

Table A.10: Actual coverage in percent with misspecified null model, bands based on model, R = 100

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	68	68	68	1	0.9	68	65	65	65	4	0.5	68	26	73	30	4	0.9	68	42	79	45
	t	1	0.5	68	73	73	73	1	0.9	68	67	67	67	4	0.5	68	37	76	41	4	0.9	68	47	81	50
	ARCH	1	0.5	68	73	73	73	1	0.9	68	68	68	68	4	0.5	68	40	77	44	4	0.9	68	48	81	51
Wald	norm	1	0.5	68	68	68	68	1	0.9	68	65	65	65	4	0.5	68	61	96	67	4	0.9	68	63	91	66
	t	1	0.5	68	73	73	73	1	0.9	68	67	67	67	4	0.5	68	68	94	72	4	0.9	68	64	91	67
	ARCH	1	0.5	68	73	73	73	1	0.9	68	68	68	68	4	0.5	68	69	92	73	4	0.9	68	66	90	69
FWPER	norm	1	0.5	95	95	95	95	1	0.9	95	93	93	93	4	0.5	95	82	94	70	4	0.9	95	85	95	77
	t	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	82	91	73	4	0.9	95	86	94	78
	ARCH	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	83	91	76	4	0.9	95	86	93	79
Wald	norm	1	0.5	95	95	95	95	1	0.9	95	93	93	93	4	0.5	95	98	100	96	4	0.9	95	94	98	90
	t	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	96	99	93	4	0.9	95	93	98	90
	ARCH	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	95	98	92	4	0.9	95	93	97	90
FWPER	norm	8	0.5	68	8	73	11	8	0.9	68	26	82	30	12	0.5	68	2	74	4	12	0.9	68	17	81	20
	t	8	0.5	68	15	72	19	8	0.9	68	30	82	34	12	0.5	68	7	67	9	12	0.9	68	21	81	25
	ARCH	8	0.5	68	20	74	24	8	0.9	68	33	80	37	12	0.5	68	11	71	15	12	0.9	68	25	79	28
Wald	norm	8	0.5	68	58	100	66	8	0.9	68	61	96	66	12	0.5	68	56	100	67	12	0.9	68	60	98	66
	t	8	0.5	68	64	97	70	8	0.9	68	63	95	67	12	0.5	68	61	99	69	12	0.9	68	61	97	66
	ARCH	8	0.5	68	66	96	71	8	0.9	68	64	94	68	12	0.5	68	65	97	71	12	0.9	68	63	95	67
FWPER	norm	8	0.5	95	68	95	41	8	0.9	95	78	94	59	12	0.5	95	57	94	22	12	0.9	95	72	94	47
	t	8	0.5	95	68	87	49	8	0.9	95	78	93	62	12	0.5	95	56	84	30	12	0.9	95	73	93	50
	ARCH	8	0.5	95	72	89	54	8	0.9	95	78	91	64	12	0.5	95	62	86	38	12	0.9	95	71	90	51
Wald	norm	8	0.5	95	99	100	96	8	0.9	95	95	99	88	12	0.5	95	100	100	96	12	0.9	95	95	100	87
	t	8	0.5	95	97	99	92	8	0.9	95	94	99	87	12	0.5	95	97	100	91	12	0.9	95	95	99	86
	ARCH	8	0.5	95	96	99	92	8	0.9	95	93	98	87	12	0.5	95	96	99	90	12	0.9	95	93	98	85

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1- α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*

Table A.11: Actual coverage in percent with misspecified null model, bands based on forecast errors, N = 40

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
	norm	1	0.5	68	67	67	67	1	0.9	68	61	61	61	4	0.5	68	26	71	49	4	0.9	68	38	75	57
FWPER	t	1	0.5	68	73	73	73	1	0.9	68	62	62	62	4	0.5	68	36	72	54	4	0.9	68	41	75	57
	ARCH	1	0.5	68	71	71	71	1	0.9	68	63	63	63	4	0.5	68	39	75	57	4	0.9	68	43	75	58
	norm	1	0.5	68	67	67	67	1	0.9	68	61	61	61	4	0.5	68	27	77	61	4	0.9	68	11	50	61
Wald	t	1	0.5	68	73	73	73	1	0.9	68	62	62	62	4	0.5	68	38	78	67	4	0.9	68	17	55	64
	ARCH	1	0.5	68	71	71	71	1	0.9	68	63	63	63	4	0.5	68	42	78	67	4	0.9	68	19	56	64
	norm	1	0.5	95	94	94	94	1	0.9	95	92	92	92	4	0.5	95	81	93	86	4	0.9	95	81	92	88
FWPER	t	1	0.5	95	94	94	94	1	0.9	95	92	92	92	4	0.5	95	80	90	84	4	0.9	95	81	91	87
	ARCH	1	0.5	95	93	93	93	1	0.9	95	92	92	92	4	0.5	95	80	89	83	4	0.9	95	80	89	87
	norm	1	0.5	95	94	94	94	1	0.9	95	92	92	92	4	0.5	95	86	96	94	4	0.9	95	60	82	87
Wald	t	1	0.5	95	94	94	94	1	0.9	95	92	92	92	4	0.5	95	85	93	90	4	0.9	95	65	81	85
	ARCH	1	0.5	95	93	93	93	1	0.9	95	92	92	92	4	0.5	95	83	91	89	4	0.9	95	65	80	84
	norm	8	0.5	68	8	71	34	8	0.9	68	24	75	55	12	0.5	68	3	71	25	12	0.9	68	17	74	54
FWPER	t	8	0.5	68	15	68	38	8	0.9	68	27	74	54	12	0.5	68	7	63	29	12	0.9	68	20	73	53
	ARCH	8	0.5	68	19	70	42	8	0.9	68	29	73	55	12	0.5	68	11	66	31	12	0.9	68	22	71	52
	norm	8	0.5	68	13	82	53	8	0.9	68	1	33	52	12	0.5	68	11	82	43	12	0.9	68	0	24	43
Wald	t	8	0.5	68	23	80	59	8	0.9	68	3	40	56	12	0.5	68	18	79	50	12	0.9	68	1	33	48
	ARCH	8	0.5	68	26	78	58	8	0.9	68	5	43	56	12	0.5	68	21	78	51	12	0.9	68	2	34	48
	norm	8	0.5	95	66	92	71	8	0.9	95	72	90	85	12	0.5	95	56	92	56	12	0.9	95	66	88	83
FWPER	t	8	0.5	95	64	85	68	8	0.9	95	72	88	84	12	0.5	95	52	81	53	12	0.9	95	64	86	81
	ARCH	8	0.5	95	67	85	69	8	0.9	95	70	86	82	12	0.5	95	57	81	54	12	0.9	95	63	83	79
	norm	8	0.5	95	78	96	91	8	0.9	95	28	67	82	12	0.5	95	72	95	86	12	0.9	95	14	54	74
Wald	t	8	0.5	95	76	92	85	8	0.9	95	37	67	78	12	0.5	95	71	92	80	12	0.9	95	22	57	71
	ARCH	8	0.5	95	76	91	83	8	0.9	95	38	66	76	12	0.5	95	70	90	78	12	0.9	95	25	56	68

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1-α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*

Table A.12: Actual coverage in percent with misspecified null model, bands based on forecast errors, N = 80

		H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S	H	ρ	1-α	M	B	S
FWPER	norm	1	0.5	68	67	67	67	1	0.9	68	64	64	64	4	0.5	68	25	74	50	4	0.9	68	41	79	60
	t	1	0.5	68	73	73	73	1	0.9	68	66	66	66	4	0.5	68	36	75	57	4	0.9	68	45	80	62
	ARCH	1	0.5	68	73	73	73	1	0.9	68	67	67	67	4	0.5	68	41	77	60	4	0.9	68	47	79	62
Wald	norm	1	0.5	68	67	67	67	1	0.9	68	64	64	64	4	0.5	68	26	79	64	4	0.9	68	11	50	65
	t	1	0.5	68	73	73	73	1	0.9	68	66	66	66	4	0.5	68	38	81	70	4	0.9	68	18	58	70
	ARCH	1	0.5	68	73	73	73	1	0.9	68	67	67	67	4	0.5	68	42	81	71	4	0.9	68	21	59	70
FWPER	norm	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	82	95	89	4	0.9	95	85	95	92
	t	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	81	90	86	4	0.9	95	85	93	90
	ARCH	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	82	91	87	4	0.9	95	84	92	89
Wald	norm	1	0.5	95	94	94	94	1	0.9	95	94	94	94	4	0.5	95	88	98	96	4	0.9	95	63	85	90
	t	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	86	94	92	4	0.9	95	67	83	88
	ARCH	1	0.5	95	94	94	94	1	0.9	95	93	93	93	4	0.5	95	86	93	91	4	0.9	95	68	83	86
FWPER	norm	8	0.5	68	7	73	37	8	0.9	68	27	80	61	12	0.5	68	2	72	28	12	0.9	68	17	80	58
	t	8	0.5	68	16	71	44	8	0.9	68	31	80	60	12	0.5	68	7	66	34	12	0.9	68	22	79	60
	ARCH	8	0.5	68	21	74	48	8	0.9	68	33	78	61	12	0.5	68	11	70	38	12	0.9	68	25	78	60
Wald	norm	8	0.5	68	11	86	60	8	0.9	68	1	33	60	12	0.5	68	7	88	54	12	0.9	68	0	23	55
	t	8	0.5	68	22	84	66	8	0.9	68	2	45	65	12	0.5	68	16	85	62	12	0.9	68	1	34	60
	ARCH	8	0.5	68	28	82	67	8	0.9	68	4	47	65	12	0.5	68	20	84	62	12	0.9	68	1	38	61
FWPER	norm	8	0.5	95	68	94	79	8	0.9	95	77	93	90	12	0.5	95	57	94	65	12	0.9	95	71	93	90
	t	8	0.5	95	67	87	75	8	0.9	95	77	92	89	12	0.5	95	56	84	62	12	0.9	95	71	91	88
	ARCH	8	0.5	95	70	87	76	8	0.9	95	76	90	87	12	0.5	95	60	83	63	12	0.9	95	70	88	86
Wald	norm	8	0.5	95	82	98	95	8	0.9	95	28	70	88	12	0.5	95	78	98	94	12	0.9	95	12	57	86
	t	8	0.5	95	81	95	91	8	0.9	95	40	73	85	12	0.5	95	76	95	88	12	0.9	95	24	64	82
	ARCH	8	0.5	95	80	93	88	8	0.9	95	42	71	82	12	0.5	95	75	92	84	12	0.9	95	27	63	79

Notes: *H refers to the forecast horizon, ρ refers to the autoregressive parameter of the DGP, 1- α refers to the coverage, M stands for "marginal," B stands for "Bonferroni", and S stands for "Scheffé."*