Bootstrap Joint Prediction Regions

Michael Wolf
Department Economics
University of Zurich
8032 Zurich, Switzerland
michael.wolf@econ.uzh.ch

Dan Wunderli
Swiss National Bank
Bundesplatz 1
3003 Bern, Switzerland
dan.wunderli@gmail.com

Abstract

Many economic and financial applications require the forecast of a random variable of interest over several periods into the future. The sequence of individual forecasts, one period at a time, is called a path forecast, where the term *path* refers to the sequence of individual future realizations of the random variable. The problem of constructing a corresponding joint prediction region has been rather neglected in the literature so far: such a region is supposed to contain the entire future path with a prespecified probability. We develop a bootstrap method to construct such a joint prediction region. The resulting region is proven to be asymptotically consistent under a mild high-level assumption. It also has better finite-sample performance than previous proposals in the literature.

KEY WORDS: Generalized error rates; path forecast; simultaneous prediction intervals.

1 Introduction

When predicting a random variable, a point forecast alone is often considered insufficient. In addition, a statement about the uncertainty contained in the point forecast, as expressed by a *prediction interval*, may also be desired.

This is similar to the situation where a point estimator of a population parameter alone is considered insufficient; and where a statement about the uncertainty contained in the point estimate, as expressed by a *confidence interval*, is also desired.

Constructing a prediction interval for a random variable is inherently more difficult than constructing a confidence interval for a population parameter.

The discussion up to this point only applies to a single (future) random variable. In many applications, however, a random variable of interest is predicted up to H periods into the future. For example, one might predict future inflation for the next $H = 12$ months. A *path* refers to the sequence of future realizations 1 to H periods into the future. A *path forecast* refers to the sequence of corresponding forecasts 1 to H periods into the future.

On the one hand, one can construct H marginal prediction intervals by using a given method to construct a prediction interval repeatedly, one period at a time. But, by design, probability statements then only apply marginally, one period at a time: the prediction interval at a specific horizon h, for some $1 \leq h \leq H$, will contain the random variable h periods into the future with prespecified probability $1 - \alpha$.

On the other hand, a more general problem is the construction of a *joint prediction region* that will contain the entire future path with the desired probability $1 - \alpha$. For example, if one would like to know with probability $1 - \alpha$ how high inflation might rise over the next $H = 12$ months, one needs to construct a joint prediction region for the future path at level $1 - \alpha$ as opposed to stringing together 12 marginal prediction intervals, each one at level $1 - \alpha$.

Using Bonferroni's inequality, a conservative joint prediction region can be constructed by stringing together marginal prediction intervals at level $1 - \alpha/H$ instead of at level $1 - \alpha$; ; for example, such an approach is already mentioned by Lütkepohl (1991, Section 2.2.3). But since Bonferroni's inequality is generally crude, such an approach results in joint confidence regions whose coverage probability is generally (much) above $1 - \alpha$ and that are thus unnecessarily wide, leading to a loss of information.

The construction of joint prediction regions for future paths of a random variable of interest with coverage probability 'near' $1 - \alpha$ has been rather neglected in the forecasting literature so far. Two notable exceptions are Jordà and Marcellino (2010) and Staszewska-Bystrova (2011). The former work proposes an 'asymptotic' method that relies on the overly strong assumption that forecast errors have, approximately, a normal distribution. The latter work proposes a bootstrap method that is of heuristic nature only. Therefore, neither of the proposed methods appears entirely safe to use in practice.

In this paper, we propose a bootstrap method to construct joint predictions regions that are proven to contain future paths of a random variable of interest with probability $1 - \alpha$, at least asymptotically, under a mild high-level assumption.

2 Path Forecast

Consider a random vector $Y = (y_1, \ldots, y_H)'$ of interest with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_H)' = \mathbb{E}(Y)$. For the purposes of this paper, Y will typically correspond to the values of a random variable 1 to H periods into the future; that is, to a future *path* of a random variable. But the discussion below applies to any random vector. The underlying probability mechanism is denoted by $\mathbb{P}$.

One can wish to predict Y or to estimate $\boldsymbol{\mu}$. Denote the forecast of Y by $\widehat{Y}$ and the estimator of $\boldsymbol{\mu}$ by $\widehat{\boldsymbol{\mu}}$. (When Y corresponds to a future path of a random variable, $\widehat{Y}$ is also called a *path forecast*.) Often, the two are actually the same, that is, $\widehat{Y} = \widehat{\boldsymbol{\mu}}$; for example, in the context of linear regression models under quadratic loss. Therefore, in terms of a (point) forecast of Y compared to a (point) estimate of $\boldsymbol{\mu}$, there often is no difference at all.

What if one desires the extension of an 'uncertainty interval' for a univariate quantity to a '(joint) uncertainty region' for a multivariate quantity? Such a region should contain the *entire* random vector Y or its mean $\boldsymbol{\mu}$, respectively, with a prespecified probability $1 - \alpha$. The two solutions are fundamentally different and the former region will have to be larger (in volume) due to the additional randomness contained in Y compared to its mean $\boldsymbol{\mu}$.

A potential complication with joint regions arises when uncertainty statements concerning the individual components y_h or μ_h, respectively, are desired. For example, this is typically the case when a joint prediction region for Y is to be constructed in addition to a path forecast $\widehat{Y}$. One desires lower and upper bounds for each component y_h in such a manner that the entire vector Y be contained in the implied rectangle with probability $1 - \alpha$. This is a trivial task if the underlying joint prediction region is already of rectangular form. Therefore, our goal is to construct joint confidence regions of rectangular form.

The joint regions discussed so far control the probability of containing the entire vector of interest to be (at least) equal to $1 - \alpha$. Equivalently, they control the probability of missing at least one component of the vector to be (at most) equal to α. Borrowing from the multiple testing literature, the latter probability can be termed the *familywise error rate* (FWE); for example, see Romano et al. (2008). So for a joint confidence region (JCR) for $\boldsymbol{\mu}$,

$$\text{FWE} = \mathbb{P}\{\text{At least one of the } \mu_h \text{ not contained in the JCR}\} \ , \tag{2.1}$$

whereas for a joint prediction region (JPR) for Y,

$$\text{FWE} = \mathbb{P}\{\text{At least one of the } y_h \text{ not contained in the JPR}\} \ . \tag{2.2}$$

The next section details how the FWE can be controlled in practice. It only does this in the context of a joint prediction region for Y. The method is similar in the context of a joint confidence region for $\boldsymbol{\mu}$ and is detailed in Romano and Wolf (2005, 2007) already.

Since the method is based on quantiles of random variables whose cumulative distribution function may not be invertible, the following remark is in order.

Remark 2.1. If the cumulative distribution function of a random variable is not invertible, then its quantiles are not necessarily uniquely defined. To be specific, we adopt the following definition for quantiles in this paper.

Let X be a random variable with cumulative distribution function $F(\cdot)$. Then, for $\lambda \in (0,1)$, the λ quantile of (the distribution) of X is defined as $\inf\{x : F(x) \geq \lambda\}$.

3 Joint Prediction Regions Based on FWE Control

Any formal analysis has to be put into a suitable framework. To this end, we borrow some notation from Jordà et al. (2010). We only discuss the case of a univariate time series; the extension to the case of a multivariate time series can be found in Wolf and Wunderli (2012).

One observes a univariate time series $\{y_1, \ldots, y_T\}$ generated from a true probability mechanism $\mathbb{P}$ and wishes to predict the future path $Y_{T,H} = (y_{T+1}, \ldots, y_{T+H})'$. At time t, denote a forecast h periods ahead by $\widehat{y}_t(h)$. Then a path forecast for $Y_{T,H}$ is given by $\widehat{Y}_T(H) = (\widehat{y}_T(1), \ldots, \widehat{y}_T(H))'$. Denote the vector of prediction errors by $\widehat{U}_T(H) = (\widehat{u}_T(1), \ldots, \widehat{u}_T(H))' = \widehat{Y}_T(H) - Y_{T,H}$. Finally, $\widehat{\sigma}_T(h)$ denotes a prediction standard error, that is, a standard error for $\widehat{u}_T(h)$: it is an estimator of the unknown standard deviation of the random variable $\widehat{u}_T(h)$.

We further assume a generic method to compute a vector of bootstrap prediction errors $\widehat{U}^*_T(H) = (\widehat{u}^*_T(1), \ldots, \widehat{u}^*_T(H))'$, based on artificial bootstrap data $\{y^*_1, \ldots, y^*_T, y^*_{T+1}, \ldots, y^*_{T+H}\}$ generated from an estimated probability mechanism $\widehat{\mathbb{P}}_T$.

Our high-level assumption below is based on the two vectors of *standardized prediction errors* $\widehat{S}_T(H) = (\widehat{u}_T(1)/\widehat{\sigma}_T(1), \ldots, \widehat{u}_T(H)/\widehat{\sigma}_T(H))'$ and $\widehat{S}^*_T(H) = (\widehat{u}^*_T(1)/\widehat{\sigma}^*_T(1), \ldots, \widehat{u}^*_T(H)/\widehat{\sigma}^*_T(H))'$, respectively. Denote the probability law under $\mathbb{P}$ of $\widehat{S}_T(H)|y_T, y_{T-1}, \ldots$ by $\widehat{J}_T$. Also denote the probability law under $\widehat{\mathbb{P}}_T$ of $\widehat{S}^*_T(H)|y^*_T, y^*_{T-1}, \ldots$ by $\widehat{J}^*_T$. In the asymptotic framework, T tends to infinity whereas H remains fixed.

Assumption 3.1. *$\widehat{J}_T$ converges in distribution to a non-random continuous limit law $\widehat{J}$. Furthermore, $\widehat{J}^*_T$ consistently estimates this limit law: $\rho(\widehat{J}_T, \widehat{J}^*_T) \to 0$ in probability, for any metric ρ metrizing weak convergence.*

Expressed in words, Assumption 3.1 states that, as the sample size T increases, the conditional distribution of the vector of standardized bootstrap prediction errors $\widehat{S}^*_T(H)$ becomes a more and more reliable approximation to the (unknown) conditional distribution of the vector of true standardized prediction errors $\widehat{S}_T(H)$.

We next specify the forms of the joint prediction regions for $Y_{T,H}$ for the two-sided case; the specification for the one-sided case can be found in Wolf and Wunderli (2012).

Some further notation is required. Suppose $X = (x_1, \ldots, x_H)'$ is a vector with H components. Then, $|X|$ denotes the vector $\big(|x_1|, \ldots, |x_H|\big)'$.

Let $d^{\max}_{|\cdot|,1-\alpha}$ denote the $1-\alpha$ quantile of the random variable $\max\big(|\widehat{S}_T(H)|\big)$. Then a two-sided joint prediction region for $Y_{T,H}$ that exactly controls the FWE is given by

$$\big[\widehat{y}_T(1) \pm d^{\max}_{|\cdot|,1-\alpha} \cdot \widehat{\sigma}_T(1)\big] \times \cdots \times \big[\widehat{y}_T(H) \pm d^{\max}_{|\cdot|,1-\alpha} \cdot \widehat{\sigma}_T(H)\big] \ . \tag{3.1}$$

The problem is that this ideal region is not feasible, since the constant $d^{\max}_{|\cdot|,1-\alpha}$ is unknown. It has to be estimated in practice by $d^{\max,*}_{|\cdot|,1-\alpha}$, which is defined as the $1-\alpha$ quantile of the random variable $\max\big(|\widehat{S}^*_T(H)|\big)$. This quantile can typically not be derived analytically, but it

can be simulated to arbitrary precision from a sufficiently large number of bootstrap samples; see Wolf and Wunderli (2012) for the details.

Then a two-sided joint prediction region for $Y_{T,H}$ that asymptotically controls the FWE is given by

$$\left[\widehat{y}_T(1) \pm d^{\max,*}_{|\cdot|,1-\alpha} \cdot \widehat{\sigma}_T(1)\right] \times \cdots \times \left[\widehat{y}_T(H) \pm d^{\max,*}_{|\cdot|,1-\alpha} \cdot \widehat{\sigma}_T(H)\right] . \tag{3.2}$$

The following proposition formally establishes the asymptotic validity of this feasible bootstrap joint prediction region; the proof can be found in Wolf and Wunderli (2012).

Proposition 3.1. *Under Assumption 3.1, the joint prediction region (3.2) for $Y_{T,H}$ satisfies*

$$\limsup_{T\to\infty} \; FWE \leq \alpha \; . \tag{3.3}$$

3.1 Comparison with Previous Methods

Jordà and Marcellino (2010) propose an alternative 'asymptotic' method to construct a joint prediction region for $Y_{T,H}$ that controls the FWE.[1] It is based on the assumption that

$$\sqrt{T}\big(\widehat{Y}_T(H) - Y_{T,H}|y_T, y_{T-1}, \ldots\big) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Xi}_H) \; , \tag{3.4}$$

where $\xrightarrow{d}$ denotes convergence in distribution, and on the availability of a consistent estimator $\widehat{\mathbf{\Xi}}_H \xrightarrow{\mathbb{P}} \mathbf{\Xi}_H$, where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability.

The proposed joint prediction region is given by

$$\widehat{Y}_T(H) \pm P \left[\sqrt{\frac{\chi^2_{h,1-\alpha}}{h}}\right]_{h=1}^{H} \; , \tag{3.5}$$

where P is the lower-triangular Cholesky decomposition of $\widehat{\mathbf{\Xi}}_H/T$, satisfying $PP' = \widehat{\mathbf{\Xi}}_H/T$, and the quantity to the right of P is a $H \times 1$ vector whose h^{th} entry is given by $\sqrt{\chi^2_{h,1-\alpha}/h}$. This approach has several theoretical flaws, as explained by Wolf and Wunderli (2012).

Staszewska-Bystrova (2011) proposes an alternative bootstrap method to construct a joint prediction region for $Y_{T,H}$ that controls the FWE. In a nutshell, the method works as follows. Conditional on the observed data, one generates B bootstrap path forecasts $\widehat{Y}^{*,b}_T(H)$, for $b = 1, \ldots, B$. One then discards αB of these bootstrap path forecasts: namely those $\widehat{Y}^{*,b}_T(H)$ that are 'furthest' away from the original path forecast $\widehat{Y}_T(H)$, where the distance between two $H \times 1$ vectors is measured by the Euclidian distance.[2] Finally, the joint prediction region is defined as the envelope of the remaining $(1-\alpha)B$ bootstrap path forecasts, where the term *envelope* refers to the smallest region containing all remaining bootstrap path forecasts. Although this *neighboring paths (NP)* method seems to perform quite well in simulation studies, it is purely heuristic and not backed up by theoretical analysis.

As evidenced by extensive Monte Carlo studies in Wolf and Wunderli (2012), our proposed bootstrap joint confidence region also has better finite-sample coverage properties than these two previous proposals, and in particular than the proposal of Jordà and Marcellino (2010).

[1] They use the term *joint confidence region* instead of *joint prediction region*.

[2] Staszewska-Bystrova (2011) also considers other distance measures, but concludes that the Euclidean distance seems to work best.

4 Conclusions

Many economic and financial applications require the forecast of a random variable of interest over several periods into the future, that is, one needs to forecast an entire future path. In addition to the resulting path forecast, one often would also like to compute a corresponding joint prediction region. Such a region is supposed to contain the entire future path with a prespecified probability $1 - \alpha$.

In this paper, we have proposed a two-sided bootstrap joint prediction region; in addition, analogous one-sided lower and one-sided upper joint prediction regions are detailed in Wolf and Wunderli (2012). This way, the applied researcher can choose the most suitable shape for the task at hand. Furthermore, the joint prediction regions are completely generic in that they allow the applied researcher to select whichever methods are deemed most appropriate by him to make forecasts, compute prediction standard errors, and generate bootstrap data.

Compared to two previous proposals in the literature, our bootstrap joint prediction regions have two important advantages. First, they are proven to be asymptotically consistent under a realistic, mild high-level assumption. Second, they enjoy superior finite-sample properties, as demonstrated via extensive Monte Carlo simulations in Wolf and Wunderli (2012).

References

Jordà, Ò., Knüppel, M., and Marcellino, M. G. (2010). Empirical simultaneous confidence regions for path-forecasts. Discussion Paper No. DP7797, CEPR. Available at SSRN: `http://ssrn.com/abstract=1611493`.

Jordà, Ò. and Marcellino, M. G. (2010). Path-forecast evaluation. *Journal of Applied Econometrics*, 25:635–662.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer, Berlin.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2008). Formalized data snooping based on generalized error rates. *Econometric Theory*, 24(2):404–447.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *Annals of Statistics*, 35(4):1378–1408.

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

Wolf, M. and Wunderli, D. (2012). Bootstrap joint prediction regions. Working Paper ECON 64, Department of Economcis, University of Zurich.