Discussion of “Bootstrap prediction intervals for linear, nonlinear, and nonparametric autoregressions”, by Li Pan and Dimitris Politis

Sílvia Gonçalves and Benoit Perron

Département de sciences économiques, Université de Montréal, CIREQ and CIRANO

July 28, 2014

We would like to start by congratulating the authors for having written this important paper. Prediction intervals are popular in economics and finance (e.g. they are often used by Central Banks to measure point forecasts uncertainty). The paper provides a unifying treatment of bootstrap prediction intervals for autoregression models, which are one of the workhorse models for economic forecasting. Therefore, the methods proposed by the authors will likely have an important impact on the economics profession.

The paper considers autoregressive models of the form $X_t = m\left(X_{t-1}, X_{t-2}, \ldots, X_{t-p}\right) + \epsilon_t$, where the errors ϵ_t satisfy the following assumption:

$$\epsilon_1, \epsilon_2, \ldots \text{ are i.i.d. } \left(0, \sigma^2\right) \text{ and } \epsilon_t \text{ is independent from } \{X_{t-1}, X_{t-2}, \ldots\} \text{ for all } t. \tag{1}$$

An heteroskedastic version of the model is also considered in the paper, but we will focus our attention on the homoskedastic version. The functions $m(\cdot)$ and $\sigma(\cdot)$ are unknown and potentially nonparametric, thus the models are rather general along this dimension. However, in economics and many other fields, it is standard to include information from additional predictors when forecasting a given variable of interest, for example predicting excess stock returns using the dividend-price ratio. Therefore, our first object of interest is how to extend the analysis of the paper (assumptions, definitions and the corresponding bootstrap methods) when extra predictors are used for forecasting future values of the target variable X. Our second concern is the construction of prediction intervals for X in the presence of *estimated* predictors. Factor-augmented prediction models are an important example in economics where estimated predictors arise. These factors are estimated from large panels of macroeconomic and financial data and provide a way to include a large set of information when making forecasts. They have become increasingly common since the work of Stock and Watson (2002).

In the sequel, we will highlight some of the issues that the presence of additional predictors create for bootstrap prediction intervals. To simplify the exposition, we will focus on the simple ARX(1) model

$$X_t = \phi_1 X_{t-1} + \alpha' F_{t-1} + \epsilon_t, \tag{2}$$

where F_{t-1} is an $r \times 1$ vector containing extra predictors, which might be observed or not. We will focus on $h = 1$ in the next two sections and reserve our discussion of multi-step forecasting for Section 3.

1 Observed F_{t-1}

Extending the analysis of the paper when F_{t-1} is observed is relatively straightforward provided we modify the assumptions and the definitions appropriately to account for the presence of F_{t-1}. In particular, suppose we observe data $\{X_t, F_t : t = 1, \ldots, n\}$ and we care about (interval) prediction of X_{n+1} given information at time n. Under the martingale difference assumption that

$$E\left(\epsilon_t | X_{t-1}, X_{t-2}, \ldots, F_{t-1}, F_{t-2}, \ldots\right) = 0, \tag{3}$$

the MSE-optimal predictor of X_{n+1} at time n is given by

$$X_{n+1|n} = E\left(X_{n+1} | X_n, X_{n-1}, \ldots, F_n, F_{n-1}, \ldots\right) = \phi_1 X_n + \alpha' F_n.$$

Its feasible version is $\hat{X}_{n+1} = \hat{\phi}_1 X_n + \hat{\alpha}' F_n$. The prediction error $X_{n+1} - \hat{X}_{n+1}$ can be decomposed as

$$\begin{aligned} X_{n+1} - \hat{X}_{n+1} &= \left(X_{n+1} - X_{n+1|n}\right) + \left(X_{n+1|n} - \hat{X}_{n+1}\right) \\ &= \epsilon_{n+1} + \left[\left(\phi_1 - \hat{\phi}_1\right) X_n + (\alpha - \hat{\alpha})' F_n\right], \end{aligned}$$

where the first term reflects the innovation error and the second term reflects the estimation error. These two terms are independent if we strengthen (1) by

Assumption 1. $\epsilon_1, \epsilon_2, \ldots$ are i.i.d.$(0, \sigma^2)$ and ϵ_t is independent from $\{X_{t-1}, X_{t-2}, \ldots, F_{t-1}, F_{t-2}, \ldots\}$ for all t.

Assumption 1 extends the causality assumption in the paper (cf. eq. (1)) in the presence of extra predictors. This assumption is very strong for economic applications because it requires the predictors $\{F_s : s < t\}$ to be independent of ϵ_t for each t (this is a much stronger form of exogeneity as compared to the martingale difference condition (3), which only requires conditional mean independence between the predictors and the error term). We wonder if it could be possible to relax this assumption and still get valid prediction intervals for X_{n+1} in the presence of extra predictors.

The forward bootstrap method that is advocated in the paper is easily generalized to this case if we condition on the values of F_t when bootstrapping X_t. Specifically, the two steps of the method for the ARX(1) model are as follows.

A. For $t = 1, \ldots, n$, let

$$X_t^* = \hat{\phi}_1 X_{t-1}^* + \hat{\alpha}' F_{t-1} + \epsilon_t^*,$$

where X_0^* is a starting value chosen appropriately and $\{\epsilon_t^* : t = 1, \ldots, n\}$ is an i.i.d. resample from $\left\{\hat{\epsilon}_t - \bar{\hat{\epsilon}} : t = 1, \ldots, n\right\}$. Get the bootstrap analogues of $\hat{\phi}_1$ and $\hat{\alpha}$, $\hat{\phi}_1^*$ and $\hat{\alpha}^*$, respectively.

B. Let $X_n^* = X_n$ and set

$$\begin{aligned}\hat{X}_{n+1}^* &= \hat{\phi}_1^* X_n + \hat{\alpha}^{*\prime} F_n \\ X_{n+1}^* &= \hat{\phi}_1 X_n + \hat{\alpha}' F_n + \epsilon_{n+1}^*,\end{aligned}$$

where ϵ_{n+1}^* is an i.i.d. resample from $\left\{\hat{\epsilon}_t - \bar{\hat{\epsilon}} : t = 1, \ldots, n\right\}$.

The bootstrap prediction error $X_{n+1}^* - \hat{X}_{n+1}^*$ can be decomposed as

$$X_{n+1}^* - \hat{X}_{n+1}^* = \epsilon_{n+1}^* + \left[\left(\hat{\phi}_1 - \hat{\phi}_1^*\right) X_n + (\hat{\alpha} - \hat{\alpha}^*)' F_n\right],$$

where the second term is independent of ϵ_{n+1}^*, conditionally on the data (as happens in the real world).

Given the exogeneity assumption on F_t, fixing the value of the extra predictors is entirely natural. It also automatically implies that the bootstrap point forecast for $n+1$ is conditional on F_n, which is very important for obtaining interval forecasts that are conditionally valid (as emphasized in the paper). Generating the observations on X_t^* recursively as is done in step A of the method is also natural as it exploits the parametric structure of the model. However, given that the main purpose of step A is to replicate the parameter estimation uncertainty in $\hat{\phi}_1$ and $\hat{\alpha}$, we believe other bootstrap methods could be used. For instance, a fixed-design bootstrap method that fixes both X_{t-1} and F_{t-1} when generating X_t^* would also allow us to capture the parameter estimation uncertainty. The asymptotic validity of this method for confidence intervals in the context of heteroskedastic autoregressive models was proven in Gonçalves and Kilian (2004, 2007). This work showed that bootstrap confidence intervals based on fixing the regressors were valid more generally than bootstrap intervals based on the recursive-design version (in particular, the fixed-design bootstrap allowed for asymmetric forms of conditional heteroskedasticity such as the popular EGARCH (Exponential Generalized AutoRegressive Conditional Heteroskedastic) model of Nelson (1991), which were ruled out when proving the validity of the recursive-design bootstrap).

One of the crucial contributions of the paper is to propose a definition of asymptotic validity of a bootstrap prediction interval that takes into account the presence of parameter estimation uncertainty. The usual notion of asymptotic validity (see Definition 2.3) requires that the prediction interval contains the future observation with a probability that converges to a given nominal level as the sample size increases. (In the autoregressive context, this probability should be conditional on the value of the last observed values of the predictors used in forming the point forecast.) Although this property is fundamental for the validity of a prediction interval, it is not stringent enough to account for parameter estimation uncertainty. Therefore, the paper proposes the notion of asymptotic pertinence (cf. Definition 2.4) as an extension of Definition 2.3 that accounts for this source of variability in the point forecast.

Under Assumption 1, extending the definitions of asymptotic validity and asymptotic pertinence to the case of ARX models is straightforward provided we extend the information set to include both X_n *and* F_n.

2 Unobserved F_{t-1}: factor-augmented regressions

Factor-augmented regression models have become very popular for forecasting economic variables since the seminal paper by Stock and Watson (2002). The main idea is that we think of the predictors F_{t-1} in equation (2) as being the unobserved common factors underlying a panel factor model:

$$Z_t = \Lambda F_t + e_t, \quad t = 1, \ldots, n,$$

where Z_t contains N observed variables, Λ is an $N \times r$ matrix containing the factor loadings, and e_t is the $N \times 1$ vector of idiosyncratic error terms. Point predictions of X_{n+1} are then obtained by a two-step procedure: first, we typically obtain $\left\{ \tilde{F}_t \right\}$ by applying the method of principal components to Z_t (although some other estimators are available), and then we regress X_t on X_{t-1} and $\tilde{F}_{t-1}$ to obtain the OLS estimators $\hat{\phi}_1$ and $\hat{\alpha}$. The point forecast of X_{n+1} is given by

$$\hat{X}_{n+1} = \hat{\phi}_1 X_n + \hat{\alpha}' \tilde{F}_n.$$

Since the above factor model is only identified up to rotation, Bai (2003) showed that the principal component of F_t converges (as N and n go to infinity jointly) to HF_t where H is a matrix that depends on population and sample quantities. This means that $\hat{\alpha}$ does not converge to α but to $H^{-1\prime}\alpha$. Note however that this lack of identification does not affect the forecast $\hat{X}_{n+1}$.

Since

$$X_{n+1} = \phi_1 X_n + \alpha' F_n + \epsilon_{n+1},$$

the prediction error $X_{n+1} - \hat{X}_{n+1}$ can now be decomposed as

$$X_{n+1} - \hat{X}_{n+1} = \epsilon_{n+1} + \left[\left(\phi_1 - \hat{\phi}_1 \right) X_n + (\alpha - \hat{\alpha})' F_n + \hat{\alpha}' \left(F_n - \tilde{F}_n \right) \right].$$

The second term reflects not only the uncertainty associated with the estimation of ϕ_1 and α, but also the estimation uncertainty that arises from having to estimate F_n with $\tilde{F}_n$.

Recently, Gonçalves, Perron and Djogbenou (2013) proposed bootstrap prediction intervals for X_{n+1} in this context. Their method consists of the following steps:

1. For $t = 1, \ldots, n$, let

   (a) $X_t^* = \hat{\phi}_1 X_{t-1} + \hat{\alpha}' \tilde{F}_{t-1} + \epsilon_t^*$, where $\{\epsilon_t^* : t = 1, \ldots, n\}$ is an i.i.d. resample from $\left\{ \hat{\epsilon}_t - \bar{\hat{\epsilon}} : t = 1, \ldots, n \right\}$.

   (b) $Z_t^* = \tilde{\Lambda} \tilde{F}_t + e_t^*$, where $\{e_t^* : t = 1, \ldots, n\}$ is a resample from $\{\tilde{e}_t\}$ obtained independently from $\{\hat{\epsilon}_t\}$. Get the bootstrap analogues of $\tilde{F}_t$, $\tilde{F}_t^*$.

   (c) Get $\hat{\phi}_1^*$ and $\hat{\alpha}^*$, the bootstrap analogues of $\hat{\phi}_1$ and $\hat{\alpha}$, respectively, by running an OLS regression of X_t^* on X_{t-1} and $\tilde{F}_{t-1}^*$.

2. Let $X_n^* = X_n$ and set

$$\begin{aligned} \hat{X}_{n+1}^* &= \hat{\phi}_1^* X_n + \hat{\alpha}^{*\prime} \tilde{F}_n^* \\ X_{n+1}^* &= \hat{\phi}_1 X_n + \hat{\alpha}' \tilde{F}_n + \epsilon_{n+1}^*, \end{aligned}$$

   where ϵ_{n+1}^* is an i.i.d. draw from $\left\{ \hat{\epsilon}_t - \bar{\hat{\epsilon}} : t = 1, \ldots, n \right\}$.

It is interesting to note that this algorithm amounts to a form of the forward bootstrap method proposed by the current paper. Steps A and B are now replaced by Steps 1 and 2. Step 1 is the analogue of Step A in the previous section. The main difference is that it contains an additional step (step b) that estimates $\tilde{F}^*_{t-1}$, which is then used as the estimated predictor for X^*_t (step c). The other difference is that step 1(a) is based on a fixed-design bootstrap instead of the recursive-design bootstrap discussed in the paper (see our previous comment). Step 2 is the analogue of Step B. As emphasized in the paper, we redefine the last bootstrap observation to be equal to X_n when constructing the bootstrap point forecast $\hat{X}^*_{n+1}$ and the bootstrap new observation X^*_{n+1}. However, and contrary to the case where F_n is observed, $\hat{X}^*_{n+1}$ now depends on $\tilde{F}^*_n$, the *estimated* version of $\tilde{F}_n$ that is used to construct X^*_{n+1}. This step is meant to capture the estimation uncertainty in the predictors.

Following previous work by Bai and Ng (2006), who showed the asymptotic validity of Gaussian prediction intervals for X_{n+1} in the factor-augmented regression model under the assumption that ϵ_t is i.i.d. $N(0,1)$, Gonçalves, Perron and Djogbenou (2013) proved that bootstrap prediction intervals, as described above, are asymptotically valid unconditionally, without requiring the Gaussianity of ϵ_t. An interesting extension of the present paper would be to provide an asymptotic framework that is able to handle both parameter estimation uncertainty and predictors estimation uncertainty, i.e. that extends the notion of asymptotic pertinence to this context.

3 Multi-horizon forecasting, $h > 1$

Finally, we consider the case where the forecasting horizon, h, is larger than 1. There are two approaches to produce such forecasts. The first one is to specify a forecasting model for horizon 1 and iterate it to the desired horizon h. This is the approach followed in the current paper. When other predictors are present, this approach has the drawback of requiring forecasts of the future values of the predictors, whether they are observed or not. This is usually accomplished in a vector autoregressive (VAR) framework when the predictors are observed or a factor augmented VAR (FAVAR) when predictors are estimated factors.

An alternative approach (the so-called direct approach) which avoids this difficulty is to make direct forecasts and specify a forecasting model at horizon h. In this scenario, the analog of our ARX model is

$$X_t = \phi_h X_{t-h} + \alpha' F_{t-h} + \epsilon_t, \tag{4}$$

which can be estimated by OLS for $t = h+1, ..., n$ with corresponding point forecast $\hat{X}_{n+h} = \hat{\phi}_h X_n + \hat{\alpha}' F_n$.

The main complication of the direct approach when $h > 1$ is the fact that the regression errors ϵ_t will generally be serially correlated to order $h-1$. This serial correlation affects the distribution of the estimated parameters and the construction of prediction intervals requires the use of an estimator of the covariance matrix that is robust to serial correlation as in Andrews (1991). Similarly, when constructing intervals using the bootstrap, a method for drawing errors that is robust to serial correlation is needed.

Gonçalves, Perron, and Djogbenou (2013) modify the algorithm from the previous section by drawing ϵ^*_t using the block wild bootstrap (BWB). The idea is to separate the sample residuals $\hat{\epsilon}_{t+h}$ into non-overlapping blocks of b consecutive observations. For simplicity, we assume that

$\frac{T-h}{b}$, the number of such blocks, is an integer. Then, we generate our bootstrap errors by multiplying each residual within a block j by the *same* draw of an external variable, i.e.

$$\epsilon^*_{i+(j-1)b} = \hat{\epsilon}_{i+(j-1)b} \cdot \eta_j$$

for $j = 1, ..., \frac{T-h}{b}$, $i = h+1, ..., h+b$, and $\eta_j \sim i.i.d. (0, 1)$. The fact that each residual within a block is multiplied by the same external draw preserves the time series dependence. Because the errors are $(h-1)-$dependent under correct specification, it is natural to set the block length to h. One final issue is how to generate X^*_{n+h}. The issue is that there is no residual $\hat{\epsilon}_{n+h}$ that can be used to draw ϵ^*_{n+h} as above. Gonçalves et al. (2013) make a draw from the empirical distribution function of $\hat{\epsilon}_t$, $t = h+1, ..., n$. This should reproduce the unconditional distribution of the error term, but it is not clear to us that this is the optimal choice. We believe that extending the notion of asymptotic pertinence to direct forecasts would be a very useful addition to the current paper.

4 Conclusion

Pan and Politis have made a significant contribution to the comparison of forecast intervals by introducing the concept of asymptotic pertinence that allows parameter estimation to matter even asymptotically. In this discussion, we have tried to extend this notion to the case where additional predictors are available. The case of estimated predictors is difficult as it is not obvious how to condition on the value of the latent predictor in the bootstrap world. We look forward to contributions and discussion on these issues.

References

[1] Andrews, D. W K, 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.Bai, J., 2003. "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-172.

[2] Bai, J. and S. Ng, 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-augmented Regressions," *Econometrica*, 74, 1133-1150.

[3] Bai, J. 2003. "Inferential theory for factor models of large dimensions", *Econometrica*, 71, 135-172.

[4] Gonçalves, S. and L. Kilian, 2007. "Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity," *Econometric Reviews*, 26 (6), 609-641.

[5] Gonçalves, S. and L. Kilian, 2004. "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89-120.

[6] Gonçalves, S. and B. Perron, 2013. "Bootstrapping Factor-Augmented Regressions", *Journal of Econometrics*, forthcoming.

[7] Gonçalves, S., B. Perron, and A. Djogbenou, 2013. "Bootstrap prediction intervals for factor models," mimeo, Université de Montréal.

[8] Djogbenou, A., S. Gonçalves, and B. Perron, 2013. "Bootstrap Inference for Regressions with Estimated Factors and Serial Correlation", mimeo.

[9] Stock, J. H. and M. Watson, 2002. "Forecasting using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167-1179.