NN3-Forecasting Competition: an Adaptive Robustified Multi-Step-Ahead Out-Of-Sample Forecasting Combination Approach

By Marc Wildi[1]

Institute of Data Analysis and Process Design, University of Applied Sciences, Winterthur and Institute of Business Cycle Research, KOF-ETH, Zurich.

Keywords: Multi-step-ahead out-of-sample forecasts, forecast combination, state-space, robustification.

1 Introduction

Extensive empirical experience suggests that traditional forecasting approaches are subject to more or less severe model misspecifications which affect true (out-of-sample) one- as well as multi-step ahead forecasting performances. The main problems are due to non-stationarity and non-Gaussianity. In order to overcome these difficulties, we propose a prototypical design derived from a traditional adaptive state-space approach which is suited for tracking non-stationarities. The proposed procedure has been heavily modified to account for true out-of-sample performances, for non-Gaussianity, for multi-step performances as well as potential misspecifications.

2 The original state-space approach

2.1 The model

In practice time series are often decomposed into components:

$$X_t = T_t + C_t + S_t + I_t$$

where T_t is the trend, C_t is stationary (sometimes C_t is called a 'cyclical' component), S_t is a seasonal process and I_t are random disturbances. Relying on that decomposition, the following 'classic' model-based approach has been chosen as **starting-point** for our method:

$$\begin{aligned} \boldsymbol{\xi_t} &= \mathbf{F}\boldsymbol{\xi_{t-1}} + \boldsymbol{\nu_t} & (1) \\ X_t &= \mathbf{H}\boldsymbol{\xi_t} + \epsilon_t & (2) \end{aligned}$$

[1]Marc Wildi, Institute of Data Analysis and Process Design, ZHW, tel. (32) 345 12 71, e-mail wia@zhwin.ch

whereby

$$\begin{aligned}
\boldsymbol{\xi_t}' &= (T_t, \Delta T_t, S_t, S_{t-1}, ..., S_{t-13}) \\
\mathbf{F} &= \begin{pmatrix}
1 & 1 & 0 & 0 & ... & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & ... & 0 & 0 & 0 \\
0 & 0 & ar_1 & ar_2 & 0 & ... & sar_1 & -ar_1 \cdot sar_1 \\
1 & 0 & 0 & ... & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & ... & 0 & 0 & 0 & 0 \\
... & ... & ... & ... & ... & ... & ... & \\
0 & 0 & 0 & ... & 1 & 0 & 0 &
\end{pmatrix} \\
\mathbf{H} &= (1, 1, 1, 0, ..., 0)
\end{aligned}$$

and the disturbance terms satisfy the usual iid and normality assumptions imposed in the literature. The first two rows of $\mathbf{F}$ define T_t as an I(2)-process

$$\begin{aligned}
T_t &= T_{t-1} + \Delta T_t + \nu_{1t} \\
\Delta T_t &= \Delta T_{t-1} + \nu_{2t}
\end{aligned}$$

where ΔT_t is a random-walk process determining the local trend growth. The third row accounts for the cyclical and seasonal structures simultaneously by specifying

$$S_t = ar_1 S_{t-1} + ar_2 S_{t-2} + sar_1 S_{t-12} - ar_1 \cdot sar_1 S_{t-13}$$

2.2 Optimization Criterion and Forecasting-Function

If all required model assumptions are satisfied, then the Kalman-filter can be used to decompose the likelihood function into conditional one-step ahead distributions of the one-step ahead forecasting errors. At the current boundary of the time series, the most recent component estimates obtained by the Kalman-filter can then be used to generate forecasts (basically, future innovations are equated to zero in the above state and observation equations 1 and 2).

Traditionally, components are estimated based on in-sample one-step ahead forecasting performances. Forecasts (one- and/or multi-step ahead) are derived accordingly. At this stage it is important to note that the above model assumptions legitimate the **in-sample one-step ahead (mean-square) error criterion**, even in a strict multi-step ahead perspective, because the resulting forecasting functions are maximum likelihood estimates of the future observations in this restricted theoretical framework.

3 'Mods'

In practice, neither of the above model assumptions are satisfied and therefore the traditional optimization criterion cannot be justified by invoking efficiency. We here propose a set of modifications of the traditional 'pure' model-based approach in order to account for misspecification issues.

3.1 Out-of-Sample One- and Multi-Step-Ahead Performances

True out-of-sample one-step ahead forecasting errors can be obtained conveniently from the updating-step of the Kalman-filter equations. More generally, multi-step ahead out-of-sample errors are obtained by letting the 'empty' updating equations (future innovations are set to zero) run until the desired forecasting horizon is attained. Therefore h-step optimization criteria can be generated by relying on the corresponding errors. More precisely, the hyperparameters (unknown innovation variances and starting values for the Kalman recursion) are optimized separately for each forecasting horizon $h = 1, ..., 18$. This allows to match the adaptivity (of the resulting forecasting functions) to the intended forecasting horizon.

3.2 Robustification

It is well-known in practice that the mean-square error criterion emphasizes extreme observations and that outliers can often not be detected by analyzing (in-sample) residuals. Fortunately, out-of-sample forecasting errors are less subject to 'smearing-effects'. Therefore, the proposed optimization criterion can be robustified by truncating the traditional up-dating formulas in the case of 'huge' unexpected out-of-sample deviations.

3.3 Numerical optimization

Determining hyperparameters in our approach is a tricky numerical task. In particular, one has to account for non-linearities induced by the above robustification step. In order to overcome problems associated to local extrema (of the generalized optimization criterion), genetic algorithms are used in conjunction with local gradient procedures. The latter enhance convergence in the vicinity of the optimum.

3.4 Forecasting combination

It is well-known that a combination of various forecasting functions often improves over the individual forecasts because of misspecification issues. In our approach, we can rely on 18 different forecasting functions that are obtained by optimizing hyperparameters for each individual forecasting horizon $h = 1, ..., 18$. The final submitted forecasts are obtained by computing the median of the 18 estimates for each $h = 1, ..., 18$.

3.5 Adaptivity

Besides the inherent adaptivity of the underlying state-space approach we emphasize more specifically forecasting issues at the **current boundary** of the time series by discounting past forecasting performances at an exponentially decreasing rate in our optimization criteria.

4 Conclusion

The proposed forecasting method starts with a classical model-based approach and evolves into an almost non-parametric design based on extensions and generalizations that account for frequently observed misspecification problems such as Non-Gaussianity (robustification), overfitting (out-of-sample forecast combination) or numerical issues (genetic algorithm). It is a pure prototype - in the sense that we do not have any practical experience up to yet - that is intended for future applications in macroeconomics and social sciences. In particular, there are some important 'tuning-parameters' whose precise settings are unknown at the present stage. Moreover, numerical problems still seem to affect performances for particular time series and particular forecasting horizons (the latter problem is tackled, at least partially, by computing medians of the available 18 forecasting functions).

Our interest in participating to the nn3-competition is to assess the relative performance of our prototype in this experimental setting, in particular when compared to traditional statistical methods as well as non-linear forecasting rules.