International Journal of Forecasting 36 (2020) 129–134

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Card forecasts for M4

Jurgen A. Doornik [a,*], Jennifer L. Castle [b], David F. Hendry [a]

[a] Nuffield College and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK
[b] Magdalen College and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK

ARTICLE INFO

Keywords:
Automatic forecasting
Calibration
Forecast intervals
Regression
M4
Seasonality
Software
Time series
Unit roots

ABSTRACT

The M4 forecast competition required forecasts of 100,000 time series at different frequencies. We provide a detailed description of the calibrated average of Rho and Delta (Card) forecasting method that we developed for this purpose. Delta estimates a dampened trend from the growth rates, while Rho estimates an adaptive but simple autoregressive model. Calibration estimates a more elaborate autoregressive model, treating the averaged forecasts from Rho and Delta as if they were observed. The proposed method is easy to understand, combining very fast execution with an excellent forecast performance.

1. Introduction

Our submission to the M4 competition, which we labelled *Card*, did well overall, particularly for hourly and weekly data, and forecast intervals. We had multiple objectives in developing *Card*. The first was based on the observation that, while it is difficult to do better than simple integrated autoregressive moving average (ARIMA) models, such models can occasionally give wild forecasts unless due care is taken. Thus, our aim was to determine what adjustments were needed to improve their forecasting performance in this setting. This was achieved by studying many aspects of the forecasting model, including transformations, seasonality, lag lengths, test significance, forcing unit roots, etc. The calibration of settings and parameters for the large M4 data set was feasible after developing a fast forecasting environment. Many of the steps below have been studied for their impact on the forecast performance using an additional holdback sample from the competition data set. As a final aim, we sought more experience in analyzing large economic databases with Ox Doornik (2013).

The remainder of this paper is as follows. Section 2 first describes how we organized the analysis, then Section 3 gives an overview of our approach, with details in the subsequent three sections. Next, Section 7 discusses the forecast intervals, and finally, Section 8 concludes. Some additional details are contained in appendices.

2. Development framework and preliminary analysis

M4 consists of 100,000 time series variables, split over six frequencies: yearly, quarterly, monthly, weekly, daily and hourly. The data are also subdivided into categories, but we ignored that information. The data are supplied in comma-separated files, with one variable per row. Our initial attempt to transform this to storing the variables in columns took several minutes to run, which we deemed too slow. We improved this by ordering the variables within each frequency by sample size and storing them in separate blocks (11 for monthly data, the largest segment). Each block forms a rectangular array with a limited number of missing values, which can be read more quickly. For yearly data, our code loads the data *and* produces 23,000 forecasts in less time than it takes for the R code supplied by the M4-team just to load the data. These implementation aspects were an important factor in the development of our methods, as they made it possible to experiment with the procedures.

With so many variables, it is not possible to explore more than a small subset of series. As a complement, we

* Corresponding author.
E-mail address: jurgen.doornik@nuffield.ox.ac.uk (J.A. Doornik).

https://doi.org/10.1016/j.ijforecast.2019.03.012

plotted distributions of means, variances, and seasonal aspects. This last indicated the presence of strong seasonality in many series. We also looked at the distributions of ordinary least squares (OLS) estimates of the autoregressive parameters from simple models: $y_t = \mu + \rho y_{t-1} + \tau t + \epsilon_t$ and

$$\log y_t = \mu + \rho \log y_{t-1} + \epsilon_t. \quad (1)$$

As would be expected for economic time series, $\widehat{\rho}$ values near one were prevalent. This corresponded to our prior expectation that a unit root model for the growth rates, i.e., Eq. (1) with $\rho = 1$, should form the basis for forecasting. However, there were many cases with $\widehat{\rho} > 1$ in the short annual series. That would result in explosive forecasts, and just one of these could distort the average forecast performance.

Forecasts from Eq. (1) with the unit root imposed have a linear trend of magnitude $\widehat{\mu}$. It is often advantageous to be conservative with this trend. For example, the *Theta(2)* method of Assimakopoulos and Nikolopoulos (2003) halves the impact of the trend, and did very well in the M3 competition (see Makridakis & Hibon, 2000).

These arguments formed the basis for the two forecasting devices that we developed, called *Delta* and *Rho*. Both are adaptive versions of simple techniques, with the adaptation criteria being evaluated after withholding H observations from the 'training' M4 data set, as well as the full M3 data set. The evaluation criteria were as used in M4, as was the number of forecasts H (6, 8, 18, 13, 14, 48 respectively). This mimics the M4 competition, but with the drawback that annual series in particular get quite short.

2.1. Preliminary stage

A particular time series is denoted by y_t, $t = 1, \ldots, T$. We limit the sample size, T, of the forecasting models to 10 years for daily data, 210 days for hourly, and 40 years for all other seasonal frequencies. Each series has a primary seasonality S and (possibly) a secondary seasonality S_2, using $1, 4, 12, 52, 5 \times 12$, and 24×7, respectively. The aim is to produce H forecasts of the original series: $\widehat{y}_{T+1}, \ldots, \widehat{y}_{T+H}$. However, *Card* usually forecasts a transformation, which is then transformed back at the end. The procedure is as follows.

1. If $\min(y_1, \ldots, y_T) > 1$, take natural logs: $x_t = \log(y_t)$, else set $x_t = y_t$.
   When logs are taken, we transform the forecasts back at the end using $\widehat{y}_{T+h} = \exp(\widehat{x}_{T+h})$.
2. Define I_ρ as an indicator that is one when the differences have the smaller sample variance:
   $$I_\rho = I(\mathrm{var}[\Delta x_t] \leq 1.2\ \mathrm{var}[x_t]). \quad (2)$$
   The factor 1.2 was determined by experimentation, looking at the forecast performance when holding back H observations, as was described in the previous section.
3. I_A indicates the presence of (additive) seasonality. An ANOVA-based test for stable seasonality is applied to Δx_t (or x_t when $I_\rho = 0$). Seasonality is assumed to be present if the F-test (see Appendix A) rejects at 10%, provided that there are enough observations:
   $$I_A = I(S > 1 \ \text{ and } \ T >= 3S + I_\rho \ \text{ and } \ \mathrm{Prob}[F > A\{\Delta x_t, S\}] < 0.1). \quad (3)$$
4. I_R indicates the presence of a seasonal autoregressive lag R. This is based on the significance of the Sth term of the autocorrelation function of Δx_t (or x_t when $I_\rho = 0$):
   $$I_R = I(S > 1 \ \text{ and } \ T >= 3S + I_\rho \ \text{ and } \ \mathrm{Prob}[\chi^2_1 > R\{\Delta x_t, S\}] < 0.1). \quad (4)$$

3. Overview of *Card*

Forecasting commences after the preliminary steps outlined in Section 2.1. For M4 (and M3), this means that all forecasting is done in logarithms. Furthermore, $I_\rho = 1$ almost always: monthly data has the lowest rate, at 94%. Additive seasonality (using logs), is detected in roughly half of the quarterly, monthly, and weekly series, in about 7% for daily, and 96% for hourly series. Similar percentages obtain for I_R, except for a higher incidence for daily data, at 28%, and almost all hourly series. The calibrated average of *Rho* and *Delta*, or *Card*, method involves five steps.

Delta The growth rate is estimated as the mean of the first differences, hence the *Delta* label, but with the largest values removed and additional dampening if the growth rate at the end of the sample is smaller than this. Seasonality is based on the 3×5 moving average (MA) of each season's growth rates. The *Delta* method applied to annual series outperforms *Theta(2)* forecasts of the levels by quite a margin, although somewhat less so in M3. This could be because the dampening of the trend is more adaptive to the data (and taking logarithms helps too).

Rho The second method estimates an autoregressive model, with seasonal dummies and a seasonal root if detected. If the estimated $\widehat{\rho}$ is 'large' (as determined by the evaluation procedure; the details are given in Section 5), then a unit root is imposed. In that case, the estimated intercept, which determines the trend, is dampened on the basis of its estimated significance.

Averaging Although *Rho* does not perform as well as *Delta* in general, it is useful in a forecast combination of the two: the combination usually has better forecasts than either alone. We take averages using the simple arithmetic mean of the forecasts for the transformed series.

Calibration The calibration stage treats the forecasts as if they were observed and estimates an autoregressive model that is richer than the one used for *Rho*, see Eq. (7) below. The fitted values of this model are used as the final forecasts. Whereas the *Delta* and *Rho* methods extrapolate the model in order to obtain forecasts, the calibrated forecasts are the in-sample fits, meaning that even an explosive root does not impact the latter. As another example, if the calibration model had a perfect fit, the new forecasts would just be the original forecasts.

Transformation The final step is to undo the logarithmic transformation. There is no bias correction, and this amounts to the median forecast of the raw variable.

The calibration step seems particularly useful in getting improved estimates of seasonality: there is no gain for annual data, but a large one for weekly and hourly data. Our calibration is somewhat similar to the X12-ARIMA procedure for the seasonal adjustment of quarterly and monthly data, although the ARIMA model comes first in that case Findley, Monsell, Bell, Otto, and Chen (1998).

The calibration model is also used to obtain 90% forecast intervals. Some further minor adjustments are made at several stages that are specific to a frequency.

4. Forecasting with the *Delta* method, $I_\rho = 1$

The *Delta* method derives growth rates (assuming that the input is in logs) d_1, d_2 from the order statistics: d_1 drops the largest growth rate, and d_2 the largest three. The 'recent' growth d_r is the mean of the final six observations. A function *amin* with two arguments is defined: it either sets the result to zero if there is positive and negative growth, or selects the smaller value of the two. Then, the final growth rate is dampened severely unless it is present in a consistent manner. Up to three large structural breaks can occur without affecting the forecasts.

The input to the *Delta* method consists of the transformed series, the status of seasonality, I_A, and differencing, I_ρ. More formally: $x_t, t = 1, \dots, T,\ S, H, I_A, I_\rho$. The $I_\rho = 0$ case is handled in Appendix B. The detailed steps of the *Delta* method for $I_\rho = 1$ are given next.

$I_A = 0$. Set $z_t = \Delta x_t$. Let $z_{(i)}$, be the sorted zs such that $|z_{(i)}| \le |z_{(i+1)}|$, $i = 1, \dots, T-1$.

Compute

$$d_1 = \frac{1}{T-2}\sum_{i=1}^{T-2} z_{(i)},$$

$$d_2 = \frac{1}{T-4}\sum_{i=1}^{T-4} z_{(i)},$$

$$d_r = \frac{1}{6}\sum_{t=T-5}^{T} z_t,$$

where $d_2 = d_r = d_1$ if $T \le 6$.

Define $\operatorname{amin}(a,b) = 0$ if $ab \le 0$, $\operatorname{amin}(a,b) = a$ if $|a| \le |b|$, otherwise $\operatorname{amin}(a,b) = b$. Compute $d_m = \bar{z}$; if $T > 2S+1$ then $d_m = \operatorname{amin}(d_m, \frac{1}{s}\overline{\Delta_s x})$. Next, $d_r^* = \operatorname{amin}(d_r, d_m)$, and finally, construct forecasts without seasonality ($s_{Y,j} = 0$):

$$\begin{aligned} \widehat{x}_{T+1} &= x_T + \operatorname{amin}(d_r^*, d_1) + s_{Y,1}, \\ \widehat{x}_{T+h} &= \widehat{x}_{T+h-1} + \operatorname{amin}(d_r^*, d_2) + s_{Y,h}, \\ & \quad h = 2, \dots, H. \end{aligned} \tag{5}$$

$I_A = 1$. Construct the seasonal table of Δx_t with each season in a column; see the example in Appendix A.4. Smoothing is applied to each season, giving for period j: $s^*_{.,j} = \mathrm{MA}_{3\times 5}\Delta x_{.,j}$. The future seasonal is taken from the last 'year': $s_{Y,j} = s^*_{Y,j} - \sum_{j=1}^{S} s^*_{Y,j}/S$. The 'annual' means $z_t = \overline{\Delta x_{i,.}}$ are used to compute d_1, d_2 and d_r in Eq. (5).

5. Forecasting with the *Rho* method

At the core of the *Rho* method is an AR(1) model. Seasonality is handled by seasonal dummies and a sine + cosine term for the optional secondary seasonality. In addition, there may be a longer autoregressive lag at length R that is usually but not necessarily equal to S, see Appendix A. The unit root is imposed when $\widehat{\rho}$ is fairly close to one, and in that case, the trend growth is dampened. The unit root cut-off and the amount of dampening were determined by measuring the impact on the forecast performance.

The input consists of $x_t, I_\rho, S, S_2, H, I_A, I_R, R$. We start with $I_r = I_\rho$, $I_\Delta = 0$, $I_\tau = 0$, $I_2 = S_2 > 1$.

(1) Define $S_{2t} = \sin[2\pi t/(SS_2)]$ and $C_{2t} = \cos[2\pi t/(SS_2)]$, with centred seasonal dummies $q_{j,t}$, and estimate

$$\begin{aligned} x_t &= \mu + (\rho x_{t-1} + \rho_R x_{t-R} I_R) I_r \\ &\quad + \tau \lfloor t/S \rfloor I_\tau + \left\{\delta_j q_{j,t}\right\} I_A + \left(\gamma_2 S_{2t} + \gamma_2^* C_{2t}\right) I_2 + \epsilon_t \end{aligned} \tag{6}$$

using OLS. Note that, when an indicator is zero, the corresponding term is excluded. One observation is lost when $I_r = 1$ and R when $I_r I_R = 1$. With hourly data, a maximum of one is lost, because the first 24 observations are duplicated for the seasonal lag.

(2) Only if $I_r = 1$: Change the model according to the OLS estimate $\widehat{\rho}$:

(2.1) if $\widehat{\rho} > 0.5$ and $\widehat{\rho} + 2\mathrm{SE}[\widehat{\rho}] > 0.9$, set $I_\Delta = 1$ and re-estimate Eq. (6) imposing $\rho = 1$;
(2.2) otherwise, if $\widehat{\rho} < 0$, set $I_r = 0$ and re-estimate Eq. (6).

(3) If $I_\Delta = 0$ and $T - k > 10$: Test whether the cumulated residuals have a zero mean using a t-test. If the zero mean is rejected using a significance level of 0.01, then add a trend to the model, $I_\tau = 1$, and re-estimate Eq. (6). Remove the trend again if $\widehat{\rho} < -0.5$.

(4a) If $I_\Delta = 0$: Forecasts are constructed in the standard way from the final version of Eq. (6).

(4b) If $I_\Delta = 1$: Construct the 90% confidence interval $(\widehat{\mu} - s, \widehat{\mu} + s)$ using $s = 1.645\widehat{\sigma}(T-1)^{-1/2}$, where $\widehat{\sigma}$ is the equation standard error of Eq. (6). Define $\widetilde{\mu} = \max(0, \widehat{\mu} - s)$ if $\widehat{\mu} > 0$, $\widetilde{\mu} = \min(0, \widehat{\mu} + s)$ otherwise. Replace $\widehat{\mu}$ with $\widetilde{\mu}$, but otherwise construct forecasts as standard.

6. Calibration

Calibration uses the averaged forecasts $\widetilde{x}_{T+h}$ to extend the series to $x_1,\ldots,x_T,\widetilde{x}_{T+1},\ldots,\widetilde{x}_{T+H}$. This has $T_c = T+H$ observations. The calibration model is similar to Eq. (6), but extended with additional autoregressive lags and a broken intercept and trend. There is no adjustment to reduce the trend or impose a unit root. The structure of the model is as follows:

$$x_t = \mu + [\text{autoregression}] + [S \text{ seasonality}] \\ + [S_2 \text{ seasonality}] + [\text{breaks}] + u_t,$$

where the inclusion of the components in square brackets depends on earlier decisions and the sample size. The fitted values of this model provide the updated forecasts $\widehat{\widetilde{x}}_{T+1},\ldots,\widehat{\widetilde{x}}_{T+H}$.

The full specification of the calibration model is given by:

$$\begin{aligned} x_t &= \mu + \rho x_{t-1} I_\rho + (\rho_R x_{t-R} + \rho_{R+1} x_{t-R-1}) I_R I_\rho I_4 \\ &\quad + \left\{\delta_j q_{j,t}\right\} I_A + \left(\gamma_1 S_{1t} + \gamma_1^* C_{1t}\right)(1-I_A) \\ &\quad + \left(\gamma_2 S_{2t} + \gamma_2^* C_{2t}\right)(1-I_3) + \rho_{SS_2} x_{t-SS_2} I_3 \\ &\quad + (\tau_1 d_t + \tau_2 t d_t I_5) I_6 + u_t, \quad t = T_0,\ldots,T_c, \end{aligned} \tag{7}$$

where the values of I_ρ, I_R and I_A are as at the start of the *Rho* procedure, i.e., based on the original x_t. We also introduce new indicators related to the sample size and frequency: $I_4 = T > 4S$, $I_3 = T_c > 3SS_2$ and $S_2 > 1$, $I_6 = 1$ when S is not 24 and $T_c - H > 3S$ and $T_c - k > 10$ (k is the number of regressors when $I_6 = 0$), and finally $I_5 = 1$ when $S = 4, 12, 13$ with $I_\rho = 1$. Also used is $S_{1t} = \sin[2\pi t/S]$, $C_{1t} = \cos[2\pi t/S]$. Some specifications use a broken intercept $d_t = I(t < T - \min[2S, T_c/2])$ or trend $D_t = td_t$. When $I_3 = 1$, the dependent variable is added with lag SS_2. We avoid losing observations by duplicating the first SS_2 observations before taking the lag.

As a first example, consider a quarterly series with dynamics, significant seasonality, and enough observations. This has the calibration model:

$$x_t = \mu + \rho x_{t-1} + \rho_4 x_{t-4} + \rho_5 x_{t-5} + \delta_1 q_{1,t} + \delta_2 q_{2,t} \\ + \delta_3 q_{3,t} + \tau_1 d_t + \tau_2 t d_t + u_t.$$

Hourly data uses lag six, and the most common specification is:

$$x_t = \mu + \rho x_{t-6} + \rho_{24} x_{t-24} + \rho_{25} x_{t-25} + \delta_1 q_{1,t} \\ + \cdots + \delta_{23} q_{23,t} + \rho_{168} x_{t-168} + u_t.$$

6.1. Further adjustments for hourly and weekly data

By default, the *Rho* and *Delta* forecasts are first averaged, then calibrated. However, a few additional changes are made at certain frequencies that have been found to improve the forecast performance. For hourly data, the *Rho* and *Delta* forecasts are calibrated, then averaged, then calibrated again. Hourly calibrations are done with an autoregressive lag of six, instead of one as in Eq. (7). For weekly data, *Rho* is applied to the four-weekly averages (giving a frequency of 13), and calibrated at the reduced frequency. These forecasts are calibrated along with the *Delta* forecasts, then averaged, and this average is then calibrated again.

7. Forecast confidence intervals

The accuracy of forecast intervals in M4 is assessed solely on the basis of the mean scaled interval score (MSIS). The penalty in MSIS consists of the width of the forecast interval together with forty times the amount that an outcome is outside the interval. We adopted a 90% interval because this performed marginally better than 95% for annual data in terms of MSIS. We discovered after submission that this is not the case at other seasonal frequencies, corresponding to the findings of Gneiting and Raftery (2007, p. 370), who show that the interval score is optimized at the true quantiles.

Our additional aim was to achieve a good coverage at each forecast horizon, not just on average. Starting with yearly data, we found that the standard forecast intervals were too wide at a horizon of one and too narrow at a horizon of six. In that case, the forecast variance grows with powers of $\widehat{\rho}^2$ as we go further ahead, but that narrows too quickly for M4. A better coverage was achieved by making the uncertainty grow with powers of $\widehat{\rho}$, where $\widehat{\rho} \geq 0$ is the estimated coefficient in the calibration model Eq. (7).

Our reported 90% forecast intervals are

$$\widehat{y}_{T+h} \pm c_S \widetilde{s}_h^F.$$

The scale factor s_C was found by minimizing the MSIS performance for the hold-back sample as $c_S = 1.9, 1.3, 1.3, 0.7, 0.9, 1.2$ for $S = 1, 4, 12, 52, 5, 24$. This calibration relied on our ability to run experiments quickly, which was helped by the way in which we organized our forecasting software environment.

The scale of the forecast uncertainty at horizon h is approximated by:

$$\begin{aligned} \widetilde{s}_h^F &= \widetilde{\sigma}_u \sum_{j=0}^{h+d} \widetilde{\rho}^j \qquad \text{where} \;\; d = -1 \; (S=1), \\ &\qquad d = 1 \; (S > 24), \;\; d = 0 \; (S \neq 24), \\ \widetilde{s}_h^F &= \widetilde{\sigma}_u \sum_{j=0}^{h+6} \widetilde{\rho}^{\lfloor j/6 \rfloor} \quad \text{if} \;\; S = 24. \end{aligned}$$

This in turn is based on the calibration residuals $\widehat{u}_t$:

$$\widetilde{\sigma}_u^2 = \sum_{T-T^*+1}^{T} \frac{\widehat{u}_t^2}{\max[T^* - k^*, 2]}, \\ k^* = 2 + 2I_4 I_\rho, \;\; T^* = \min(\max[SS_2, 80], T).$$

The degrees of freedom adjustment only counts the mean and the main autoregressive parameters. We take the 10% lower bound of the autoregressive coefficient:

$$\widehat{\rho}_L = \widehat{\rho} - 1.645 \widetilde{\sigma}_u \left(\left[X'X\right]_\rho^{-1}\right)^{1/2}$$

and restrict it to $\widetilde{\rho} = \min[\max[\widehat{\rho}_L, 0.6], 0.9]$.

While this method worked well with M4, it could be improved in two ways. The first would be to allow other intervals, especially 95%. The second would be to derive an approximation that works for all sample sizes, collapsing to the asymptotic interval as the sample size gets large. These issues are addressed by Castle, Doornik, and Hendry (2019).

8. Conclusions

We created the *Card* method for constructing forecasts and 90% forecast intervals for the M4 data. Producing 100,000 forecasts from start to finish takes less than half a minute on our desktop computer. Each series is handled in isolation, so there is no danger of making infeasible forecasts. The forecast accuracy is good, because we tend to overdifference, but then restrict the growth rate. Differencing changes a break into an outlier, and therefore acts as a form of robustification. On the other hand, breaks lead the autoregressive coefficient towards a unit root. Differencing is also used to remove the trend prior to seasonal smoothing.

The basic forecast models are adaptive but simple, preventing issues with forecasts running away. This is important because the evaluation is based on the mean performance rather than the median performance. The subsequent calibration allows for a richer seasonality and structure, which is particularly useful for weekly and hourly data (the daily data are mainly financial, and therefore much less susceptible to modelling). If hourly data have holiday effects, as with the electricity load, then it would be useful to extend *Card* to capture these annual features.

A companion paper, Castle et al. (2019), studies the properties of M4 and *Card* in more detail, and provides some further improvements to our methods.

Acknowledgments

Financial support from the Robertson Foundation, United States (award 9907422) and the Institute for New Economic Thinking, United States (grant 20029822) is gratefully acknowledged.

Appendix A. Seasonality

A.1. Testing seasonality

The data for seasonal frequency S can be presented in tabular form, organized by 'year'. Using the indices (i, j) to denote (year, period), write $y_{i,j}$, $i = 1, \dots, Y$, $j = 1, \dots, S$, in such a way that $y_{Y,S} = y_T$. Thus, by construction, the *last 'year'* is always complete. There are $Y = \lfloor T/S \rfloor$ complete years. An incomplete initial year is dropped in this representation, so that the first observation used is $T_s = 1 + T - SY$ and $y_{i,j} = y_{T_s+(i-1)S+j-1}$. From this, we can compute:

$$\bar{y}_{.,j} = \sum_{i=1}^{Y} y_{i,j}, \; j = 1, \dots, S,$$

which are the means for each season, and, similarly, the means for each year $\bar{y}_{i,.}$.

X-11, see Ladiray and Quenneville (2001), incorporates an ANOVA-based test for stable seasonality, which has the null hypothesis that the seasonal means are equal:

$$A\{y_t, S\} = \frac{S-1}{S(Y-1)} \frac{\mathrm{var}[\bar{y}_{.,j}]}{\mathrm{var}[y_{i,j}] - \mathrm{var}[\bar{y}_{.,j}]} \sim \mathsf{F}[S-1, S(Y-1)]. \tag{A.1}$$

Let x_t be the transformed variable after the logarithm decision (step 1 in Section 2.1). The test in Eq. (A.1) is applied to x_t if $I_\rho = 0$, and to Δx_t otherwise. Additive seasonality is assumed if the test rejects at the 10% level.

A.2. Seasonal autoregression

In the M4 benchmark methods, the seasonality decision is based on the Sth term in the ACF of y_t according to:

$$R\{y_t, S\} = T \frac{r_S^2}{1 + 2\sum_{j=1}^{S-1} r_i^2} \sim \chi^2(1). \tag{A.2}$$

The *Rho* method considers a potential seasonal lag length for Δx_t (or x_t if $I_\rho = 0$; by default, $x_t = \log y_t$, see Section 2.1). If the test in Eq. (A.2) on the Rth term in the ACF is significant at 10%, then $R = S$. Otherwise, R is set to the most significant lag up to S, provided that the corresponding test is significant at 1%. If neither condition holds, we set $I_R = 0$.

A.3. Extension for monthly data

For monthly data, we consider the possibility that there is seasonality at a lower frequency when none is detected at frequency $S = 12$. If both $A\{z_t, S\}$ and $R\{z_t, S\}$ are accepted at the 5% level, where z_t is the transformed variable, then we try all integer lags $s = S-1, \dots, S/2$. For the first s that has $A\{z_t, s\}$ rejected at the 0.1% level, we set $S = s$ and $I_A = 1$. If, in addition, $R\{z_t, s\}$ rejects at the 1% level, we set $R = s$ and $I_R = 1$.

A.4. Seasonal estimates in Delta method

Ignore the differencing, and, as an illustration, consider a quarterly time series x_t with T observations that is two more than a multiple of four. The seasonal table is given in Box I. The first column $x_{i,1}$ is smoothed with an $\mathrm{MA}_{3\times5}$, and the last value provides the seasonal estimate: $s^*_{Y,1} = (x_{Y-3,1} + 2x_{Y-2,1} + 3x_{Y-1,1} + 9x_{Y,1})/15$. The final estimate of the first quarter seasonal is to give this a zero mean by subtracting $(s^*_{Y,1} + s^*_{Y,2} + s^*_{Y,3} + s^*_{Y,4})/4$. The *Delta* method continues with Y observations $\bar{x}_{1,.}, \dots, \bar{x}_{Y,.}$ and the seasonals.

For quarterly data, there is also a small amount of smoothing with adjacent seasonals: $(s^*_{Y,j-1} + 6s^*_{Y,j} + s^*_{Y,j+1})/8$.

Appendix B. Forecasting with the *Delta* method, $I_\rho = 0$

The input consists of x_t, $t = 1, \dots, T$, S, H, I_A.

If $I_A = 0$, set $z_t = x_t$ and define $m(r) = \frac{1}{r}\sum_{t=T-r+1}^{T} z_t$, with $m(r) = \bar{z}$ for $r \geq T$. Use $\widetilde{S} = \max\{2, S\}$ to compute forecasts:

$$\widehat{x}_{T+1} = m(\widetilde{S}), \quad \widehat{x}_{T+h} = \frac{1}{2}\left[m(\widetilde{S}) + m(6\widetilde{S})\right], \quad h = 2, \dots, H.$$

If $I_A = 1$, handle seasonality as for $I_\rho = 0$, but using the seasonal table of x_t and an $\mathrm{MA}_{7\times5}$. Compute $m(r)$ from

	$Q1$	$Q2$	$Q3$	$Q4$	means
$i=0$	–	–	x_1	x_2	–
$i=1$	$x_{1,1}=x_3$	$x_{1,2}=x_4$	$x_{1,3}=x_5$	$x_{1,4}=x_6$	$\bar{x}_{1,.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i=Y$	$x_{Y,1}=x_{T-3}$	$x_{Y,2}=x_{T-2}$	$x_{Y,3}=x_{T-1}$	$x_{Y,4}=x_T$	$\bar{x}_{Y,.}$

Box I.

the 'annual' means z_t and forecasts:

$$\widehat{x}_{T+1} = m(1) + \widehat{s}_{Y1}, \quad \widehat{x}_{T+h} = \frac{1}{2}\left[m(1) + m(6)\right] + \widehat{s}_{Yh}, \quad h = 2, \ldots, H.$$

References

Assimakopoulos, V., & Nikolopoulos, K. (2003). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, *16*, 521–530.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2019). *Some forecasting principles from the M4 competition*: *Economics discussion paper 2019-W01*, Nuffield College, University of Oxford.

Doornik, J. A. (2013). *Object-oriented matrix programming using Ox* (7th ed.). London: Timberlake Consultants Press.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, W. R., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program (with discussion). *Journal of Business & Economic Statistics*, *16*, 127–177.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Ladiray, D., & Quenneville, B. (2001). *Seasonal adjustment with the X-11 method*. Berlin: Springer Verlag.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451–476.