Forecast Selection and Representativeness: Electronic companion

Appendix A. Derivations

Suppose the data generating process can be described by

$$y_t = c + t\beta + m_t\gamma + \epsilon_t, \tag{1}$$

where the variable m_t describes a series of binary variables, indicating an increase in y_t in every other period, while the term $t\beta$ indicates a linear trend.

Case 1. Assume $\beta = 0$ and $\gamma \neq 0$. Also assume we have two different forecasting models, F_1 and F_2. F_1 is misspecified by forecasting only a level, but F_2 is correctly specified by forecasting level and seasonality. Both forecasting models estimate their parameters without an estimation error, resulting in

$$F_1 : \hat{c}_1 = c + \frac{\gamma}{2},$$
$$F_2 : \hat{c}_2 = c,\ \hat{\gamma}_2 = \gamma.$$

Forecasts made in period t for period $t+1$ are accordingly

$$f_{1,t+1} = c + \frac{\gamma}{2},$$
$$f_{2,t+1} = c + m_{t+1}\gamma.$$

The residuals sum of squares for model F_1 can be written as

$$\begin{aligned}\mathrm{RSS}_1 &= \sum_{t=1}^{n} (y_t - f_{1,t})^2 \\ &= \sum_{t=1}^{n} \epsilon_t^2 + \gamma \sum_{t=1}^{n} \epsilon_t (-1)^t + n\frac{\gamma^2}{4}.\end{aligned}$$

The expectation of the RSS for model F_1 is $\mathbb{E}[\mathrm{RSS}_1] = n\sigma^2 + n\frac{\gamma^2}{4}$ and because model F_1 has only one parameter ($k=1$), we can write $\mathbb{E}[\mathrm{AIC}_c^1] = n \ln\left(\sigma^2 + \frac{\gamma^2}{4}\right) + 2 + \frac{4}{n-2} - 2C$. A similar analysis for model F_2 leads to $\mathbb{E}[\mathrm{AIC}_c^2] = n\ln(\sigma^2) + 4 + \frac{12}{n-3} - 2C$. Selecting by AIC_c will lead to the correct model only if $\mathbb{E}[\mathrm{AIC}_c^1] > \mathbb{E}[\mathrm{AIC}_c^2]$, or

$$\ln\left(1 + \frac{\gamma^2}{4\sigma^2}\right) > \frac{2(n-1)}{(n-2)(n-3)}.$$

Case 2. Assume $\beta \neq 0$ and $\gamma = 0$. Also assume we have two different forecasting models, F_1 and F_2. F_1 is misspecified by forecasting only a level, but F_2 is correctly specified by forecasting level and trend. After T observations, both forecasting models estimate their parameters without an estimation error, resulting in

$$F_1 : \hat{c}_1 = c + \frac{T\beta}{2},$$
$$F_2 : \hat{c}_2 = c,\ \hat{\beta}_2 = \beta.$$

Forecasts made in period t for period $t+1$ are accordingly

$$\begin{aligned} f_{1,t+1} &= c + \frac{t\beta}{2}, \\ f_{2,t+1} &= c + (t+1)\beta. \end{aligned}$$

The residuals sum of squares for model F_1 can be written as

$$\begin{aligned} \mathrm{RSS}_1 &= \sum_{t=1}^{n} \left(c + t\beta + \epsilon_t - c - \frac{n\beta}{2} \right)^2 \\ &= \sum_{t=1}^{n} \epsilon_t^2 - n\beta \sum_{t=1}^{n} \epsilon_t + 2\beta \sum_{t=1}^{n} t\epsilon_t + \frac{n\beta^2(n^2+2)}{12}. \end{aligned}$$

The expectation of the RSS for model F_1 is $\mathbb{E}[\mathrm{RSS}_1] = n\sigma^2 + n\frac{\beta^2(n^2+2)}{12}$ and because model F_1 has only one parameter ($k=1$), we can write $\mathbb{E}[\mathrm{AIC}_c^1] = n\ln\left(\sigma^2 + \frac{\beta^2(n^2+2)}{12}\right) + 2 + \frac{4}{n-2} - 2C$. A similar analysis for model F_2 leads to $\mathbb{E}[\mathrm{AIC}_c^2] = n\ln(\sigma^2) + 4 + \frac{12}{n-3} - 2C$. Selecting by AIC_c will lead to the correct model only if $\mathbb{E}[\mathrm{AIC}_c^1] > \mathbb{E}[\mathrm{AIC}_c^2]$, or

$$\ln\left(1 + (n^2+2)\frac{\beta^2}{12\sigma^2}\right) > \frac{2\,(n-1)}{(n-2)\,(n-3)}.$$

Both sides of the inequality depend on n, but the left side is strictly increasing in n, and the right side is strictly decreasing in n (for $n > 3$).

Case 3. Assume $\beta, \gamma \neq 0$. Also assume we have two different forecasting models, F_1 and F_2. F_1 is misspecified by forecasting only level and trend, but F_2 is correctly specified by forecasting level, trend, and a seasonal change with the according periodicity. Both forecasting models perfectly estimate their parameters without error, resulting in

$$F_1 : \hat{c}_1 = c + \frac{\gamma}{2},\ \hat{\beta}_1 = \beta,$$
$$F_2 : \hat{c}_2 = c,\ \hat{\beta}_2 = \beta,\ \hat{\gamma}_2 = \gamma.$$

Forecasts made in period t for period $t+1$ are accordingly

$$\begin{aligned} f_{1,t+1} &= c + \frac{\gamma}{2} + (t+1)\beta, \\ f_{2,t+1} &= c + (t+1)\beta + m_{t+1}\gamma. \end{aligned}$$

The residuals sum of squares for model F_1 can be written as

$$\begin{aligned}\text{RSS}_1 &= \sum_{t=1}^{n}\Big(c+t\beta+m_t\gamma+\epsilon_t-c-\frac{\gamma}{2}-t\beta\Big)^2\\ &= \sum_{t=1}^{n}\epsilon_t^2+\gamma\sum_{t=1}^{n}\epsilon_t(-1)^t+n\frac{\gamma^2}{4}.\end{aligned}$$

The expectation of the RSS for model F_1 is $\mathbb{E}[\text{RSS}_1]=n\sigma^2+n\frac{\gamma^2}{4}$ and because model F_1 has two parameters ($k=2$), we can write $\mathbb{E}[\text{AIC}_c^1]=n\ln\left(\sigma^2+\frac{\gamma^2}{4}\right)+4+\frac{12}{n-3}-2C$. A similar analysis for model F_2 leads to $\mathbb{E}[\text{AIC}_c^2]=n\ln(\sigma^2)+6+\frac{24}{n-4}-2C$. Selecting by AIC_c will lead to the correct model only if $\mathbb{E}[\text{AIC}_c^1]>\mathbb{E}[\text{AIC}_c^2]$, or

$$\ln\left(1+\frac{\gamma^2}{4\sigma^2}\right)>\frac{2\,(n-1)}{(n-3)\,(n-4)}.$$

Case 4. Assume $\beta,\gamma\neq 0$. Also assume we have two different forecasting models, F_1 and F_2. F_1 is misspecified by only forecasting level and seasonality only, but F_2 is correctly specified by forecasting level, trend, and a seasonal change with the according periodicity.

For our further considerations, we draw upon the following simplifications: $\sum_{t=1}^{T}t=\frac{T(T+1)}{2}$, $\sum_{t=1}^{T}m_t=\frac{T}{2}$, and $\sum_{t=1}^{T}tm_t=\frac{T}{2}\left(\frac{T}{2}+1\right)$. The latter two simplifications assume $m_1=0$ and $m_T=1$. Omitting the trend in F_1 leads to parameter bias in both the level and the seasonality component. To see this, consider that for data without error, the estimators for the level and seasonality in F_1 need to satisfy

$$\underset{\hat{c}_1,\hat{\gamma}_1}{\operatorname{argmin}}\, RSS=\sum_{t=1}^{T}(c+t\beta+m_t\gamma-\hat{c}_1-m_t\hat{\gamma}_1)^2$$

Therefore,

$$\frac{\partial RSS}{\partial\hat{c}_1}=-2\sum_{t=1}^{T}(c+t\beta+m_t\gamma-\hat{c}_1-m_t\hat{\gamma}_1)=0\Rightarrow$$

$$c+\beta\frac{T+1}{2}+\frac{\gamma}{2}-\frac{\hat{\gamma}_1}{2}=\hat{c}_1,$$

and

$$\frac{\partial RSS}{\partial\hat{\gamma}_1}=-2\sum_{t=1}^{T}m_t(c+t\beta+m_t\gamma-\hat{c}_1-m_t\hat{\gamma}_1)=0\Rightarrow$$

$$c+\beta\left(\frac{T}{2}+1\right)+\gamma-\hat{c}_1=\hat{\gamma}_1\Rightarrow$$

$$\hat{\gamma}_1=\beta+\gamma.$$

Therefore,

$$\hat{c}_1=c+\beta\frac{T}{2}.$$

After T observations, both forecasting models estimate their parameters without an estimation error, resulting in

$$F_1 : \hat{c}_1 = c + \beta\frac{T}{2},\ \hat{\gamma}_1 = \gamma + \beta,$$
$$F_2 : \hat{c}_2 = c,\ \hat{\beta}_2 = \beta,\ \hat{\gamma}_2 = \gamma.$$

Forecasts made in period t for period $t+1$ are accordingly

$$\begin{aligned} f_{1,t+1} &= c + \frac{t\beta}{2} + m_{t+1}(\gamma+\beta), \\ f_{2,t+1} &= c + (t+1)\beta + m_{t+1}\gamma. \end{aligned}$$

The residuals sum of squares for model F_1 can be written as

$$\begin{aligned} \mathrm{RSS}_1 &= \sum_{t=1}^{n}\left(c + t\beta + m_t\gamma + \epsilon_t - c - \frac{n\beta}{2} - m_t(\gamma+\beta)\right)^2 \\ &= \sum_{t=1}^{n}\left(\epsilon_t - \beta\frac{n-2t-2m_t}{2}\right)^2 \\ &= \sum_{t=1}^{n}\epsilon_t^2 - n\beta\sum_{t=1}^{n}\epsilon_t + 2\beta\sum_{t=1}^{n}t\epsilon_t + 2\beta\sum_{t=1}^{n}m_t\epsilon_t + n\beta^2\frac{(n^2+20)}{12}. \end{aligned}$$

The expectation of the RSS for model F_1 is $\mathbb{E}[\mathrm{RSS}_1] = n\sigma^2 + n\beta^2\frac{(n^2+20)}{12}$ and because model F_1 has two parameters ($k=2$), we can write $\mathbb{E}[\mathrm{AIC}_c^1] = n\ln\left(\sigma^2 + \frac{\beta^2(n^2+20)}{12}\right) + 4 + \frac{12}{n-3} - 2C$. A similar analysis for model F_2 leads to $\mathbb{E}[\mathrm{AIC}_c^2] = n\ln(\sigma^2) + 6 + \frac{24}{n-4} - 2C$. Selecting by AIC_c will lead to the correct model only if $\mathbb{E}[\mathrm{AIC}_c^1] > \mathbb{E}[\mathrm{AIC}_c^2]$, or

$$\ln\left(1 + (n^2+20)\frac{\beta^2}{12\sigma^2}\right) > \frac{2\,(n-1)}{(n-3)\,(n-4)}.$$

Similar to Case 2, both sides of the inequality depend on n, but the left side is strictly increasing in n, and the right side is strictly decreasing in n (for $n > 4$).

Appendix B. Simulation set-up

We use the process described in equation 6 of section 3.3 in a simulation setting to examine the (in)ability of AIC_c, CV, and REP to select the better of the two models across five cases, as presented in table 1. In the first case, we assumed a data generating process with $\beta = 0$ and $\gamma = 10$ (no trend but deterministic seasonality). The first model is Simple Exponential Smoothing (SES), a model that is able to estimate only the level of a series. The second model is an exponential smoothing model with additive seasonality and no trend. This second model estimates the level and the seasonality of a signal through separate equations (components). The second model is correctly specified when $\gamma \neq 0$, whereas the first model is correctly specified when $\gamma = 0$ (i.e., there is no seasonality). Case 2 focuses on the data generating process with $\beta = 2$ and $\gamma = 0$ (trend but no seasonality). The misspecified model is the same as in Case 1, whereas the correctly specified model is the Holt's linear trend exponential smoothing model. Cases 3 and 4 examine the case of trend and seasonality, that is, $\beta = 2$ and $\gamma = 10$. In both cases, the correctly specified model is the Holt-Winters exponential smoothing model, which includes a level, a trend, and a seasonal component. The misspecified models are missing the components for estimating trend and seasonality respectively. Finally, Case 5 assumes a data generating process with neither trend nor seasonality, aiming to examine if the various selection criteria are susceptible to accidentally selecting more complex models.

Table 1 Models in the simulation setting.

Case	Misspecified model	Correctly specified model
Case 1 $\beta = 0$, $\gamma = 10$	SES Level only ets(model="ANN", ...)	SES with seasonality Level and seasonality ets(model="ANA", ...)
Case 2 $\beta = 2$, $\gamma = 0$	SES Level only ets(model="ANN", ...)	Holt's linear trend Level and linear trend ets(model="AAN", damped=F, ...)
Case 3 $\beta = 2$, $\gamma = 10$	Holt's linear trend Level and linear trend ets(model="AAN", damped=F, ...)	Holt-Winters Level, linear trend and seasonality ets(model="AAA", damped=F, ...)
Case 4 $\beta = 2$, $\gamma = 10$	SES with seasonality Level and seasonality ets(model="ANA", ...)	Holt-Winters Level, linear trend and seasonality ets(model="AAA", damped=F, ...)
Case 5 $\beta = 0$, $\gamma = 0$	Holt-Winters Level, linear trend and seasonality ets(model="AAA", damped=F, ...)	SES Level only ets(model="ANN", ...)

Note that in all five of these cases, none of the models with correctly specified forms is strictly optimal because the model parameters are estimated from the data. Also, the estimates for the level, trend, and seasonal components are updated every period based on the formulation of the exponential smoothing models, so the out-of-sample and in-sample forecasts will be different. As a result, (a) REP and CV can select an incorrect model, and (b) the selections via REP and CV can differ. We used the *forecast* package for the R statistical software to produce the forecasts. We present the specific functions used in the third row of each case within table 1 of electronic Appendix A.

For each of the cases described in table 1, we simulated series with $c = 100$. We varied the length of the series in such a way that $n \in \{20, 80\}$, and noise so that $\sigma \in \{0, 1, \ldots 50\}$. We generated 1,000 series for each possible combination of n and σ. The results are presented in figure 1. We calculated CV and REP assuming a $\mathcal{L}_1$, and a forecast horizon of 10 periods ahead. We set the discounting factor to our standard of $\delta = 0.5$. The results for each sample size are presented in a separate panel. The horizontal axis of each panel refers to the standard deviation of the noise term of the process. The vertical axis shows the percentage of the correct selections for each criterion. For example, in Case 1, the vertical axes show how often each criterion selects SES with seasonality.

In general, REP performs well. When the data generating process has a seasonal term but the misspecified model lacks a seasonal component (Cases 1 and 3), REP outperformed AIC$_c$ and CV in selecting the correct model. This result is especially true for shorter time series. When the process exhibited a trend and the misspecified model was unable to capture this trend (Cases 2 and 4), REP also outperformed the other two selection criteria for shorter time series. As the length of the series increased, the difference between the three criteria became negligible. These simulation results confirmed our earlier theoretical insights: When the data generation process contains a trend and/or seasonality, but also a lot of noise, selection by REP or CV will outperform selection by IC.

Finally, we simulated series with $\beta = \gamma = 0$ (Case 5; last two panels of figure 1) to consider the case in which trend and seasonality do not exist. In this case, SES is the correctly specified model. We measure the percentage of cases in which each criterion selects the level-only model over the level and seasonal model (Holt-Winters). When a trend or seasonal signal did not exist, then REP and CV tended to select the incorrect (more complex) model more often than AIC$_c$ did. Information criteria selection benefits from the complexity penalty in this instance. The performance of CV and REP improves as the sample size increases, but it is still worse than selection by AIC$_c$.

We have also tested whether AIC selects differently compared with AIC$_c$ in these cases. When a signal exists and the noise is high, AIC will also result in a much lower percentage of correct selections than either CV or REP. Although AIC selects slightly better than AIC$_c$ when n is small, the differences between the two IC decreases as the sample size increases.

In practice, we would expect smaller differences between REP and the two other criteria, CV and AIC$_c$. Real-life series do not consist of deterministic signals, like the one that we assumed in this section. Changing data patterns may render the identification of the correctly specified model more challenging for any criterion. The discounting factor, δ, in the proposed selection criterion REP may be critical when dealing with more stochastic processes.

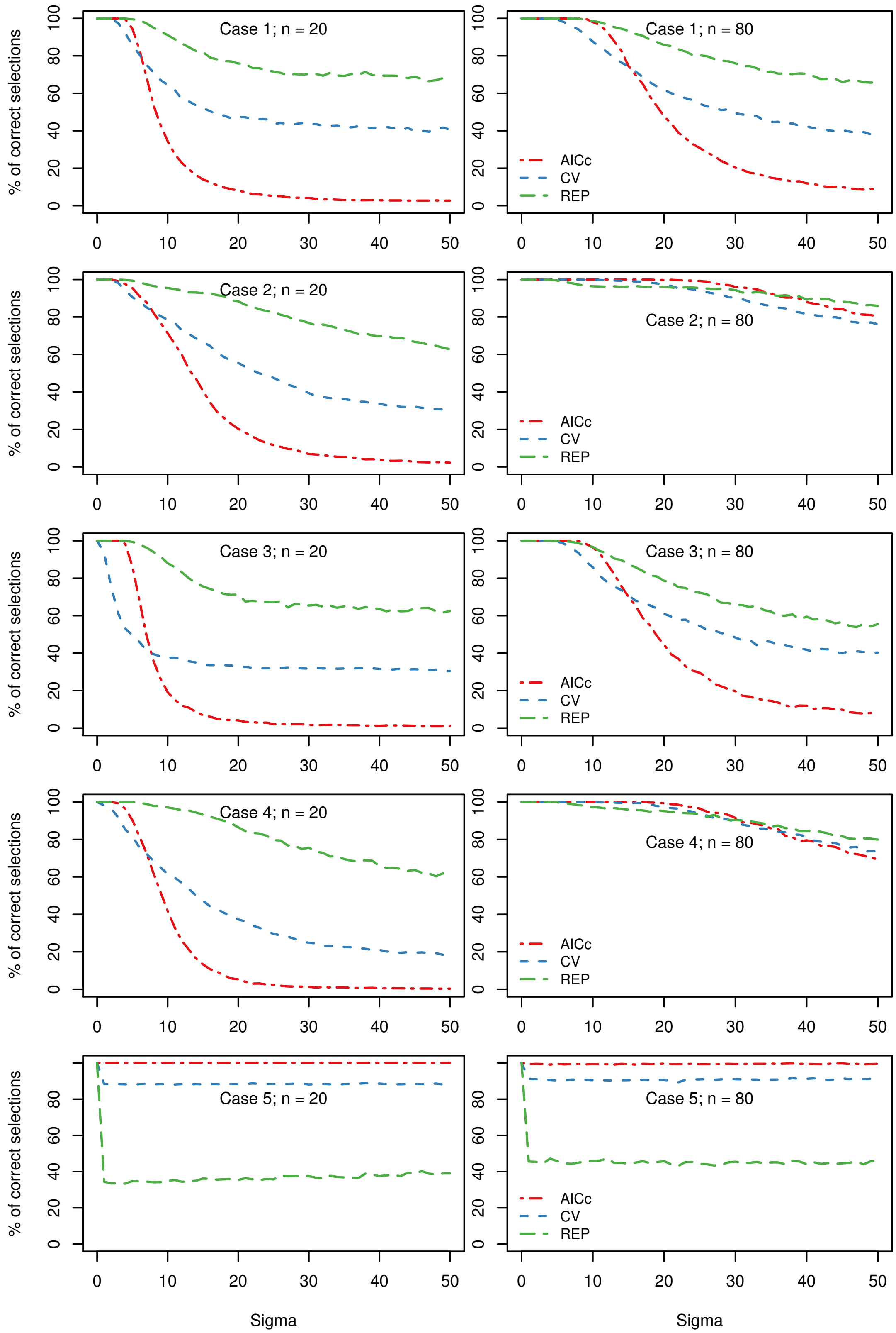

Figure 1 **Simulation results for the five cases considered in table 1.**

Appendix C. Empirical data

Table 2 The number of series, their lengths, and the forecasting horizon for each data frequency.

Frequency	Number of series	In-sample observations					s	h
		Min	Q1	Q2	Q3	Max		
Yearly	23826	9	20	29	40	835	1	6
Quarterly	24959	10	58	85	114	866	4	8
Monthly	50045	30	82	196	306	2794	12	18
Weekly	359	80	379	934	1603	2597	52	13
Daily	4227	93	323	2940	4197	9919	7	14
Hourly	414	700	700	960	960	960	168	48
Total	103830							