Online Technical Notes

This is the online technical notes to the article: Syntetos, A and Wang, S. (2026). Categorical Forecasting for Garage Management. *FORESIGHT: The International Journal of Applied Forecasting*.

TECHNICAL NOTE A: PREDICTOR VARIABLES

We elaborate the definitions of the selected predictors from four categories, namely vehicle condition, manufacturing, geographical, and calendar.

Vehicle Condition Variables

- Age_{i+1}: age at t_{i+1}
- M_{i+1}: mileage at t_{i+1}
- $AvgM_{i+1}$: average mileage per year at t_{i+1}
- $PorC_{i+1}$: preventive or corrective maintenance at t_{i+1}
- $isAcc_{i+1}$: whether it is due to an accident at t_{i+1}
- N_i^P: total number of preventive maintenance jobs up to t_i
- N_i^C : total number of corrective maintenance jobs up to t_i
- CRT_i^P : cumulative repair time due to preventive maintenance up to t_i
- CRT_i^C : cumulative repair time due to corrective maintenance up to t_i
- $PorC_i$: preventive or corrective maintenance at t_i
- RT_i : repair time at t_i
- $isPart_i$: whether any parts are replaced at t_i
- TBF_{i+1}: time between failures between t_i and t_{i+1}
- MBF_{i+1}: mileage between failures between t_i and t_{i+1}

Manufacturing Variables

- $Make$: vehicle make
- $Model$: vehicle model
- MY: vehicle model year
- $Type$: vehicle type

Geographical Variables

- $Garage_{i+1}$: which garage the vehicle is assigned to between t_i and t_{i+1}

- $Urban_{i+1}$: whether the affiliated garage is situated in an urban area
- $Seaside_{i+1}$: whether the affiliated garage is situated by the seaside
- $Region_{i+1}$: which region the affiliated garage is located in

Calendar Variables

- $Year_{i+1}$: number of years after 2010 at t_{i+1}
- $Month_{i+1}$: month of the year at t_{i+1}
- $Weekend_{i+1}$: whether it is a weekday or in a weekend at t_{i+1}

TECHNICAL NOTE B: ORDINAL REPAIR TIMES FORECASTING METHOD

Suppose that each observation $y_i,\ i = 1, \dots, n$, belongs to one of the ordinal categories $k = 1, \dots, K$, and $\boldsymbol{x}_i = \left(x_1, \dots, x_p\right)^{\mathrm{T}}$ represents a p-dimensional vector containing the predictor variables; we model the logit of the conditional cumulative probability

$$\begin{aligned} logit(P(y_i \leq k)|\boldsymbol{X}_i = x_i) &= \log\left(\frac{P(y_i \leq k)}{1 - P(y_i \leq k)}\middle|\ \boldsymbol{X}_i = \boldsymbol{x}_i\right) \\ &= \beta_{0,k} + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon_i, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta} = \left(\beta_{0,k}, \beta_1, \dots, \beta_p\right)$ is the set of unknown parameters to be estimated and ε_i is the error term. In our case, there are three categories (minor, medium, major), and we only need two cumulative probabilities to obtain the full probability of the three categories.

The key idea of LASSO is to maximize the likelihood function subject to the sum of the absolute value of the coefficients being less than a constant. By imposing such constraint, the estimated parameters are shrinking and some of them tend to be exactly zero, which then serves the purpose of variable selection. The direct advantage of LASSO is the reduction of the variance of the estimated value and the increase of the accuracy of the regression prediction. Meanwhile, the resulting model is parsimonious and hence tends to be more interpretable.

Technically, the LASSO estimator resolves the ℓ_1-penalized problem of estimating parameters $\boldsymbol{\beta}$ by maximizing the likelihood of the ordinal logit model, $\ell(\boldsymbol{\beta}|y_i, \boldsymbol{x}_i)$, subject to the constraint $\sum_{j=1}^{p}\left|\beta_j\right| \leq s$, as shown in Equation (2).

$$\widehat{\boldsymbol{\beta}} = argmax_{\boldsymbol{\beta}}\left(\ell(\boldsymbol{\beta}|y_i, \boldsymbol{x}_i) - \lambda \sum_{j=1}^{p}\left|\beta_j\right|\right), \tag{2}$$

where λ is a tuning parameter to control the strength of shrinkage. Intuitively, a larger value of λ leads to a stronger penalization on the sum of absolute values of estimated parameters, which shrinks the values closer to zero. If λ is beyond a threshold value, then some of the estimated parameters are forced to zero, which is equivalent to leaving the corresponding predictor variables out of the model. Keep increasing the value of λ beyond the threshold leaves out more predictor variables, but it may also suffer from loss of predictive power. Thus, the value of λ should be carefully chosen. In this study, we use a 5-fold cross-validation technique to choose an optimal value of λ. This is based on the fact that cross-validation is intuitively appealing and can provide a good estimate of the expected forecasting error (Hastie et al., 2009, Chapter 7). It should be highlighted that the predictor variables are at different scales. To fairly select variables by LASSO, it is necessary to rescale all predictor variables at the same level. In such way, we give the same importance to all variables, rather than giving high weights to ones at small scales (i.e., larger magnitude of parameter values). Thus, all predictor variables are standardized before applying LASSO.

TECHNICAL NOTE C: FORECASTING PERFORMANCE EVALUATION

The Brier score (BS) is one of the standard metrics to assess and compare probability forecasts for unordered categories. It is a quadratic rule defined as:

$$BS = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{f}_k - o_k\right)^2 \tag{3}$$

where K is the number of possible categories, $\hat{f}_k$ is the forecasted probability for category k, and o_k takes the value 1 or 0, according to whether the true category is category k or not. The range of BS is between 0 and 1. A lower BS indicates a better forecast, and a perfect forecast has BS of 0. As the BS measures only one observation, it is common to report the average BS over a given number of forecasted observations, denoted as $\overline{BS}$.

The ranked probability score (RPS) is a strictly proper scoring rule that considers the ordering of events by assigning higher scores for assessments if higher predicted probabilities are given for events close to the actual event. The RPS is also a quadratic rule computed by:

$$RPS = \frac{1}{K-1}\sum_{k=1}^{K}\left(\sum_{s=1}^{k}\hat{f}_k - \sum_{s=1}^{k} o_k\right)^2. \tag{4}$$

Similar to BS, RPS is also in the range of 0 and 1, and a better forecast is associated with a lower RPS. In the special case of only two categories, the RPS is equivalent to the BS. Again, we report the average RPS over a given number of forecast observations, denoted as $\overline{RPS}$. Further, and when the task is to evaluate probabilistic forecasts in comparison with those produced by another method, skill scores may be particularly useful. A skills score is associated with a particular scoring rule, and amongst many of them, the Brier skill score (BSS) and the ranked probability skill score (RPSS) are widely used to quantify improvement over a reference method (Weigel et al., 2007). The BSS and the RPSS are defined as:

$$BSS = 1 - \frac{\overline{BS}}{\overline{BS}_{ref}} \tag{5}$$

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS}_{ref}} \tag{6}$$

where $\overline{BS}_{ref}$ and $\overline{RPS}_{ref}$ correspond to the average BS and the average RPS of a chosen reference method. The range of BSS and $RPSS$ is from minus infinity to 1. 0 indicates no skill comparing to the reference method, while 1 indicates a method with perfect skill. Positive values of BSS and $RPSS$ indicate a more skilled method with respect to the reference one, while negative values suggesting a less skilled method.

TECHNICAL NOTE D: DECISION-MAKING UNDER PROBABILITIC FORECSTING

Denote the manager's decision as d and the actual category as k. Recall that $d, k \in \{minor, medium, major\}$ in our context and the total number of categories $K = 3$. A loss function $\mathfrak{L}(d, k)$ is introduced to quantify the cost of incorrect classification, which is defined as follows:

$$\mathfrak{L}(d,k) = \begin{cases} 0 & if\ d = k \\ \ell_{dk} & if\ d \neq k \end{cases} \tag{7}$$

where $\ell_{dk} > 0$. The probabilistic forecasting model, such as the ordinal logit model, provides the predicted probabilities associated with each category, denoted as $\hat{f}_k$, where $\sum_1^K \hat{f}_k = 1$. The cost of a decision on each category is a random variable denoted as ℓ_d and its expectation is

$$E(\ell_d) = \sum_{k=1}^{K} \hat{f}_k \ell_{dk} \tag{8}$$

In line with Taylor and Jeon (2018), the objective of rational decision making is to minimize the (long run) expected cost. Thus, the optimal decision based on the loss function associated with the predicted probabilities is shown below.

$$d^{\ell,*} = \arg\min_d \sum_{k=1}^{K} \hat{f}_k \ell_{dk} \tag{9}$$

The traditional way to make decisions by point forecasting is expressed as follows:

$$d^{f,*} = \arg\max_d \hat{f}_d \tag{10}$$

By combining the loss function and the predicted probabilities, the optimal decision $d^{\ell,*}$ is not always the same as the decision $d^{f,*}$.