University of Reading

Improving automotive garage operations by categorical forecasts using a large number of variables

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Wang, S. ORCID: https://orcid.org/0000-0003-2113-5521, Syntetos, A. A., Liu, Y., Di Cairano-Gilfedder, C. and Naim, M. M. (2023) Improving automotive garage operations by categorical forecasts using a large number of variables. European Journal of Operational Research, 306 (2). pp. 893-908. ISSN 0377-2217 doi: 10.1016/j.ejor.2022.06.062 Available at https://centaur.reading.ac.uk/106019/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.ejor.2022.06.062

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

European Journal of Operational Research 306 (2023) 893–908

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

Improving automotive garage operations by categorical forecasts using a large number of variables

Shixuan Wang[a], Aris A. Syntetos[b,*], Ying Liu[c], Carla Di Cairano-Gilfedder[d], Mohamed M. Naim[b]

[a] Department of Economics, University of Reading, Reading, RG6 6EL, United Kingdom
[b] PARC Institute of Manufacturing, Logistics & Inventory, Cardiff University, Cardiff, CF10 3EU, United Kingdom
[c] Institute of Mechanical and Manufacturing Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, United Kingdom
[d] BT Technology Applied Research, Ipswich, IP5 3RE, United Kingdom

ARTICLE INFO

Article history:
Received 9 August 2021
Accepted 30 June 2022
Available online 7 July 2022

Keywords:
Forecasting
Maintenance
Repair time
LASSO
Automotive garage

ABSTRACT

Cost effective job scheduling for garage management relies upon assigning repair times into appropriate categories rather than using exact repair time lengths. In this paper, we employ an ordinal logit model with least absolute shrinkage and selection operator (LASSO) to forecast such repair time categories for automotive engines. Our study is based on a unique dataset of maintenance records from the network of 64 UK garages of *BT Fleet Solutions*, and we consider a large number of predictor variables, with condition, manufacturing, geographical, and calendar-related information. The application of LASSO enables the identification of relevant predictor variables for forecasting purposes. Based on the Brier score and the ranked probability score (and their skill scores), we document substantial predictive ability of our method which outperforms five benchmarks, including the method used by the company. More importantly, we demonstrate explicitly how to associate the predicted probabilities with a loss function in order to make operational decisions in garages. We find that the best choice of job scheduling does not always correspond to the predicted categories, especially when the loss function is asymmetric. We show that scheduling jobs on the basis of our method can help the company reduce loss value. Finally, we identify opportunities for further improvements in the operations of the company and for garage maintenance operations in general.

© 2022 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

Operational research (OR) work in the automotive industry mostly focuses on demand and sales forecasting (Murry & Zhou, 2020), production systems (Volling, Matzke, Grunewald & Spengler, 2013), remanufacturing (Schultmann, Zumkeller & Rentz, 2006), order-to-delivery processes (Brabazon & MacCarthy, 2017), and supply chain management (Zhang, Shang & Li, 2011). Less attention has been paid to automotive garage (repair shop) management which is still mainly performed by humans relying on judgmental decision making. Academic work in this area does not reflect the size and importance of the automotive maintenance industry. For instance, the annual repair bill related to the US federal fleet of 600,000 vehicles is estimated to be $1B (US General Services Administration, 2015). There appears to be some gap in the literature when it comes to innovative applications of OR to garage management.

In automotive garage management, one of the most important operations is to make predictions of repair times[1] (also known as Time to Repair, TTR), which is crucially related to scheduling and planning (Pilar, e Silva & Borges, 2021). Meanwhile, garage customers also need the information of repair times for their own planning purposes (Ryan, 2015). Repair times that are longer than accounted for can cause disruption to the garage's job scheduling and the availability offered to customers. On the contrary, shorter repair times induce costs related to idle time and rearrangement of jobs. In practice, all too often garage managers make judgmental forecasts of the repair times based on a quick inspection. Such judgmental forecasts may not be consistent over time, and their accuracy highly depends on experience and skills

* Corresponding author.
E-mail address: SyntetosA@cardiff.ac.uk (A.A. Syntetos).

[1] The repair time excludes the inspection time, waiting time, and logistics time; only the time of repair actions is considered.

https://doi.org/10.1016/j.ejor.2022.06.062
0377-2217/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

(Goodwin, Gonul & Onkal, 2019; Lawrence, Goodwin, O'Connor & Onkal, 2006). If repair times of automotive parts can be accurately predicted based on some formal approach, garage managers may considerably improve the efficiency of job scheduling and better inform customers about their waiting times.

To this end, we propose an ordinal logit model with the least absolute shrinkage and selection operator (LASSO) to forecast the ordinal categories of repair times for automotive parts in order to improve automotive garage operations. The uniqueness of our work lies on the categorical prediction of repair times, rather than precise time lengths because of two reasons. First, small differences on the exact predicted time length have negligible impact on operational decisions. In terms of job scheduling, repair jobs are typically 'bucketed' into categories for practical purposes (Lewandowski & Olszewska, 2020). Second, the records of repair times are subject to measurement errors which further substantiates the rationale for relying upon categories rather than exact time lengths. Based on consultation with experts, we classify automotive engine repair times in this study as: minor (less than 0.5 hour), medium (between 0.5 hour and 2 hours), and major (more than 2 hours). Such classification is also consistent with the classical planning and control principle of "runners, repeaters, and strangers" (Aitken, Childerhouse & Towill, 2003). The minor category corresponds to the "runners" which are the frequent routine tasks with standardized procedures; the medium category matches the "repeaters" which contain less frequent tasks but with relatively well understood resource requirements; and the major category resembles the "strangers" which are uncertain tasks associated with limited insight.

We consider a large number of predictor variables, with condition, manufacturing, geographical, and calendar-related information. To tackle the high dimensionality problem, we employ LASSO to regularize the estimation results and automatically select the most informative predictor variables. Forecasting performance is examined by means of utilizing an empirical dataset of repair times for automotive engines, provided by our industrial partner, *BT Fleet Solutions* (which has since been restructured and is now operated by a different company), which is a sizeable fleet company in the United Kingdom. We compare our method with five benchmarks: k-nearest neighbors, deep learning, random guess, sample majority, and historical averages.[2] Our method consistently outperforms the benchmarks in the test sample based on the Brier score and the ranked probability score (and their skill scores). The superior predictive power is also confirmed by robustness checks. More importantly, we explicitly demonstrate how to use our proposed method to: *i*) support garage managers towards more efficient job scheduling; and *ii*) facilitate some better management of the garages.

The contribution of this study is two-fold. First, our study provides an innovative OR application to forecast categories of repair times for automotive engines by LASSO using a large number of predictor variables. Based on a sizeable dataset over eight years, our ordinal logit model with LASSO method shows superior predictive power and consistently outperforms five benchmarks. Second, we further contribute to the OR field by means of utility assessment, i.e. by linking the predicted probabilities to operational decision making in garages in order to reduce loss value and enhance efficiency. We find that best decisions do not always correspond to point forecasts, especially if we deal with asymmetric loss functions. Our framework applied to automotive engines, can be naturally generalized to other contexts (such as vessels and aircrafts).

[2] While confidentiality agreements preclude disclosing the exact method used by the case company, we can reveal that their method is one of the five benchmark methods considered in this study.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss the related literature. In Section 3, we describe our empirical data and explain the choice of predictor variables. Section 4 presents a description of our methodology, including the forecasting method, benchmarks, and measures of forecasting evaluation. The empirical forecasting performance, and some associated robustness checks are provided in Section 5. Section 6 demonstrates how to use probabilistic forecasting to improve garage management in practice. Section 7 provides a summary of this study and discusses potential extensions for future research. There is also a Supplement that provides additional results of robustness checks.

2. Research background

This study sits in the intersection of three strands of literature: *i*) maintainability analysis; *ii*) application of LASSO in OR; *iii*) probabilistic forecasting for decision making. We discuss below the most relevant studies in each of these strands.

Maintainability analysis is concerned with the amount of time required to repair a system when it fails. Empirical studies dedicated to forecasting repair times are limited, and most of the relevant work is designed around very specific systems. Keizers, Bertrand and Wessels (2003) used a logistic regression to investigate the repair performance for the Royal Netherlands Navy. They selected eight predictor variables related to planning behavior, and found five of them to be significant at 10% level. Based on the concept of the proportional hazard model, Gao, Barabady and Markeset (2010) developed a proportional repair model (PRM) to analyze repair rate (i.e. the reciprocal of repair time) by using explanatory variables. In the PRM, the repair rate function is a product of the baseline repair function and a term incorporating the effect of predictor variables. Empirically, they used the PRM to predict repair rates of oil and gas production facilities in Arctic conditions. They considered three predictor variables: maintenance design, maintenance crew skill, and climatic conditions. Based on 15 records of TTR data of turbo-compressors, their PRM showed two variables, maintenance design and climatic condition, significantly (at 10%) affecting the repair rate. However, their model is based on the proportionality assumption, which needs to be carefully checked in order to ensure the applicability of such model. A limitation of their study is that the proportionality assumption was not checked. Additionally, the effect of predictor variables must be time independent. To enhance the applicability of PRM, Barabadi, Barabady and Markeset (2011) further extended it by incorporating time-independent and time-dependent variables with an application in predicting the repair rate of screen mesh (equipment used by mineral crushing plants). They considered six operational variables, namely temperature, shift (day or night), location, wind, icing, and rain. Based on 16 records of TTR data of screen mesh, their baseline model, with the proportionality assumption, showed that three variables (temperature, shifts, and rain) have a significant effect (at the 5% level). To check the proportionality assumption, they used a graphical method of plotting repair function versus time for the different strata. The proportionality assumption was violated for temperature, and they proceeded to use a stratification approach to resolve the problem. The limitation of the stratification approach is that the break point to divide the strata is arbitrary and typically based on experience. Ozturk and Fthenakis (2020) employed a Bayesian updating technique to predict TTR of wind turbines. They selected six predictor variables related to operational and environmental conditions. Based on 753 records, they found that TTR is higher in high inland locations, higher number of previous failures, and geared-drive turbine types with medium capacity. Overall, the existing literature on repair times typically uses a small number of predictor variables, while our study utilizes a much larger number of (potential) predictors.

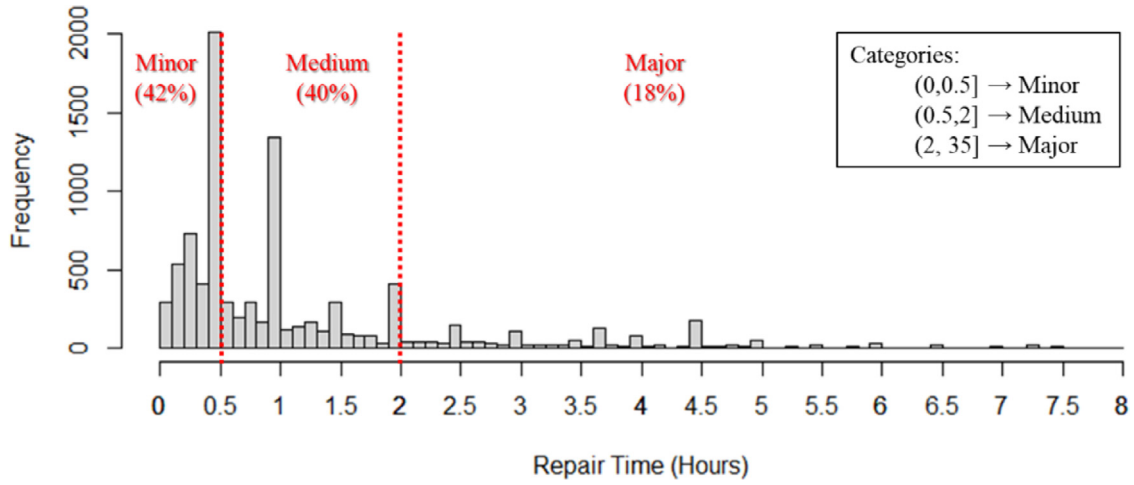

Fig. 1. Histogram of Repair Times and their Categories.

For problems involving a large number of variables, it is necessary to use a variable selection technique to reduce dimensionality. Treating all variables as potential predictors is inefficient and leads to the well-known 'curse of dimensionality' and poor prediction performance by using irrelevant variables (Jursa & Rohrig, 2008). LASSO was firstly proposed by Tibshirani (1996) as the ℓ_1-penalized method and has been successfully developed over the last decades to deal with high dimensionality. Due to the ℓ_1-penalty, LASSO shrinks the coefficients of irrelevant variables to zero and, thus, automatically serves the purpose of variable selection, where only informative variables with nonzero coefficients are selected. Applications of LASSO in forecasting shows its robust predictive power and it has been a prevailing tool in variable selection for forecasting in many areas, such as retail demand forecasting (Ma, Fildes & Huang, 2016), macroeconomics (Smeekes & Wijler, 2018), credit rating (Sermpinis, Tsoukas & Zhang, 2018), marketing analytics (Sun, Zheng, Jin, Jiang & Wang, 2019), and electric load (Ziel & Liu, 2016). One drawback of LASSO relates to variable selection instability, because of the parameter uncertainty in estimating a large covariance matrix. As an extension of LASSO, Zou and Hastie (2005) proposed a revision of the penalization term as the combination of the ℓ_1- penalization and ℓ_2- penalization, which is referred to as the elastic net in the literature. The elastic net has the advantage that the estimated coefficients and variable selection are more stable, and it can further improve the forecasting performance in the presence of highly correlated variables. There are also a number of applications in forecasting by using the elastic net (e.g., Eickmeier & Ng, 2011; Huck, 2019; Ozturk & Fthenakis, 2020). Despite the popularity of LASSO and the elastic net in various fields, they are rarely used in the field of maintainability analysis. This is mainly because the number of available predictor variables in the previous studies is limited, and thus variable selection techniques are not needed. Our study is based on a sizeable dataset with a large number of predictors, and we show that LASSO is associated with a superior predictive power in forecasting repair times of automotive engines.

In the third strand of related literature, probabilistic forecasting is a well-developed field and has many applications in OR. Unlike point forecasting, probabilistic forecasting provides more information as it takes "the form of a predictive probability distribution over future quantities or events" (Gneiting & Katzfuss, 2014, p. 126). Using probabilistic forecasting may bring additional benefits (over and above point forecasts) in decision making in operations management. In the context of winter road maintenance, it is important to make a decision on whether to use the pre-emptive application of chemicals for anti-icing, and such decision is usually taken on the basis of point forecasts. Berrocal, Raftery, Gneiting and Steed (2010) developed two methods for forecasting the probability of ice formation. One important issue highlighted by the authors is that the costs of two types of errors are highly asymmetric: the cost of a road closure is much higher than taking anti-icing measures. Based on the data of Interstate Highway in Washington State, they found that the use of probabilistic forecasts can help reduce 50% of the total cost, compared to point forecasts. In the context of the offshore wind turbine maintenance, it is complicated to make decisions on whether to send over a service vessel because access to the equipment is restricted by the wave height, which depends on weather. Taylor and Jeon (2018) used time series methods to produce forecasts of probabilities of wave height density, and then they incorporated probabilistic forecasting to enable rational decision making on the part of the maintenance engineers. Their recommendation on whether to send out the service vessel relies on the probability of wave height falling below the safe limit. In a follow-up study, Gilbert, Browell and McMillan (2020) employed a boosted semi-parametric model to generate probabilistic forecasts of significant wave height and peak wave period, which further predicted the vessel motion during crew transfer up to 5-days ahead in order to support operations of wind offshore farms. Our study expands the application of probabilistic forecasting to the automotive industry. We follow the work of Berrocal et al. (2010)), to consider asymmetric loss functions, and Taylor and Jeon (2018), to minimize the expected loss related to decision making.

3. Data

The research described in this paper has been motivated by, and successfully applied to, our industrial collaborator, *BT Fleet Solutions* (now restructured and belonging to a different organization[3]), which is a UK-based fleet company providing service, maintenance, and repair for the fleet of *BT Group plc.*, and commercial vehicles of external companies in various industries. The maintenance records studied in this article were collected in their network of 64 garages across the UK. Although each garage and workshop were managed independently by their managers, they all shared the same web-based information system to collect and consolidate the maintenance data, ensuring consistency in data collection across all of

[3] The data from the period we analyzed belongs to *BT Fleet Solutions.* In September 2019, the fleet company was sold to, and is currently owned by a different business entity.

them. This study considers repair records for four fleets of vehicles, including light commercial vehicles, heavy goods vehicles, four-wheel drive vehicles, and personal cars. Following discussions with the company, we focus on the repair time of automotive engines, because engines are deemed to be the most critical vehicle component. The repair times of other vehicle components can be forecasted in the same manner.

3.1. Categories of repair times

There are two strong motivations for us to forecast the categories of repair times, rather than the exact length. First, the purpose of forecasting repair times of automotive parts is to support job scheduling in the garage. Because the repair jobs are scheduled based on the categories of repair time, there is no fundamental difference in terms of job scheduling if the predicted repair time falls anywhere within a category. For example, the scheduling is exactly the same for a predicted repair time of 0.35 hour and 0.45 hour, if they both fall within the same category. Therefore, pursuing a (precise) point forecast is neither the ultimate goal of the company nor, in turn, of this study. The second motivation is due to measurement error. As a demonstration, Fig. 1 presents the histogram of repair times for automotive engines from the case company. It can be clearly observed that spikes occur at the multiples of 0.5 hour. This is because technicians in garages sometimes input the records of repair times rounding up to the nearest 0.5 hour. Thus, the precise length of repair times is not even available, and discussions with our industrial collaborator reveal that this was generally the case. On the basis of the two aforementioned issues, we convert the repair hours into ordinal categories (shown in Fig. 1) and further focus on forecasting such categories (minor: less than 0.5 hour; medium: between 0.5 hour and 2 hours; major: more than 2 hours). Such classification has been undertaken in consultation with experts in the fleet company. The empirical results presented in this study utilize categories of repair times decided upon consultation with experts, which is exogenously determined. But it is worth noting that our work is applicable to any other arbitrary classification of categories, based on the context of other industries.

3.2. Life cycle of automotive engines

Fig. 2 schematically presents the life cycle of an automotive engine, along with the notation of the variables which will be discussed in detail in Section 3.3. A specific engine (characterized by manufacturing information), in brand new condition, starts to be used in a vehicle at time t_0, with mileage reading $M_0 = 0$. During its lifetime, an engine may encounter multiple failures and needs to undergo maintenance in a garage.[4] At the beginning of the i^{th} maintenance, the garage records the time (as t_i), mileage (as M_i), and condition-related information, and the same procedure is conducted for the $(i+1)^{th}$ maintenance. Between t_i and t_{i+1}, the engine is used in a vehicle that operates within a specific region, which provides the geographical information. We can also compute the time between failure (TBF) between the i^{th} and the $(i+1)^{th}$ maintenance as $TBF_{i+1} = t_{i+1} - t_i$, and the mileage between failure (MBF) as $MBF_{i+1} = M_{i+1} - M_i$. The calendar information at $(i+1)^{th}$ maintenance can be directly retrieved from the date of t_{i+1}. At the end of life cycle, the company will sell or scrap an engine.

[4] Generally, the engine needs a maintenance either when it is not functioning (corrective maintenance) or when the technician believes that there is a substantial failure risk associated with the engine (preventive maintenance).

3.3. Choice of predictor variables

We are interested in forecasting the repair time RT_{i+1} of an automotive engine in the $(i+1)^{th}$ maintenance. The ultimate purpose of our forecasting method is to support decision making of the garage manager when the vehicle arrives in the garage for the maintenance of its engine. Thus, the information set $\mathbf{\Phi}_{i+1}$ contains all the information when the vehicle arrives in the garage for its $(i+1)^{th}$ engine maintenance, which is available for us to use in order to forecast RT_{i+1}. We systematically select and categorize the predictor variables into four sets: vehicle condition variables, manufacturing variables, geographical variables, and calendar variables, which are all explained in detail below.

3.3.1. Condition variables

The first set of predictor variables measures the engine condition and are collected at both time t_{i+1} and t_i. It is standard in the automotive industry to use a two-dimensional warranty which depends on the time and the mileage (e.g. Huang, Gau & Ho, 2015; Majeske, 2007; Murthy & Blischke, 2006). Thus at time t_{i+1}, it is undoubtable that the most important condition variables are the age (Age_{i+1}) and the mileage (M_{i+1}). The average mileage ($AvgM_{i+1}$), defined as M_{i+1}/Age_{i+1}, is also useful to measure the engine condition because it gives an indication of whether the engine is over or under-used. Next, it is crucial to distinguish between preventive maintenance and corrective maintenance ($PorC_{i+1}$). In the UK, the routine service on vehicles is closely related to safety inspections regulated by the Ministry of Transport (MOT). This is typically done on an annual basis.[5] As a result of MOT, preventive maintenance is performed to satisfy the safety requirements and to avoid any potential future malfunctions, while no action is needed if there is no problem found. As for corrective maintenance, it is performed in response to an actual malfunction occurred, rather than a potential one. Further, we also consider whether the $(i+1)^{th}$ maintenance is due to an accident ($isAcc_{i+1}$) or not. At time t_i, we collect a range of variables to measure the historical usage of the engine, including the total number of preventive maintenances up to T_i (N_i^P), the total number of corrective maintenances up to T_i (N_i^C), the cumulative repair time due to preventive maintenance up to T_i (CRT_i^P), the cumulative repair time due to corrective maintenance up to T_i (CRT_i^C), whether the i^{th} maintenance is preventive or corrective ($PorC_i$), the repair time at the i^{th} maintenance (RT_i), and whether any parts within the engine are replaced at the i^{th} maintenance ($isPart_i$). Finally, we also include information in relation to the time between t_i and t_{i+1}, which is TBF_{i+1} and MBF_{i+1}, defined in Section 3.2.

3.3.2. Manufacturing variables

The literature found quality difference in automotive parts by different vehicle makers (see, e.g. Sawyer-Beaulieu & Tam, 2015). Thus, the repair time is likely to be determined by the manufacturing information of the engine. Thus, the second set of predictor variables is related to the manufacturing information, which is fixed and not time varying. Specifically, we employ four manufacturing variables: vehicle make (*Make*), vehicle model (*Model*), model year (*MY*), and vehicle type (*Type*). The vehicle *make* refers to the company that manufactured the vehicle, while the vehicle *model* is the name of the range of cars. For example, in the case of a Honda Civic, the make is Honda, and the model is Civic. The additional attribute to further differentiate the same make and model is the model year, which describes the year when a specific version of a model was launched in the market for the first

[5] In addition to MOT, more frequent (optional) service may be done for prudential reasons.

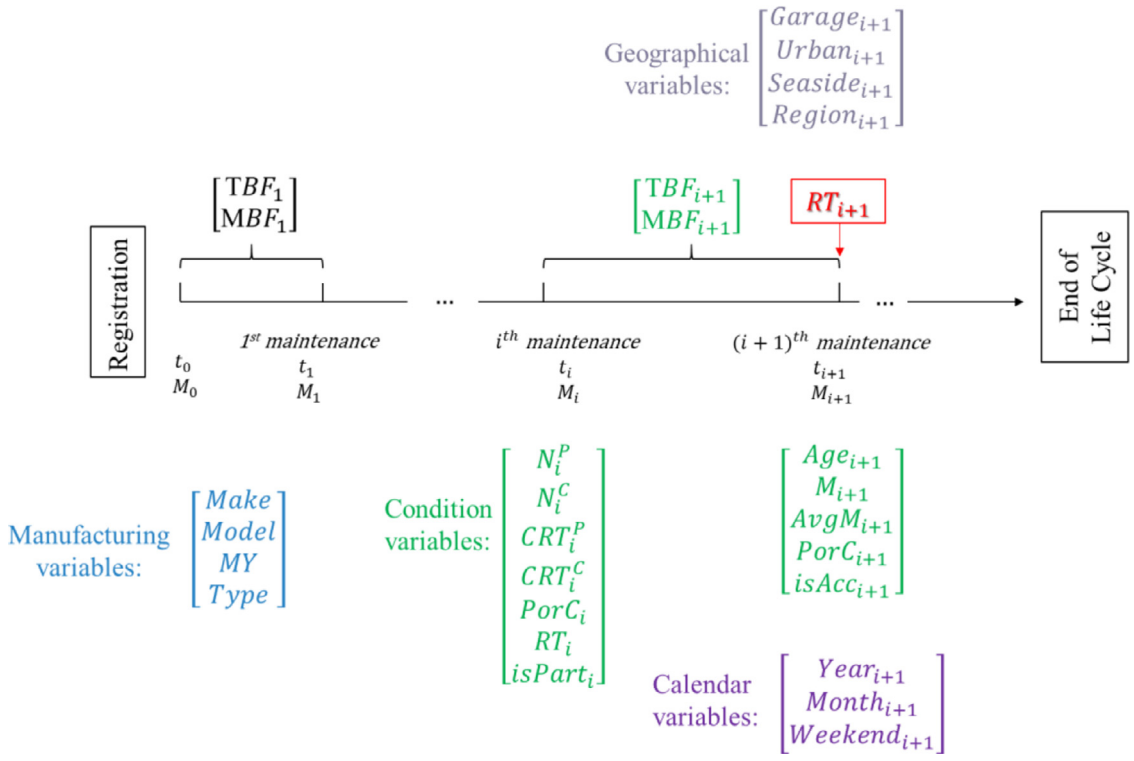

Fig. 2. Life Cycle of an Automotive Engine and Notation of Variables.

time. Lastly, the four fleets of the vehicles are further classified by our collaborator in 10 types, encoded as *A* (small vans), *B* (light vans), *C* (medium vans), *D* (large vans), *E* (medium goods vehicles), *F* (large goods vehicles), *G* (four wheel drive vehicles), *H* (vehicular mechanical aids), *J* (personal cars), and *N* (special medium vans).

3.3.3. Geographical variables

Each vehicle in the dataset is affiliated with one of the garages during a certain period[6] and operates within a certain mileage of its affiliated garage. Thus, the specific garage the vehicle is assigned to between t_i and t_{i+1} ($Garage_{i+1}$) is potentially informative in forecasting the repair time due to the characteristics associated with the garage. After consulting with the garage technicians, the nature of the road around the garage also plays an important role. Thus, we classify the 64 garages according to whether they are situated in an urban area ($Urban_{i+1}$) and/or by the seaside ($Seaside_{i+1}$). Roads in urban areas tend to be flat with higher capacity and good condition, whereas rural roads are typically in a worse condition and located in craggier terrains. Comparing to inland, the air in coastal areas has a higher level of saltness humidity, which could cause corrosion of some certain engine components. Additionally, we use the regional variable ($Region_{i+1}$) to indicate the region the affiliated garage is in. This is because different regions can have different traffic rules on transportation.

3.3.4. Calendar variables

The fourth set of variables considers trend, seasonality, and weekend effects. In more detail, we employ the variable ($Year_{i+1}$), defined as the number of years after 2010 (which is the start of our dataset – please refer to Section 3.4), to capture any potential (linear) trends in the dataset. Additionally, previous studies (e.g., Rai, 2009) have documented the monthly seasonality in repair times as a possible predictor variable, which motivates us to include a relevant variable in our study ($Month_{i+1}$). Moreover, there could be a weekend effect due to differences in traffic patterns on weekdays and weekends (Gao & Niemeier, 2007; Liu, Ge & Gao, 2014). As such, it is also worthwhile including the potentially relevant variable ($Weekend_{i+1}$).

Overall, we consider four sets of predictor variables, which could potentially contribute to forecasting the ordinal category of repair times. From an implementation perspective, the predictor variables may also be further classified as: i) numerical variables, and ii) nominal variables. While it is straightforward to use the numerical variables, the nominal ones need to be encoded as a set of dummy variables before inputting them into the models described in Section 4. For example, $Garage_{t+1}$ is used to indicate which specific garage, out of the 64, is used (at time $t+1$), and it is encoded as 63 dummy variables. After encoding the nominal variables, there are 439 variables in total. This motivates the use of LASSO and the elastic net to conduct a variable selection.

3.4. Summary statistics

Our dataset contains 9511 maintenance records of automotive engines in new vehicles registered with the fleet company after November 2009. The maintenance records were collected from the company's network of 64 garages between February 2010 and September 2017. There are 5291 unique engines in our dataset, and it should be noted that an engine can have multiple maintenances in its lifetime. A histogram of the number of maintenances is presented in Fig. 3. As can be seen, most vehicles (97.8%) have no more than six records of engine maintenances.

Table 1 reports the categories of repair time (RT_{i+1}) in terms of calendar year, month, and weekdays. There are only 6 records in 2010, and most cases of engine repairs are reported after that. This is because automotive engines are reliable in the first few years of their lifetime and rarely have failures. We also observe a higher percentage of major cases in later years. This is also related

[6] It is possible that a vehicle may, occasionally, be assigned to a different garage due to changes in demand. However, such changes are rare.

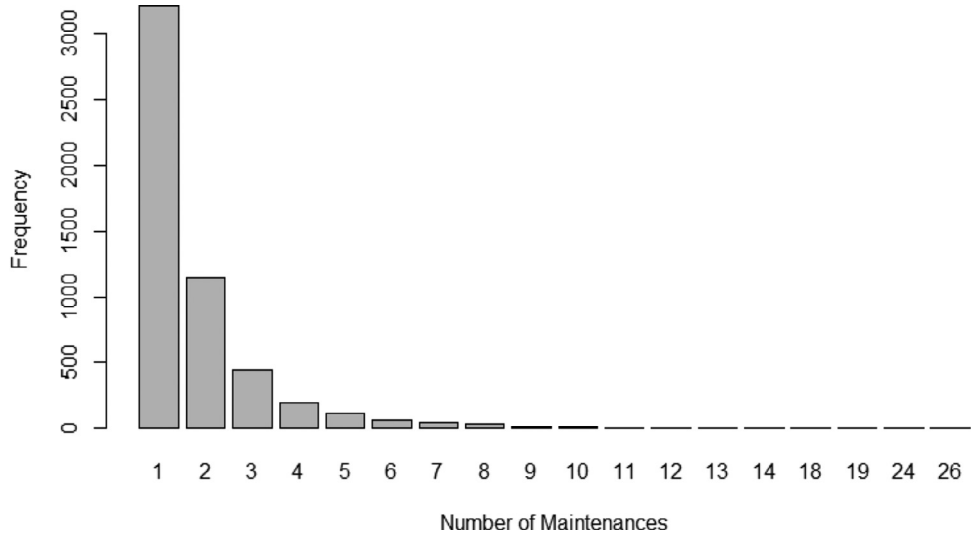

Fig. 3. Histogram of Number of Maintenances (Feb. 2010 – Sep. 2017).

Table 1
Categories of Repair Time by Year, Month, and Weekday.

Year	Minor	Medium	Major	Total
2010	6	0	0	6
2011	42	11	6	59
2012	105	45	12	162
2013	222	187	46	455
2014	434	387	99	920
2015	668	550	172	1390
2016	1147	1335	621	3103
2017	1348	1314	754	3416
Month	**Minor**	**Medium**	**Major**	**Total**
1	353	304	172	829
2	338	332	144	814
3	388	337	146	871
4	289	328	149	766
5	326	312	143	781
6	355	342	142	839
7	357	344	133	834
8	322	377	151	850
9	382	396	156	934
10	282	257	123	662
11	311	260	139	710
12	269	240	112	621
Weekday	**Minor**	**Medium**	**Major**	**Total**
Monday	797	754	368	1919
Tuesday	833	828	375	2036
Wednesday	809	822	357	1988
Thursday	789	750	347	1886
Friday	727	652	257	1636
Saturday	17	22	6	45
Sunday	0	1	0	1

Note: there are 9511 maintenance records for the automotive engines in the dataset. We report the categories of repair times in terms of calendar year, month, and weekdays. There is one special case reported on a Sunday, when the garages are typically closed.

to the reliability of the engines which deteriorates with age and mileage. Regarding the total number of cases in different months, there are fewer cases in April, November, and December. This can be (partly) explained by the holiday effect of Easter and Christmas breaks when the commercial vehicles are less frequently used. We observe substantially fewer cases in weekends as the garages are running in low capacity on Saturday and are closed on Sunday. There is one special case reported on a Sunday. We can observe a relatively even distribution of minor, medium, and major cases across different months and different workdays.

Table 2 shows the summary statistics of the conditional variables. At the time t_{i+1} (in Fig. 2) when the engine has its $(i+1)^{th}$ maintenance, the average age is 4.26 years and the average mileage is 100.72 thousand miles; maintenances due to accidents are rare; and 93% of maintenances are corrective with 7% being preventive. Focusing on the time t_i (in Fig. 2) when the engine has the i^{th} maintenance, the number and cumulative time of corrective maintenances is higher than that of preventive maintenances; the average repair time at the i^{th} maintenance is 1.26 hours; there are 92% corrective maintenances; and the variable $isPart_i$ shows that about half of the cases are associated with engine parts being replaced. Between t_i and t_{i+1}, the mean TBF is 316.88 days and the mean MBF is 17.59 thousand miles. As for the rest of the predictor variables, there are 26 different vehicle makes and 287 vehicle models. Among 64 garages, there are 23 in the urban areas, 16 by the seaside, and the rest are in the inland and rural areas. Finally, the company employs its own classification of four UK regions: Scotland, Northern Ireland, Northern England, the rest of the UK.

4. Method

In this section, we explain the method employed in our study. Our purpose is to forecast the ordinal categories of repair times by a large number of predictor variables. To this end, we employ the ordinal logit model with LASSO, which aims at selecting the most important predictors for forecasting such categories. We also consider the elastic net as another way of regularization. As benchmarks, we choose two machine learning approaches and three naïve methods (including the method currently used by the company). To evaluate the forecasting performance of the considered approaches, we employ four metrics: the Brier Score (BS), the ranked probability score (RPS), and their skill scores, i.e. the Brier skill score (BSS) and the ranked probability skill score (RPSS).

4.1. Ordinal logit model with LASSO

We organize the presentation of the method into two parts. The first part explains the ordinal logit model, and the second part is dedicated to illustrating the variable selection technique via LASSO.

The ordinal logit model is one of many models under the framework of generalized linear models for ordinal data. The

Table 2
Descriptive Statistics of the Conditional Variables.

	Mean	Std.	Min.	25% Q.	Median	75% Q.	Max.
Age_{i+1}	4.26	1.51	0.00	3.23	4.33	5.48	7.81
M_{i+1}	100.72	110.16	0.00	45.97	73.42	110.34	1106.08
$AvgM_{i+1}$	24.13	24.63	0.00	11.84	16.97	25.92	231.76
N_i^P	0.13	0.35	0.00	0.00	0.00	0.00	2.00
N_i^C	1.96	2.00	0.00	1.00	1.00	2.00	26.00
CRT_i^P	0.23	0.87	0.00	0.00	0.00	0.00	11.90
CRT_i^C	2.23	3.79	0.00	0.50	1.00	2.40	54.60
RT_i	1.26	2.15	0.00	0.40	0.60	1.20	51.50
TBF_{i+1}	316.88	322.13	1.00	83.00	211.00	436.50	2694.00
MBF_{i+1}	17.59	23.08	0.00	3.61	9.87	22.14	187.78
$PorC_{i+1}$	0.07	0.25					
$isAcc_{i+1}$	0.00	0.03					
$PorC_i$	0.08	0.27					
$isPart_i$	0.55	0.50					

Note: 25% Q. and 75% Q. denote the respective quantile values. The unit of Age_{i+1} is years, the unit of TBF_{i+1} is days, the unit of M_{i+1} and MBF_{i+1} is 1000 miles, the unit of $AvgM_{i+1}$ is 1000 miles per year, the unit of CRT_i^P, CRT_i^C, RT_i is hours. The distributional statistics are reported for the numerical variables, but not for the binary variables. The minimum of $AvgM_{i+1}$, TBF_{i+1}, and $isAcc_{i+1}$ is very small and shown as zero due to rounding of two decimal places.

model has the assumption that there is a latent continuous variable, and the observed ordinal category of the outcome is from the discretization of the underlying latent variable. Then a link function is used to convert the latent continuous variable explained by the causal variables (on the right-hand-side of model) into the probability of the ordinal categories (on the left-hand-side of model). We choose the logit link function in our specification. This is because the cumulative probability of ordinal categories of repair times has gradual changes. There are other choices of link function, such as the probit or log-log functions, which are less suitable for our task. For instance, the probit link function is for explicit normally distributed latent variables, and the log-log link function is for unevenly distributed categories, such as the situation where higher categories have higher probabilities (Harrell, 2015, Chapter 4).

In terms of the mathematical notation, suppose that each observation y_i, $i = 1, \ldots, n$, belongs to one of the ordinal categories $k = 1, \ldots, K$, and $\boldsymbol{x_i} = (x_1, \ldots, x_p)^{\mathrm{T}}$ represents a p-dimensional vector containing the predictor variables; we model the logit of the conditional cumulative probability

$$\begin{aligned} logit\left(P(y_i \le k)|\boldsymbol{X_i} = \boldsymbol{x_i}\right) &= \log\left(\frac{P(y_i \le k)}{1 - P(y_i \le k)} | \boldsymbol{X_i} = \boldsymbol{x_i}\right) \\ &= \beta_{0,k} + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon_i, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta} = (\beta_{0,k}, \beta_1, \ldots, \beta_p)$ 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 $\sum_{j=1}^{p} |\beta_j| \le s$, as shown in Eq. (2).

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

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, Tibshirani & Friedman, 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.

4.2. Ordinal logit model with the elastic net

In terms of the ordinal logit model with the elastic net, the difference with what was described above is in the form of the penalization when estimating the parameters. It has been shown by Zou and Hastie (2005) that the elastic net can outperform LASSO when the data is highly correlated. In our dataset, some of the predictor variables are correlated, for instance, Age_{i+1} and M_{i+1}. Thus, it is worthwhile considering the elastic net in addition to LASSO.

Regarding the optimization setting, the elastic net aims to estimate parameters $\boldsymbol{\beta}$ by maximizing the likelihood of the ordinal logit model, $\ell(\boldsymbol{\beta}|y_i, \boldsymbol{x_i})$, subject to the ℓ_1-penalization constraint $\sum_{j=1}^{p} |\beta_j| \le s_1$ and the ℓ_2-penalization constraint $\sum_{j=1}^{p} (\beta_j)^2 \le s_s$, as presented in Eq. (3)

$$\hat{\boldsymbol{\beta}} = argmax_{\boldsymbol{\beta}} \left(\ell\left(\boldsymbol{\beta}|y_i, \boldsymbol{x_i}\right) - \lambda \sum_{j=1}^{p} \left\{ \alpha \left|\beta_j\right| + \frac{1}{2}(1-\alpha)\beta_j^2 \right\} \right), \tag{3}$$

where $0 \leq \alpha \leq 1$ is the second tuning parameter to balance the ℓ_1 and ℓ_2-penalization. Considering the extreme two values of α, the elastic net becomes LASSO if $\alpha = 1$, and reduces to ridge regression if $\alpha = 0$. Ideally, the values of α and λ can be simultaneously chosen by 5-fold cross-validation. However, a two-dimensional cross-validation is computationally expensive. As such, we set $\alpha = 0.5$ in this study.[7] As before, all predictor variables are standardized before applying the elastic net.

4.3. Benchmark models

The benchmarks we consider are discussed below.

- ***K*-Nearest Neighbors (KNN)**: KNN is a popular supervised machine learning approach which is typically used as the baseline against more complex methods (see, e.g. Bertsimas & Kallus, 2020; Bertsimas, Kallus & Hussain, 2016). The basic idea of KNN is to "do as your neighbor does", which is simple and intuitive and can perform well for the classification problem (Barber, 2012). Given the predictor variables of a new observation, the algorithm searches the training sample and finds K number of observations which have similar values of predictor variables. Then the new observation is predicted to have an ordinal category of the most frequent class within the K neighbours. We also use the 5-fold cross validation to choose the value of K.
- **Deep Learning (DL)**: Recent developments in the field of deep learning have shown its great predictive power in operations management and maintenance (Chen et al., 2020; Choi, Wallace & Wang, 2018; Kumar, Mookerjee & Shubham, 2018). As a benchmark for this study, we choose to use an artificial neural network with 9 dense layers of 128 hidden units.[8] Although very deep and wide networks have proven effective in general, they come at a high computation cost and may have many redundant parameters (Alvarez & Salzmann, 2016). It is common to encounter overfitting issues for large models with many parameters, and deep learning is not an exception. Developed by Srivastava, Hinton, Krizhevsky, Sutskever and Salakhutdinov (2014), dropout is one of the most effective and widely used regularization techniques for neural networks. Thus, we also introduce two dropout layers to our deep network in order to mitigate the overfitting issue.
- **Random Guess (RG)**: The first naïve method, RG, predicts the categories by randomly picking one from minor, medium, and major categories. The frequency of guessing is set to be the same as that observed in the training sample, i.e. $P(minor) = 42\%$, $P(medium) = 40\%$, and $P(major) = 18\%$.
- **Sample Majority (SM)**: The second naïve method, SM, always predicts the categories as the one with the highest frequency in the training sample. In our dataset, the minor category is the most frequently encountered one in the training sample. Thus, the frequency of SM prediction for the test sample is $P(minor) = 100\%$, $P(medium) = 0\%$, and $P(major) = 0\%$.
- **Historical Average (HA)**: The third naïve method, HA, computes the historical average of repair times in the training sample, depending only on the vehicle types, and uses that historical average as a prediction. Past studies show that the HA method is an effective prediction tool which has been widely used in operations management practices, such as forecasting manufacturing lead time (Marlin, 1986), assisting staff scheduling (Taylor, 2008), and anchoring project duration (Lorko, Servatka & Zhang, 2019).

For the two machine learning methods, we also standardize all predictor variables before applying them. We note that there are other potential candidate benchmark models in the field of machine learning, such as random forest, boosting, adaptive boosting, support vector machine, and a large variation of deep learning models, which could also possibly provide satisfactory predictive power. The selected two machine learning methods are among the most popular ones and could sufficiently serve the purpose of comparison.

4.4. Performance evaluation

In terms of the probabilistic forecasting, there are many choices of measures to evaluate forecast performance. One of the most straightforward measures to apply is the accuracy ratio, which is the percentage of the correct forecasts over the total number of forecasts. To reveal more information about which category is associated with higher accuracy, it is common to use the confusion matrix,[9] which is a two-way table of forecasted categories over true categories. However, the accuracy ratio has been criticized in the literature as not being a suitable measure for probabilistic forecasting (Harrell, 2015), especially for unbalanced data in which the sample majority method can achieve high accuracy, without being useful.

Instead, scoring rules are more suitable measures to evaluate probabilistic forecasting. For example, the field of weather forecasting has been at the forefront of that, through using (proper) scoring rules to assess forecasting accuracy since 1950s (Brier, 1950). Scoring rules are loss functions that map the predicted probabilities and corresponding observed outcomes to loss values. There are three types: improper, proper, and strictly proper scoring rules. A scoring rule is proper when its expectation is minimized if the predicted density is the true density.[10] It is strictly proper if the minimization is unique.

Among many strictly proper score rules, 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:

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

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 $\overline{BS}$.

Despite the popularity of Brier score, it does not take account the ordering of the categories, which is what we face when dealing with ordinal categories of repair times. Developed by Epstein (1969) originally in the field of meteorology, 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:

$$RPS = \frac{1}{K-1}\sum_{k=1}^{K}\left(\sum_{s=1}^{k}\hat{f}_k - \sum_{s=1}^{k}o_k\right)^2. \tag{5}$$

[7] Other values of α, including 0.1, 0.3, 0.7, 0.9, have also been considered. The forecasting performance is similar to the case of $\alpha = 0.5$ and thus the corresponding results are not presented here.

[8] The number of layers and hidden units in each layer of a deep network are typically manually set.

[9] The confusion matrix is also known as the contingency table.

[10] In some literature, a scoring rule is defined as a rewarding function, which flips the sign of the loss function. In such case, the scoring rule is proper if its expectation is maximized, rather than minimized.

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, Liniger & Appenzeller, 2007). The BSS and the RPSS are defined as:

$$BSS = 1 - \frac{\overline{BS}}{\overline{BS}_{ref}}, \tag{6}$$

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS}_{ref}}, \tag{7}$$

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 method while negative values suggesting a less skilled method.

In our empirical analysis, we will report the average of *BSS* and *RPSS* over the training sample and the test sample (see next subsection), with the reference method being the Historical Average. The choice of the reference method is based on recommendations from the Project Management Institute (2013) and International Project Management Association (2015), which suggest the prediction of project duration by using actual durations of similar projects in the past. The rationale behind such suggestion is that the Historical Average method naturally considers the impact of various relevant issues, such as omissions in the project specification, procurement lead time, and misunderstanding of requirements. Lastly, the Historical Average is commonly used as a reference method in the forecasting literature (see Kahneman & Lovallo, 1993; Lorko et al., 2019).

4.5. Forecasting scheme and practical considerations

We demonstrate our forecasting scheme in Fig. 4. We choose the first θ% of the dataset as the training sample and the rest as the test sample. The training sample is divided into five equal subsamples to perform the 5-fold cross validation in order to choose an optimal value for the tuning parameters, including the λ for LASSO and the elastic net. Then, we further K for KNN. Then we use the training sample to estimate the ordinal logit model with LASSO and the elastic net, and also train the KNN and the deep network. Next, the estimated ordinal logit models and the trained machine learning models are employed to produce forecasts in the test sample. There are a number of practical considerations listed below.

- We choose three values of θ (60%, 70%, and 80%) to organize the training and test samples.
- We decide to use two settings for the ordering of the maintenance records. In our main setting (Section 5.1), the maintenance record is the chronological order of arrival, i.e., t_{i+1} (as defined in Fig. 2). In our first robustness check (Section 5.2.1), the maintenance record is ordered by the vehicle identifier.
- Since there is no probability directly generated from the two machine learning methods and the two naïve methods of SM and HA, we set their predicted category to 100% probability and other categories to 0% probability.
- Our implementation is based on the R package “ordinalNet” (Wurm, Rathouz & Hanlon, 2017) and the application programming interface (API) “Keras”.

5. Results

In this section, we present and discuss the forecast performance of the ordinal logit models with LASSO and the elastic net and compare them to the five benchmarks by using the four previously discussed measures for the training and test sample. The performance in the training sample shows how good a model can fit the given data, and the performance in the test sample indicates the true forecast performance of different methods. We present the main results associated with the forecast performance in Section 5.1, followed by three robustness checks in Section 5.2.

5.1. Forecast performance

Table 3 presents the forecast performance of $\theta = 60\%$, 70%, and 80% for the data based on the chronological order of arrival. In the training sample, DL is associated with the lowest BS and RPS. However, this is not necessarily reflected in the test sample, as discussed later. The LASSO and the elastic net have very similar performance, which is slightly inferior to that of DL but better than the other three methods. Next comes KNN, which is better only than the three naïve methods. SM, RG, and HA all perform poorly in the training sample. Using the BSS and the RPSS, it can be seen that ordinal models and machine learning methods do much better than the benchmark, HA.

Now, we turn to the assessment of forecast performance in the test sample, which is the ultimate evaluation of predictability. The first observation is that DL is no longer best in terms of BS and RPS; although it achieves high accuracy in the training sample it may not extrapolate well in the test sample. This is an indication of overfitting, though we have attempted to mitigate such effect by adding two dropout layers. There are some other techniques to further prevent overfitting for deep networks, such as weight regularization. However, we feel that what we have done is adequate given the consideration of deep learning methods only as a benchmark. The second observation is that KNN has high levels of BS and RPS in the test sample, which indicates that some certain patterns are not captured through fitting the training sample by this method. This observation suggests that KNN is underfitting. The third observation is that the two ordinal logit models outperform the other five benchmark models in terms of much lower BS and RPS. Most importantly, the BS and the RPS of the two ordinal logit models in the test sample are similar to those in the training sample. This means that the techniques of LASSO and the elastic net achieve a high balance between underfitting and overfitting, and they can capture the patterns in the training sample and generalize well to the new observations in the test sample. The fourth observation is that the three naïve methods are associated with high BS and RPS (low predictive power). Additionally, the BSS and the RPSS show the superior predictive skills of the two ordinal logit models over the HA method.

We further examine the performance of the ordinal logit model with LASSO for different groups of data in the test sample. As a demonstration, we calculate the average RPS (when $\theta = 70\%$) over different vehicle types, shown in Fig. 5. The ordinal logit model achieves a similar level of forecast performance across different groups of vehicle types, indicating the lack of any systematic bias when it comes to this method. Additionally, it is worthwhile investigating the forecast performance against reliable and unreliable groups of vehicles. To that end, we use information on when the vehicle was registered, and such information is a good proxy for whether a vehicle is (un)reliable. Fig. 6 plots the average RPS

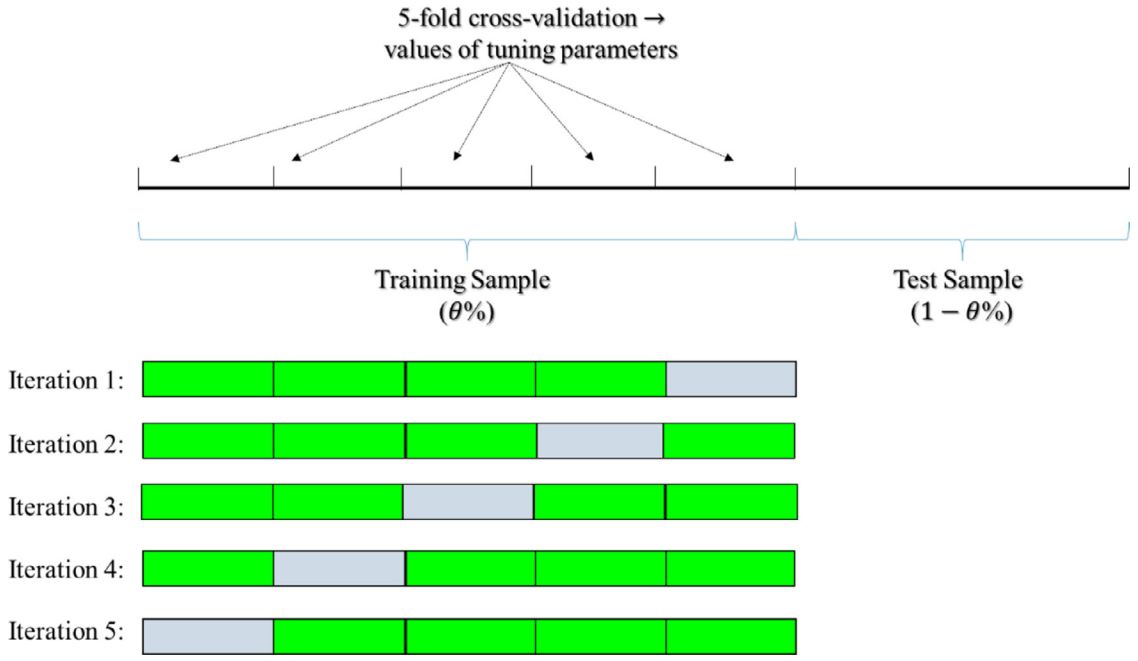

Fig. 4. Forecasting Scheme.
Note: The green part denotes the data used in one of the iterations of 5-fold cross-validation procedure, while the grey part represents the leave-out part in the respective iteration. The cross-validation, estimation, and training of the models are only based on the training sample. The estimated models with the chosen value of tuning parameter are then used to produce forecasts for the test sample.

Table 3
Forecast Performance for chronological order of arrival.

$\theta = 60\%$	Average BS		Average RPS		BSS		RPSS	
	Training Sample	Test Sample	Training Sample	Test Sample	Training Sample	Test Sample	Training Sample	Test Sample
LASSO	0.288	***0.300***	0.170	***0.182***	0.508	***0.514***	0.419	***0.411***
Elastic Net	0.288	0.300	0.170	0.182	0.508	0.513	0.419	0.410
KNN	0.512	0.543	0.295	0.315	0.123	0.120	−0.010	−0.023
DL	***0.265***	0.656	***0.154***	0.399	***0.546***	−0.063	***0.473***	−0.295
RG	0.608	0.637	0.369	0.397	−0.041	−0.032	−0.262	−0.288
SM	0.567	0.605	0.359	0.414	0.029	0.019	−0.230	−0.341
HA	0.584	0.617	0.292	0.308	0.000	0.000	0.000	0.000
$\theta = 70\%$	Average BS		Average RPS		BSS		RPSS	
	Training Sample	Test Sample	Training Sample	Test Sample	Training Sample	Test Sample	Training Sample	Test Sample
LASSO	0.287	***0.295***	0.170	***0.179***	0.515	***0.516***	0.426	***0.414***
Elastic Net	0.287	0.296	0.170	0.179	0.515	0.516	0.426	0.414
KNN	0.512	0.538	0.297	0.315	0.134	0.119	−0.004	−0.032
DL	***0.204***	0.560	***0.120***	0.341	***0.655***	0.082	***0.594***	−0.117
RG	0.617	0.639	0.379	0.395	−0.043	−0.047	−0.282	−0.294
SM	0.574	0.601	0.370	0.406	0.029	0.016	−0.252	−0.330
HA	0.592	0.610	0.296	0.305	0.000	0.000	0.000	0.000
$\theta = 80\%$	Average BS		Average RPS		BSS		RPSS	
	Training Sample	Test Sample	Training Sample	Test Sample	Training Sample	Test Sample	Training Sample	Test Sample
LASSO	0.286	0.296	0.170	0.179	0.520	0.507	0.430	0.404
Elastic Net	0.287	***0.296***	0.171	***0.179***	0.518	***0.508***	0.428	***0.405***
KNN	0.509	0.519	0.293	0.301	0.146	0.136	0.018	0.000
DL	***0.224***	0.593	***0.136***	0.376	***0.625***	0.015	***0.544***	−0.252
RG	0.630	0.635	0.389	0.395	−0.057	−0.055	−0.303	−0.314
SM	0.577	0.604	0.375	0.405	0.033	−0.004	−0.258	−0.346
HA	0.596	0.601	0.298	0.301	0.000	0.000	0.000	0.000

Note: DL denotes deep learning, RG represents random guess, SM means sample majority, and HA indicates historical average. The BSS and RPSS are calculated with respect to HA as the reference method. The **bold** numbers denote the best model. Results are presented to the third decimal place. The data is in chronological order of arrival.

(when $\theta = 70\%$) of vehicles registered[11] in different years. As can be observed, the forecast performance is at a similar level for vehicles registered between 2009 and 2016, while there is a slight improvement (lower RPS) for those registered in 2017. This is mainly because vehicles registered in 2017 are in an almost perfect condition and only need minor maintenance in most cases.

Lastly, it is worthwhile to inspect which variables are selected by LASSO. Under the setting of $\theta = 70\%$, there are a total of 111 variables selected from all 439 variables. As for the breakdown,

[11] The company registers a brand-new vehicle when it starts to serve the company.

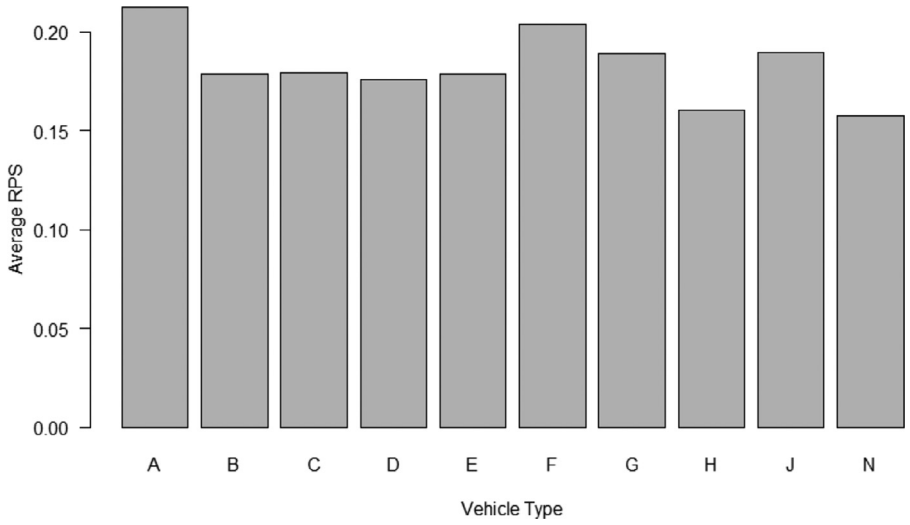

Fig. 5. Average Ranked Probability Score in Different Vehicle Types.

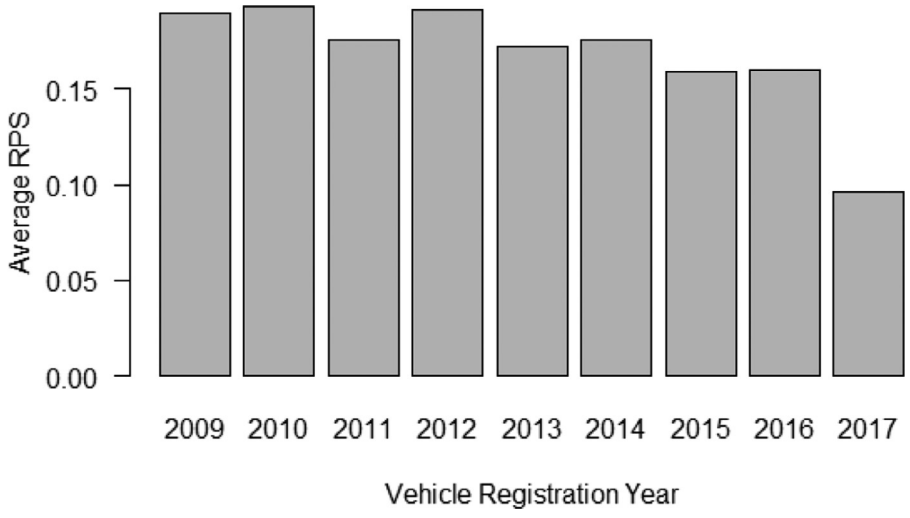

Fig. 6. Average Ranked Probability Score in Vehicles Registered in Different Years.

there are 10 conditional variables, 53 manufacturing variables, 10 calendar variables, and 38 geographical variables.

5.2. Robustness checks

To check whether the forecast performance is robust to different settings, we carried out three robustness checks, including 1) ordering of the data by the vehicle identifier; 2) inclusion of squared and interactive terms as predictors; and 3) inclusion of the vehicle identifier as a predictor.

5.2.1. Data ordered by the vehicle identifier

The main results discussed in Section 5.1 are based on the chronological order of arrival, i.e. t_{i+1} (as defined in Fig. 2). However, there might be a flaw in such a setting due to the unbalanced data in different years. As shown in Table 1, a large proportion (35.91%) of all maintenance records is in 2017, containing 3146 out of a total of 9511 cases. This may lead to a situation where the testing sample contains data only in 2017. Thus, we also order the maintenance records with respect to the vehicle identifier.[12] Table S1 in the Supplement shows the forecast performance for this setting. As can be seen, the methods based on LASSO and the elastic net still have superior forecast performance among all methods. It is worthwhile noting that the setting under concern is considered solely for the purpose of robustness checks, and the chronological order is certainly more realistic because the data is recorded in the right chronological order.

[12] Because the company randomly anonymizes vehicle identifiers, the ordering of data in this setting can be considered as random.

5.2.2. Inclusion of squared and interaction terms

It is possible that some of the predictor variables could have a nonlinear effect on the categories of repair times. In addition, there could also be interaction effects between the predictor variables. To explore these two possibilities, we include the squared terms and two-way interactions for all continuous predictor variables into all forecasting methods and reevaluate their performance (based on the chronological order of arrival). The results related to this setting are presented in Table S2 in the Supplement. First, the BS and the RPS of LASSO are very close to the results obtained without squared and interaction terms in Table 3. This is mainly because the squared and interaction terms are rarely chosen by LASSO. Thus, there is no obvious improvement in forecast performance by including those terms. Second, there is a marginal improvement for the two machine learning methods in the train-

Table 4
Numerical Examples.

Based on an asymmetric Loss Function				
Decision\Actual	k = Minor	k = Medium	k = Major	Expected Loss $E(\ell_d)$
d = Minor	0	2	5	3
d = Medium	1	0	4	1.7
d = Major	2	1	0	**0.7**
Based on a symmetric Loss Function				
Decision\Actual	k = Minor	k = Medium	k = Major	Expected Loss $E(\ell_d)$
d = Minor	0	1	2	1.3
d = Medium	1	0	1	**0.5**
d = Major	2	1	0	0.7
Predicted Probability $\hat{f}_k$	10%	50%	40%	

Note: The numerical values of losses are used for illustration purposes. Upper panel refers to the asymmetric loss function, while the lower panel to the symmetric one. Taking the same predicted probabilities, the optimal decision based on the asymmetric loss function is $d^{\ell,*}$= Major, while it is $d^{\ell,*}$= Medium based on the symmetric loss function.

ing sample, but this is not reflected in the forecast performance in the test sample. Overall, our main conclusion remains the same. The methods based on LASSO and the elastic net still provide the best forecast performance, compared to the five benchmarks.

5.2.3. Inclusion of the vehicle identifier

One might argue that there is heterogeneity associated with engines in different vehicles, which should be controlled for. To investigate this possibility, we incorporate the vehicle identifier (encoded as dummy variables) into all forecasting methods and reevaluate their performance (based on the chronological order of arrival). The results related to this setting are reported in Table S3 in the Supplement. While there is an improvement for LASSO and the elastic net in the training sample, we get nearly the same results of their forecasting performance in the test sample, with or without the inclusion of the vehicle identifier. But occasionally, it is interesting to observe that the forecast performance of LASSO with the inclusion of the vehicle identifier is slightly poorer than the results obtained without it, e.g. the average BS of LASSO in test sample when $\theta = 70\%$. This could be due to the sensitivity of LASSO to the number of predictor variables, which has been theoretically studied by Flynn, Hurvich and Simonoff (2017). They find that the predictive performance of LASSO could deteriorate with more predictors, given a sufficiently high signal to noise ratio and a sufficiently large number of predictors. Another intuitive explanation is that most of the engines have less than 3 maintenance records in our dataset (see Fig. 3). Thus, even though there are significant heterogeneous effects in some certain engines, it is unlikely that they are being picked up by LASSO, based on small number of occurrences for one specific vehicle. Lastly, compared with the benchmarks, the methods based on LASSO and the elastic net are still associated with the best forecast performance.

6. Utility

In this section, we illustrate how garage managers may utilize probabilistic forecasting information to support decision making towards better scheduling of repair jobs. The garage managers' decisions are rather insufficiently informed by a model that only forecasts categories of repair times without providing associated probabilities, such as KNN and deep learning. We show that such decisions can be improved by using information related to forecasted probabilities, along with the loss functions.

We start by introducing a typical scenario in the garages of the fleet company, which is based on the logs completed by the authors following their visits to them. Each day, first, vehicles arrive in the garage (due to malfunction or preventative caution) and the drivers discuss with the garage manager about the specific repair or maintenance request. Next, the managers decide the category that each repair job falls within the three categories (minor, medium, or major); this is done on the basis of the information provided by one of the naïve methods considered in this study. After that, the manager allocates jobs to different technicians. The key issue is that the actual repair time spent by the technician on the job may be different from the manager's original planning. The actual repair time can be longer or shorter than expected.

Wrong decisions lead to inefficient use of staff time and additional (labor) costs. From a customer perspective, costs may relate to additional driver waiting time and unavailability of the vehicles to serve their tasks. It should be pointed out that costs associated with the two types of errors could be asymmetric (e.g. Berrocal et al., 2010). That is, the implications of an actual major category misspecified as minor, are much more severe than the opposite. This is because under-forecasting leads to a shortfall of technicians and the vehicle which is supposed to serve its own company being unavailable for longer time than planned. Over-forecasting implies that the job is finished earlier, in which the manager may inform the driver to collect the vehicle and allocate (if possible) other jobs to the available technician.

Next, we set-up the framework for the analysis. 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 $\mathcal{L}(d, k)$ is introduced to quantify the cost of incorrect classification, which is defined as follows:

$$\mathcal{L}(d,k) = \begin{cases} 0, & if\ d = k, \\ \ell_{dk}, & if\ d \neq k, \end{cases} \tag{8}$$

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

$$E(\ell_d) = \sum_{k=1}^{K} \hat{f}_k \ell_{dk}. \tag{9}$$

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:

$$d^{\ell,*} = \arg\min_d \sum_{k=1}^{K} \hat{f}_k \ell_{dk}. \tag{10}$$

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

$$d^{f,*} = \arg\max_{d} \hat{f}_d. \tag{11}$$

By combining the loss function and the predicted probabilities, the optimal decision $d^{\ell,*}$ is not always the same as the decision $d^{f,*}$. We demonstrate this by two numerical examples presented in Table 4. The upper panel refers to the asymmetric loss function, while the lower panel to the symmetric loss function. Taking the same predicted probabilities, $\hat{f}_k = \{10\%, 50\%, 40\%\}$, the optimal decision is $d^{\ell,*} =$ Major based on the asymmetric loss function, and $d^{\ell,*} =$ Medium based on the symmetric loss function. If the decision purely depends on the predicated probabilities, then $d^{f,*} = Medium$ because its 50% probability is the highest among three categories, and this decision is regardless of whether the loss function is asymmetric or symmetric.

It can be argued that the example in Table 4 is too restrictive because it is based only on one set of predicted probabilities. To reveal the full picture, we find the optimal decision $d^{\ell,*}$ for all possible probabilities based on the asymmetric loss function (as in the upper panel of Table 4). Since there are only three categories, this can be easily accomplished by exhausting all possible combinations of $\hat{f}(minor)$, $\hat{f}(medium)$, and $\hat{f}(major)$. Given the constraint that $\hat{f}(minor) + \hat{f}(medium) + \hat{f}(major) = 1$, the procedure can be simplified by exhausting the combination of any two of them. Without loss of generality, we choose to exhaust $\hat{f}(medium)$ and $\hat{f}(major)$. For comparison purposes, we repeat the same procedure to obtain the decision $d^{f,*}$ based on the predicted probabilities. Fig. 7 presents the area of $d^{\ell,*}$ (upper panel) and $d^{f,*}$ (lower panel) with respect to $\hat{f}(medium)$ and $\hat{f}(major)$. It can be observed that the area of managerial decision $d^{\ell,*}$on minor and medium categories is suppressed, compared to the area of $d^{f,*}$. This can be explained intuitively as follows. The asymmetric loss function results in higher loss if the actual category turns out to be higher than the predicted category. When $\hat{f}(medium)$ is marginally higher than $\hat{f}(major)$, managerial decisions are shifted to the major category due to the higher loss. The same rationale also applies for the shift from minor to medium.

The next question is how much loss a garage may save, for one scheduled repair job, if the probabilistic forecasting is taken into consideration. To answer this question, Fig. 8 plots the difference between loss based on the optimal decision $d^{\ell,*}$ and decision $d^{f,*}$, i.e. $E(\ell_{d^{\ell,*}}) - E(\ell_{d^{f,*}})$. There are three important observations. First, the decision based on $d^{\ell,*}$ is always associated with less or equal loss compared to the decision resulting from $d^{f,*}$. Second, there is no difference in terms of loss when $d^{\ell,*}$ is the same as $d^{f,*}$. Third, the optimal decision $d^{\ell,*}$ may prevent loss mainly in the areas where the predicted probabilities are marginally close between two categories.

The final question is how much loss the fleet company could have saved in total if they were to use what we propose in this paper. To tackle this question, we couple the two loss functions in Table 4 with the results of probabilistic forecasting in Section 5 to compute the total loss for the seven methods. In accordance with our confidentiality agreement with our industrial contributor *BT Fleet Solutions*, the numerical values of losses in Table 4 are not expressed in monetary units, though they approximately represent actual relative cost differences in the various categories.

Table 5 shows the percentage cost savings with respect to HA, and positive (negative) percentage values denote cost reduction (increase) with respect to the reference method (HA). It is clear that the ordinal logit model with LASSO or the elastic net is associated with a substantial benefit in terms of cost savings (under both asymmetric and symmetric loss functions). Additionally, the benefits of using the ordinal logit model over HA is substantially

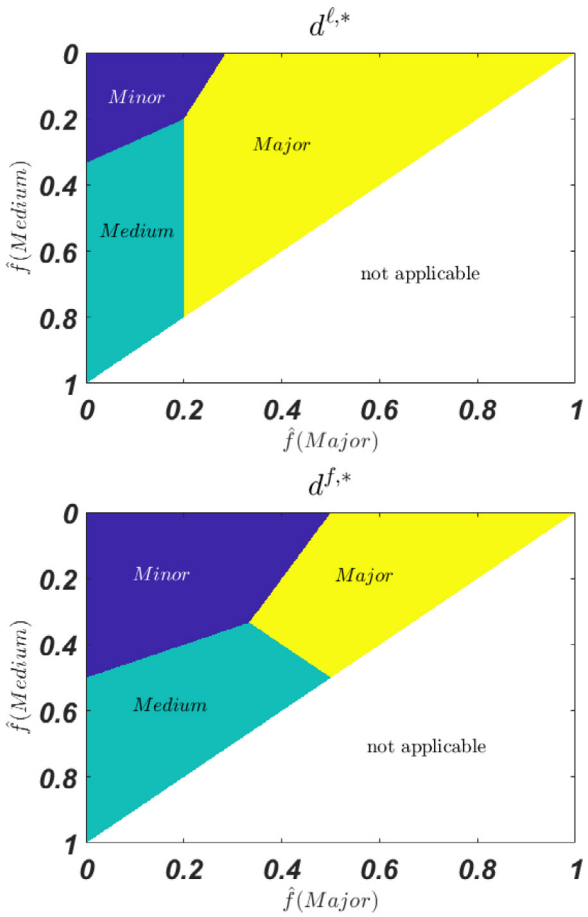

Fig. 7. Comparison of Optimal Decisions.
Note: Upper panel: optimal decision $d^{\ell,*}$ based on the expected loss. Lower panel: decision $d^{f,*}$ based only on the predicted probabilities. Blue: decision of minor category; Green: decision of medium category; Yellow: decision of major category; White: not applicable because $\sum_{k=1}^{3} \hat{f}_k = 1$.

Table 5
Percentage cost savings with respect to Historical Average in the Test Sample.

	Asymmetric Loss Function	Symmetric Loss Function
$\theta = 60\%$		
LASSO	26%	11%
Elastic Net	26%	10%
KNN	−1%	−2%
DL	−4%	−29%
RG	−19%	−29%
SM	−46%	−34%
HA	0%	0%
$\theta = 70\%$		
LASSO	24%	13%
Elastic Net	25%	12%
KNN	−1%	−3%
DL	3%	−12%
RG	−20%	−29%
SM	−48%	−33%
HA	0%	0%
$\theta = 80\%$		
LASSO	21%	11%
Elastic Net	22%	11%
KNN	2%	0%
DL	4%	−25%
RG	−19%	−31%
SM	−50%	−35%
HA	0%	0%

Note: Positive (negative) percentage values denote cost reduction (increase) with respect to the HA method.

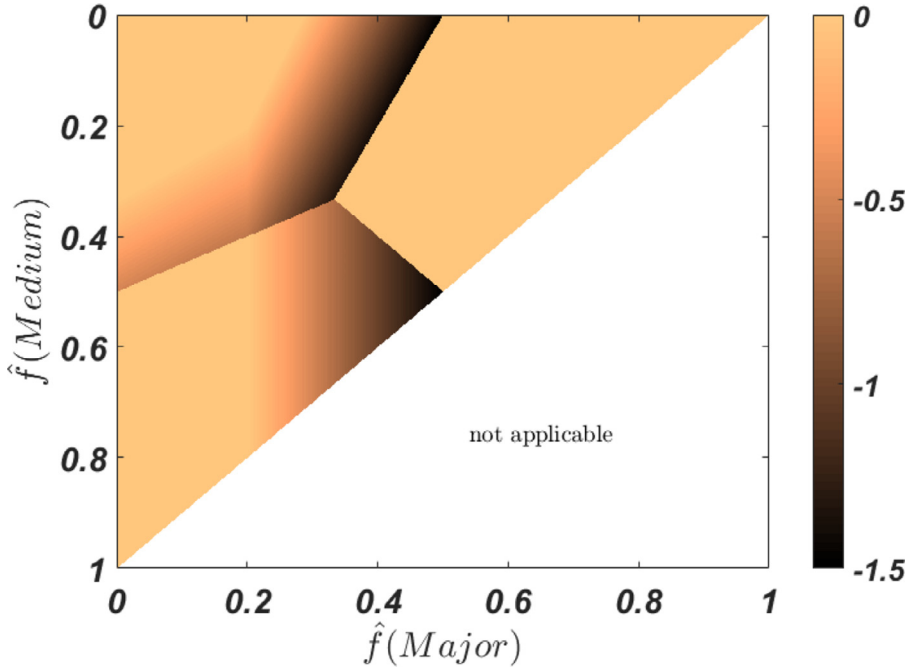

Fig. 8. The difference between the loss from optimal decision $d^{\ell,*}$ and decision $d^{f,*}$.
Note: The white area is not applicable because $\sum_{k=1}^{3} \hat{f}_k = 1$.

larger under the asymmetric loss function. Overall, using $\theta = 60\%$ for demonstration purposes, the cost can be reduced by 26% under the asymmetric loss function if the ordinal logit model with LASSO is used, rather than the HA.

Having observed the potential benefits of the proposed methods, it is worthwhile to briefly discuss implementation in practice. A decision support system (DSS) can be developed to embed the whole framework (including the probabilistic forecast models and the loss function) in order to support garage managers' job allocations. The values of most predictor variables can be automatically retrieved from the database containing the historical maintenance records for registered vehicles. Some information, such as M_{i+1}, $PorC_{i+1}$, and $isAcc_{i+1}$, needs to be collected upon a new maintenance request made by the driver. Then the DSS can provide the predicted probabilities for the three categories (minor, medium, major) and the suggested job allocation with the consideration of the loss functions. Ultimately, the garage manager can decide whether to use the suggestion from the DSS or judgmentally intervene in exceptional circumstances.

7. Conclusion

This study employs LASSO to forecast categories of repair times for automotive engines, based on a large number of predictor variables. The forecast performance is examined on a sizable dataset provided by our industrial partner, *BT Fleet Solutions*. Our method shows superior predictive power and outperforms five benchmarks, including KNN deep learning, and three naïve methods. We further explicitly demonstrate how to use the predicted probabilities in operational decision making in garages. We find that the best decision is not always the same as the point forecast, if the loss function is asymmetric. Our results demonstrate that our method outperforms the company's existing practice and may help them achieve substantial cost savings, especially under asymmetric loss functions.

Our framework, demonstrated by using automotive engines, can be naturally generalized to any repairable automotive parts or machinery parts in other industries, such as vessels and aircrafts. Subject to the data availability, operations and maintenance managers may choose several predictor variables based on their experience and in accordance with the industrial context under concern, and LASSO can help select the most relevant predictors in order to produce probabilistic forecasts. Additionally, practitioners may customize their own loss functions, either symmetric or asymmetric, and link them with the predicted probabilities to support decision making for their own operations.

One limitation of our study is the choice of predictor variables, which is mainly subject to the data availability. Further work can consider more variables collected from other sources. For example, an increasing number of vehicles are equipped with Internet of Things (IoT) sensors nowadays, and such modern technologies can send real-time technical information via the mobile network. The information collected from IoT can generate a vast number of predictor variables, which can be potentially selected by LASSO and further enhance forecasting accuracy performance. Another limitation is that we mainly focused on the ordinal logit regression with standard LASSO and elastic net, which already serves well for the purpose of categorical forecasting for repair time. Other generalizations of LASSO might also be applied, such as fussed LASSO (Tibshirani, Saunders, Rosset, Zhu & Knight, 2005), group LASSO (Yuan & Lin, 2006), and adaptive LASSO (Zhang & Lu, 2007) and further studies could consider such applications. Additionally, some recently developed techniques in machine learning for ordinal data may also be considered, such as random forest for the ordered choice model (Lechner & Okasa, 2019). Lastly, the empirical results presented in this study utilize categories of repair times decided upon consultation with experts, which is exogenously determined.

It should be noted that our framework is general and can accommodate any arbitrary classification, including different numbers of categories and their associated thresholds. This offers an opportunity to find the optimal classifications (number of categories and their thresholds), with the aim to: 1) minimize the scoring rule; and/or 2) minimize the final loss value. Thus, future work may consider the classification as endogenous variable and employ operational research techniques to find an optimal such classification. Another promising research line is to incorporate the loss

function of misclassification into the estimation of forecast models, which allows the models to be directly optimized with respect to the loss function. A third future line of enquiry is to use textual analysis and topic analysis to gather information from the qualitative comments from drivers and turn them (quantify them) into useable predicator variables in the forecast models.

Declaration of competing interest

None.

Acknowledgements

We would like to thank *BT Technology Applied Research* for all the information they have shared with us and for allowing us the opportunity to publish this research. This work has been funded by UK's Engineering and Physical Sciences Research Council (EPSRC) (grant no. EP/P008925/1). For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising. We sincerely appreciate the valuable support provided by our partners and advisers. Finally, we are very grateful for the valuable and constructive comments received from the Editorial Team of this journal and three anonymous reviewers.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2022.06.062.

Data access statement

As part of the confidentiality agreement with *BT Fleet Solutions*, data supporting this paper was made available to the research team subject to a non-disclosure agreement.

References

Aitken, J., Childerhouse, P., & Towill, D. (2003). The impact of product life cycle on supply chain strategy. *International Journal of Production Economics, 85*(2), 127–140.
Alvarez, J. M., & Salzmann, M. (2016). Learning the number of neurons in deep networks. *Paper presented at the advances in neural information processing systems*.
Barabadi, A., Barabady, J., & Markeset, T. (2011). Maintainability analysis considering time-dependent and time-independent covariates. *Reliability Engineering & System Safety, 96*(1), 210–217.
Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.
Berrocal, V. J., Raftery, A. E., Gneiting, T., & Steed, R. C. (2010). Probabilistic Weather Forecasting for Winter Road Maintenance. *Journal of the American statistical Association, 105*(490), 522–537.
Bertsimas, D., & Kallus, N. (2020). From Predictive to Prescriptive Analytics. *Management Science, 66*(3), 1025–1044.
Bertsimas, D., Kallus, N., & Hussain, A. (2016), *Inventory management in the era of big data* (25, pp. 2006–2009). Production and Operations Management.
Brabazon, P. G., & MacCarthy, B. L. (2017). The automotive Order-to-Delivery process: How should it be configured for different markets? *European Journal of Operational Research, 263*(1), 142–157.
Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review, 78*(1), 1–3.
Chen, C., Liu, Y., Wang, S., Sun, X., Di Cairano-Gilfedder, C., Titmus, S., et al., (2020). Predictive maintenance using cox proportional hazard deep learning. *Advanced Engineering Informatics, 44*, Article 101054.
Choi, T. M., Wallace, S. W., & Wang, Y. (2018), *Big data analytics in operations management* (27, pp. 1868–1883). Production and Operations Management.
Eickmeier, S., & Ng, T. (2011). Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting, 27*(2), 496–511.
Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology, 8*(6), 985–987.
Flynn, C. J., Hurvich, C. M., & Simonoff, J. S. (2017). On the sensitivity of the lasso to the number of predictor variables. *Statistical Science, 32*(1), 88–105.
Gao, H. O., & Niemeier, D. A. (2007). The impact of rush hour traffic and mix on the ozone weekend effect in southern California. *Transportation Research Part D-Transport and Environment, 12*(2), 83–98.
Gao, X., Barabady, J., & Markeset, T. (2010). An approach for prediction of petroleum production facility performance considering Arctic influence factors. *Reliability Engineering & System Safety, 95*(8), 837–846.
Gilbert, C., Browell, J., & McMillan, D. (2020). Probabilistic access forecasting for improved offshore operations. *International Journal of Forecasting*.
Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application, 1*, 125–151.
Goodwin, P., Gonul, M. S., & Onkal, D. (2019). When providing optimistic and pessimistic scenarios can be detrimental to judgmental demand forecasts and production decisions. *European Journal of Operational Research, 273*(3), 992–1004.
Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
Huang, Y.-. S., Gau, W.-. Y., & Ho, J.-. W. (2015). Cost analysis of two-dimensional warranty for products with periodic preventive maintenance. *Reliability Engineering & System Safety, 134*, 51–58.
Huck, N. (2019). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research, 278*(1), 330–342.
International Project Management Association. (2015). *IPMA competence baseline: ICB; version 4.0*. International Project Management Association.
Jursa, R., & Rohrig, K. (2008). Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. *International Journal of Forecasting, 24*(4), 694–709.
Kahneman, D., & Lovallo, D. (1993). Timid Choices and Bold Forecasts - a Cognitive Perspective on Risk-Taking. *Management Science, 39*(1), 17–31.
Keizers, J. M., Bertrand, J. W. M., & Wessels, J. (2003), *Diagnosing order planning performance at a navy maintenance and repair organization, using logistic regression* (12, pp. 445–463). Production and Operations Management.
Kumar, S., Mookerjee, V., & Shubham, A. (2018), *Research in operations management and information systems interface* (27, pp. 1893–1905). Production and Operations Management.
Lawrence, M., Goodwin, P., O'Connor, M., & Onkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting, 22*(3), 493–518.
Lechner, M., & Okasa, G. (2019). Random forest estimation of the ordered choice model. *arXiv preprint*. ⟨13:italic ⟩arXiv:1907.02436⟨/13:italic⟩.
Lewandowski, R., & Olszewska, J. I. (2020). Automated Task Scheduling for Automotive Industry. In *2020 IEEE 24th International Conference on Intelligent Engineering Systems* (pp. 159–164).
Liu, Y. L., Ge, Y. E., & Gao, H. O. (2014). Improving estimates of transportation emissions: Modeling hourly truck traffic using period-based car volume data. *Transportation Research Part D-Transport and Environment, 26*, 32–41.
Lorko, M., Servatka, M., & Zhang, L. (2019). Anchoring in project duration estimation. *Journal of Economic Behavior & Organization, 162*, 49–65.
Ma, S. H., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research, 249*(1), 245–257.
Majeske, K. D. (2007). A non-homogeneous Poisson process predictive model for automobile warranty claims. *Reliability Engineering & System Safety, 92*(2), 243–251.
Marlin, P. G. (1986). Manufacturing lead time accuracy. *Journal of operations Management, 6*(2), 179–202.
Murry, C., & Zhou, Y. Y. (2020). Consumer Search and Automobile Dealer Colocation. *Management Science, 66*(5), 1909–1934.
Murthy, D. N. P., & Blischke, W. R. (2006). *Warranty management and product manufacture*. London: Springer-Verlag.
Ozturk, S., & Fthenakis, V. (2020). Predicting Frequency, Time-To-Repair and Costs of Wind Turbine Failures. *Energies, 13*(5), 1149.
Pilar, M. F., e Silva, E. C., & Borges, A. (2021). A Case Study on Scheduling of Repairs in an Automobile Shop. In *Paper presented at the International Conference Innovation in Engineering*.
Project Management Institute. (2013). *A guide to the project management body of knowledge*.
Rai, B. K. (2009). Warranty Spend Forecasting for Subsystem Failures Influenced by Calendar Month Seasonality. *Ieee Transactions on Reliability, 58*(4), 649–657.
Ryan, A. (2015). *KPIs are vital to fleet strategy, but be selective* Retrieved from https://www.fleetnews.co.uk/costs/fleet-managers-guides-and-whitepapers/fleet-and-strategy-2015 [Accessed April 25, 2022].
Sawyer-Beaulieu, S., & Tam, E. K. L. (2015). Maximizing Automotive Parts Reuse, Remanufacturing, and Recycling Through Effective End-of-Life Vehicle Management: A Different Perspective on What Needs to be Done. *Sae International Journal of Materials and Manufacturing, 8*(1), 118–127.
Schultmann, F., Zumkeller, M., & Rentz, O. (2006). Modeling reverse logistic tasks within closed-loop supply chains: An example from the automotive industry. *European Journal of Operational Research, 171*(3), 1033–1050.
Sermpinis, G., Tsoukas, S., & Zhang, P. (2018). Modelling market implied ratings using LASSO variable selection techniques. *Journal of Empirical Finance, 48*, 19–35.
Smeekes, S., & Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting, 34*(3), 408–430.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). *Dropout: A simple way to prevent neural networks from overfitting. The journal of machine learning research, 15*(1), 1929–1958.
Sun, L. P., Zheng, X. N., Jin, Y., Jiang, M. H., & Wang, H. S. (2019). Estimating Promotion Effects Using Big Data: A Partially Profiled LASSO Model with Endogeneity Correction*. *Decision Sciences, 50*(4), 816–846.

Taylor, J. W. (2008). A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science, 54*(2), 253–265.
Taylor, J. W., & Jeon, J. (2018). Probabilistic forecasting of wave height for offshore wind turbine maintenance. *European Journal of Operational Research, 267*(3), 877–890.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288.
Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 67*, 91–108.
US General Services Administration. (2015). *Federal fleet report* from https://www.gsa.gov/policy-regulations/policy/vehicle-management-policy/federal-fleet-report .
Volling, T., Matzke, A., Grunewald, M., & Spengler, T. S. (2013). Planning of capacities and orders in build-to-order automobile production: A review. *European Journal of Operational Research, 224*(2), 240–260.
Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2007). The discrete Brier and ranked probability skill scores. *Monthly weather review, 135*(1), 118–124.
Wurm, M. J., Rathouz, P. J., & Hanlon, B. M. (2017). Regularized ordinal regression and the ordinalNet R package. *arXiv preprint*. ⟨13:italic ⟩arXiv:1706.05003⟨/13:italic⟩.
Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 68*, 49–67.
Zhang, G., Shang, J., & Li, W. (2011). Collaborative production planning of supply chain under price and demand uncertainty. *European Journal of Operational Research, 215*(3), 590–603.
Zhang, H. H., & Lu, W. B. (2007). Adaptive lasso for Cox's proportional hazards model. *Biometrika, 94*(3), 691–703.
Ziel, F., & Liu, B. D. (2016). Lasso estimation for GEFCom2014 probabilistic electric load forecasting. *International Journal of Forecasting, 32*(3), 1029–1037.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: Series B (statistical methodology), 67*(2), 301–320.