Judge's Commentary

MathWorks Math Modeling Challenge 2023

Kelly Black, Ph.D.
Department of Mathematics, University of Georgia

Introduction

The sales and use of e-bikes was the focus of this year's M3 Challenge. Teams were asked to address three tasks: the first was to estimate e-bike sales two and five years in the future; the second was to estimate the strength of association between different factors and the adoption of e-bikes; and the last task was to estimate how the use of e-bikes impacts carbon emissions, traffic congestion, and health.

Some observations about the teams' approaches to each task are given in separate sections below. Additionally, some comments about the fundamental nature of mathematical modeling are discussed in a separate section. The topic of this section was prompted by a concern that an increasing number of teams are including sophisticated techniques in their papers to address questions that can be addressed just as well using simpler approaches.

I think it is worth noting that this year's submissions did not address a factor that impacted the data: supply chain issues resulting from COVID-19 lockdowns. The focus of this year's problem was to estimate sales of a product over a time span that includes the COVID-19 lockdown, and yet I do not recall any teams that included or noted limitations associated with the supply chain in their paper. The recurring supply chain issues imply that a mathematical model could be broken into two parts, before and after the lockdowns that occurred in 2020. The idea that a model could be separated into two parts represents a key insight that was not addressed by any of the teams (at least none that I know of).

Finally, I wish to thank all the students who took part in M3 Challenge this year as well as all the people who supported them. M3 Challenge is a rare opportunity for students to explore a set of open-ended questions and create, present, and analyze a response that can illuminate and provide insight. To see so many students willingly engage in a challenging endeavor is inspiring, and we are grateful for your efforts and dedication. All of the people who support the teams also inspire and are greatly appreciated. We recognize that all of you are subject to immense pressures and continue to endure difficult demands. Thank you for everything you are doing.

Task One

The first task of M3 Challenge 2023 required teams to create a model of e-bike sales. The specific task was to estimate total sales two years and five years in the future. The data provided included sales figures for e-bikes in several different countries. In this section, the potential issues with the data are discussed first, followed by some issues associated with regression and extrapolation. And finally, some of the more complicated models developed by teams and the potential issues associated with them are given.

For the first question, teams were asked to predict future sales of e-bikes in either the United States (US), the United Kingdom (UK), or both. The majority of teams that made predictions for US sales only used the data for the US and did not try to incorporate data from other countries. No UK sales figures were included in the dataset, and many teams adapted the data from France to estimate UK sales. Most teams provided good explanations as to why the sales from other countries were not appropriate for estimating sales in the US or the UK, which is a good demonstration that the teams carefully considered the information and made an informed decision. A critical aspect of modeling is knowing how to evaluate information and decide what to use and what to ignore.

One problem with the US data was that it only included sales figures for five years. The sales for one of those years, 2020, does not follow the same pattern as the other years. Teams reacted to this discrepancy in several

different ways. Some teams censured the data by removing the sales for 2020 while other teams kept the sales for 2020 noting the potential issue. Censuring data should only be done if there is a very good reason. In this case 2020 coincided with the beginning of the lockdowns associated with the COVID-19 pandemic, and a case could be made that 2020 was different. However, few teams recognized that this conclusion also implies that the years following 2020 may also follow a different trend. Regardless of the approach, teams that noted the issue, even if they kept the data, demonstrated that they examined the data and looked for trends and possible issues.

The lack of data for the UK and the sparsity of data for the US represented a difficulty in creating a model. In such circumstances it is good to keep things simple and try to focus on how to best answer the immediate question. The question posed required an approximation of future sales over a relatively short time span. Extrapolation is a difficult task and should be done with a multitude of caveats, but it can be done with more confidence over a short time. The issue was whether a team recognized and stated the potential problems, made rational decisions with explicit justifications, accounted for a minimum number of important factors, and then performed a detailed analysis of the model. In this situation it is good to create a simple model and then perform a wide-ranging analysis to give the reader a better understanding of the potential concerns associated with the model.

Most teams assumed a simple model and then used regression to estimate the parameters in their model. This is a good approach as long as a reasonable justification is given. Many teams tried multiple models. For example, many tried a linear model for sales as a function of time, and then explored the use of an exponential model. To decide which one was better, the vast majority of teams simply looked at the coefficient of determination. A few teams provided a more introspective approach and examined the residuals associated with the model and noticed that the pattern present in the residuals indicated that an exponential model better captures the nature of the data as well as better meets the assumptions inherent in the regression.

Many teams tried to make use of more complicated models. For example, some teams used high order polynomials which is problematic. A high order polynomial may provide better interpolation of the data, but for extrapolation it is not reliable. The use of a particular model should be grounded in some basis beyond simply fitting the data. For example, many teams noted that e-bikes are a new technology and should follow rapid growth in the early stages of its adoption. Other teams noted that sales figures should grow quickly initially but eventually plateau as the market for e-bikes becomes saturated. A common model used in response to this observation is a logistic model. Other teams decided that a Bass model is appropriate. Every year a small number of teams makes use of a Bass model, and this is one of the times it is an appropriate model.

All these choices are good if ample justification is given, and it is preferable to make use of as simple a model as possible. Task one provided a good context for using something basic like an exponential model due to the short time for the forecast. What is more important, though, is the analysis associated with the model. I tend to focus more on the resulting analysis of a model, and a team that chooses a simple model, like an exponential model, and then conducts an extensive analysis can make a positive impact.

A thorough examination of the residuals, appropriate plots, and different points of view, as well as exploring different kinds of "what if" scenarios can make a good impression. For example, some teams used a logistic function for their first model. One of the difficulties with a logistic model is approximating the coefficients, and many teams struggled to fully explain how they obtained their results. Some teams made assumptions about the saturation level (the carrying capacity) based on information from other sources or used the initial year for the initial value. They then used the data to approximate the other parameters. Such approaches can be problematic when only a small number of data points are available, especially when the data does not include the inflection point or any points near the saturation level.

A structured exploration of how the final results are impacted for different values of the assumed quantities can help the reader better understand the potential issues and pitfalls associated with the model. For example, a team using a logistic model that did not estimate the carrying capacity using the data may try changing the carrying capacity by a non-trivial amount and then recalculate the other parameters. If there is little change in the final predictions then this may indicate that the phenomenon of interest is still experiencing fast, initial growth, and a simpler exponential model may be adequate.

Another important part of the analysis of the model is to get a broader understanding of the sensitivity of the model. This can be done in a wide variety of ways and, given the short time period associated with the event, teams are not expected to perform an extensive sensitivity analysis for multiple aspects of the model. It is beneficial, though, for a team to demonstrate the general idea. For example, a team could examine the differences in the final estimate that occur when the 2020 data is censured or changed by some small amount. A team could explore each data point to see which one has the greatest influence on the final approximation. Another option is to simply make a small change to the parameters, determine the differences in the final estimates, and then explicitly state which parameter has the biggest impact in the final estimates.

While most teams opted for a simple model for task one, some teams opted to examine more complicated models. One example of a complicated model is an SIR model used to approximate the spread of a disease. A small number of teams divided the population into people who do not have an e-bike, those who adopt an e-bike, and those who stop using their e-bike. The assumption is that people who decide to use an e-bike do so after being influenced by other people. This is a bit problematic in that it creates the answer to the second question by assuming the primary influence associated with adopting an e-bike. Additionally, the SIR model can be quite complicated creating numerous issues in trying to approximate the solutions as well as trying to determine the coefficients to the system of equations.

Another example of a more complicated model is the use of a Markov chain. This is a more reasonable model than an SIR model. One advantage to this model is that it creates a context to handle the phenomena of people who have an e-bike but decide to stop using it as well as being able to accommodate people who purchase a replacement and continue to use an e-bike. The downside to this approach is that it can be difficult to approximate the transition probabilities, which is an especially acute issue given the small number of data points.

Task Two

Addressing the second task turned out to be the most difficult aspect of this year's problem. To address the second task, teams had to determine one or more factors associated with the adoption of e-bikes and then estimate how important the factors are with respect to people using e-bikes. Most teams examined more than one factor and compared the factors. Many teams looked at two or three factors, which is quite good given the enormous time constraints associated with M3 Challenge.

Many teams examined the factors that were associated with the data provided. The data included information on the following topics:

- Cost of e-bikes

- Disposable income

- Environmental perceptions

- Improvement in battery technologies

- Cost of energy

Some teams were able to find other data sources and examined other factors. Examples of other factors include the following:

- Local bicycle infrastructure

- Commute distance

- Socioeconomic status

The judges did not express a preference for which factor a team should examine—as long as they examined at least one factor they met the requirements for the question. Examining more than one is good, though, because it

allows the team to compare how their method works with multiple factors, but anything beyond a few factors is not providing much additional context. The resulting analysis and discussion about the methodology is more important.

With respect to the methodologies, many teams estimated the strength of the associations between the different variables by calculating their correlations. To use correlation to test the association assumes there is a linear relationship between the two factors; however, in some cases that may not be true. For example, it may be that for a person with a low disposable income an increase in income will result in an increased likelihood of purchasing an e-bike because the cost of a bicycle is more affordable, but at some point once the disposable income becomes large enough the likelihood of purchasing an e-bike may decrease since the person can more easily afford the cost of fuel for a car. If this is the case, the relationship between disposable income and adopting an e-bike may be a quadratic function, and it will not necessarily be detected by solely looking at the correlation. Some justification for using correlation for each variable is beneficial in understanding why it is an appropriate way to compare each of the different factors.

Some teams used more sophisticated methods. Several teams made use of a machine learning technique, and a common choice was to use a random forest model. These kinds of approaches can be difficult to describe even though they are easy to implement, and a team should not assume the reader will figure out the details by reading code. The description should be complete enough that the reader should be able to reproduce the team's results. For example, the data that is used should be carefully documented, including citations in the text. Simply putting a link in the reference section is not adequate. Additionally, it should be made clear how the data was subdivided for the training as well as testing. In this situation, the small number of data points made some of the approaches problematic. Performing an analysis of the approach can be difficult and reporting the results can be a non-trivial task.

Another approach used to address this question was to implement a Monte Carlo simulation. A Monte Carlo simulation can provide insights that other techniques cannot, but the approach suffers from some of the same disadvantages as machine learning. The implementation of a Monte Carlo method requires many assumptions and a set of very specific rules must be defined. Every factor and aspect describing an individual agent within the simulation must be drawn from a probability distribution, and each distribution should be explicitly stated and justified. Describing the probability distributions and rules can be tedious to write and difficult to read, but a team should not expect a judge to read their code to get the details. With respect to the analysis of the approach, at the least the team should explore how their conclusions change as different parameters are changed, which can be an immense task considering that every rule and every probability distribution may have multiple parameters.

Regardless of which approach a team used, presenting the results proved to be a difficult task. For example, some of the teams that examined correlations showed plots of the different factors as functions of time but did not plot one factor with respect to another. It was not uncommon to see plots of battery costs versus time next to plots of sales figures versus time, rather than seeing sales plotted versus the battery costs. Viewing both factors as functions of time made it difficult to understand the relationship between the two factors. Additionally, reporting the results of a large number of calculations of correlations is awkward, and organizing a large number of results is difficult, especially when numbers are reported to more than a few digits of precision. Teams that were able to graph their results or present their results in a more easily read format tended to make a more favorable impression.

Reporting the results of a machine learning model or a Monte Carlo simulation is even more difficult. The different machine learning methodologies result in different kinds of weights associated with the results. Teams that carefully described the weights, provided context to their meaning, and discussed the uncertainty associated with the results tended to make a more positive impact, and also demonstrated that they had a better understanding of the methodologies employed. The results of a Monte Carlo simulation represent a sample from a stochastic system, and the results should be reported in a way that gives a sense of the distribution of the output. For example, histograms, box plots, and statistical summaries should be used to demonstrate the distributions associated with the results, and summaries of the results should indicate the central tendency as well as the spread in the data.

Task Three

To address the third task teams had to predict various impacts that result from a change in the usage of e-bikes. The problem statement indicated that the teams should address carbon emissions, traffic congestion, and health impacts, and some teams also examined other factors. There is no expectation to examine other factors and, given the time constraints, it is generally good to try to maintain a focused discussion.

Regardless of the number of impacts explored, the third task represented an opportunity to tie the previous tasks together. To demonstrate how different models can be used together and weave a story about the relationships between different variables creates a compelling case to the reader and directly demonstrates that a team can use, build on, and adapt mathematical models to answer important questions.

In this case, most of the impacts tend to form relatively straight-forward relationships. For example, the difference in carbon production can be approximated using a simple linear function comparing the reduction in the number of cars and the increase in the number of e-bikes. One subtlety, though, is determining the number of miles commuters drive or bike. It is reasonable to assume that people with a shorter commute are more likely to ride their bike. Determining the factors to use and determining which people will commute by bike is a non-trivial task, and teams accomplished this in a wide variety of ways.

Most teams used relatively simple models, and one of the primary differences between submissions is how carefully a team documented the expressions employed in their model as well as discussed and provided motivation for each term within their model. A team that discussed why a linear function was used, stated how they determined the values of constants, included citations and references, and then provided a detailed analysis of the resulting model clearly demonstrated an understanding of the interplay between insight, analysis, and presentation.

While the models employed for the third question are important and tended to be relatively straightforward, an equally important factor was the analysis employed to examine the ramifications of the model. When taken together, the different models included many parameters and many different parts. Discussing the results as well as discussing the robustness of the model is a difficult and time-consuming task. It is also a difficult writing exercise to discuss the analysis when the team is under time constraints. The teams that were able to do so tended to make a more favorable impression.

General Comments

In most of the previous Judge's Commentaries the focus of this section has been on basic practices and presentation of results—I recommend that you look at the general comments from previous years. This year, however, I want to focus on a topic that I heard about in multiple conversations with other judges. It is a topic that gets to the heart of what modeling is and why it is used. Mathematical models are generally used to gain insight and enhance our understanding of a complex topic, and a premium is placed on the development of the simplest model possible that captures the fundamental trends that can be found in a given phenomenon.

Over the past few years there has been a growing trend to include more sophisticated models in the papers submitted to M3 Challenge. This year, I personally read many papers that included models with very complicated methodologies to model a relatively simple phenomenon, and I heard from other judges who shared a similar experience. The use of a complicated model for the first question in this year's problem can cause confusion for the reader and does not result in a better approximation. A cursory examination of the data indicates that sales of e-bikes appear to be in the earliest stage, and there is likely to be similar growth for a short time. Combined with the small number of data points, it is difficult to draw conclusions with much certainty. Spending time on a complicated model for the first question will likely result in only a marginally better approximation, and given the time constraints, it can be more difficult for a team to spend vital time on other questions as well as on the analyses of the models.

What does a judge look for? The development of a model and the analysis of the resulting model are equally important. Furthermore, a simple, more elegant model provides a better context to understand the phenomenon

of interest. Having said that, a simpler model does not imply it is obvious. A team should provide good reasons why a model is appropriate. For example, if regression is used, a reason why a relationship is linear or exponential should be explicitly stated. Also, once the regression is completed, the residuals should be inspected to ensure that there are no obvious patterns, and the residuals should meet the conditions of the underlying assumptions of the regression method. Simply comparing coefficients of determination is not enough to fully evaluate the results of regression.

When a mathematical model is stated, a reason for the choice should be provided. Also, the individual terms within a model have meaning, and a team that demonstrates an understanding of what the terms are and how they interact with each other will more clearly demonstrate an understanding of the modeling process. For example, if a team decides to use a logistic equation then a well-reasoned explanation for the choice should be given. There is no need to go into a discussion about the differential equation which leads to the logistic equation, but some discussion as to how the different parameters relate to the current situation is helpful. For example, a discussion about terms within the equation that relate to the carrying capacity, the initial rate of growth, and the inflection point in the context of the situation is vital to help the reader better understand why certain choices are made and why they are good choices. You should assume that the reader is not familiar with the situation and it is your job to describe the context as well as provide a convincing argument why your choices are good options.

With respect to the development of a model, it can be a fine line between being too verbose and skipping too many steps. It is helpful to the reader to see the development of each part of the model before putting everything together. Explanations should be given for each conclusion and assumption; at the same time, one should be careful about sharing algebra among polite company. Choices and assumptions need to be explicitly stated because the reader does not know what you are thinking. The algebraic steps used to reach a conclusion can only be duplicated if a rough scaffolding of the steps is explicitly provided.

Once a general model has been chosen and the specific model developed, it is vital that a careful analysis of the model be conducted. The short-term trends, long-term trends, and other important features should be explicitly stated. A mathematical model represents a rough approximation, and the limitations and deficiencies of the model should be explored and stated. A careful error analysis that goes beyond just examining the coefficient of determination should be discussed. A full sensitivity analysis is not possible due to the time constraints, but some effort should be made to determine the potential pitfalls. This can be done by examining what happens to the final conclusions for a small change in a set of parameters, "jittering" the data points, determining if there are any influential points in the data set, or comparing the results under different assumptions. The censoring of any data points should be done with extreme caution, and the removal of a data point should be carefully documented and justified.

Finally, the issue of citations versus references comes up every year, and this year is no exception. Teams are quite good at providing references at the end of their report that list the sources used. However, a consistent reference style should be used, and a simple list of URLs to resources is not sufficient to document sources. The inclusion of citations within the text that indicate the context in which a certain reference influenced the work is absolutely vital. When a paper has both citations and references it sends an immediate message to the reader that the authors respect the work of others and understand how the development of a model requires adapting existing models to new contexts.

Conclusions

The growing uncertainty about lifestyle, the cost of energy, and concern for our long-term health motivate introspection into many of our most basic choices. A central aspect of our lives is how we travel, and the use of e-bikes has increased dramatically. The tasks in this year's problem centered on the short-term adoption of e-bikes, what is influencing their use, and the impact of their use.

Some observations of how students responded to each of the three tasks is presented, and some introspection is included on the central idea of what a mathematical model is and why mathematical models are used. The

primary use of mathematical models is not necessarily to provide the most accurate answer, rather it is to provide insight and enhance our understanding of a given phenomenon. The simplest model that captures the most important aspects of a situation can best help develop intuition. An essential part of developing that intuition is an exploration and a stress test of a model to determine if it really does mimic important features.

Finally, I once again wish to thank all the people associated with MathWorks Math Modeling Challenge. The students inspire us. The faculty, parents, and others who serve and provide the context to the students make M3 Challenge a worthwhile and positive experience. The organizers at SIAM ensure that it happens and put everything in place to allow students to excel. The support from MathWorks provides the foundation and resources to make everything possible. Thank you all!

Acknowledgements