# Judge's Commentary

MathWorks Math Modeling Challenge 2022

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

## Introduction

The 2022 M3 Challenge posed three questions centered around the future trends of remote work. The start of the COVID-19 pandemic initiated an abrupt change in how work from home is perceived, and it resulted in a broad evaluation of our shared assumptions about the efficacy of working from home. To answer the three questions, students had to predict how many jobs could be performed at home in the near future, decide how to predict whether or not a person would work from home, and then combine their findings from the first two questions to predict future trends for five different cities.

In the commentary that follows, I will provide some insights into what I observed for each of the three questions as well as provide some general comments that are relevant to any modeling effort. The list of observations given here is not complete, and I recommend reading the commentaries from previous years for a more exhaustive list of things to consider.

Once again, student teams submitted an outstanding collection of entries. Despite the challenges of the last two years, students still amaze us as we read their incredible work. We are also reminded that this event would not be possible without the efforts of the students and the support they receive from their teacher-coaches. We recognize that the last two years have been difficult and the pandemic continues to place considerable stresses on everybody. To see students and advisers continue to push themselves to achieve and take part in these kinds of exercises is inspiring, and we are grateful for the example you set.

## Question One

To answer the first question, teams had to develop a model to approximate the percentage of workers whose jobs can be performed remotely. The teams were required to implement their model using data from five cities: Seattle, Washington (U.S.); Omaha, Nebraska (U.S.); Scranton, Pennsylvania (U.S.); Liverpool, England (U.K.); and Barry, Wales (U.K.). The data that was provided included information about different sectors of the economy as well as details about sectors of the job market, such as manufacturing and retail.

Teams had to decide how to make use of the information and in some cases integrate it with additional information obtained elsewhere. Most teams used all the sectors as defined in the provided data, while some teams used a subset of the sectors they felt were most important. Other teams aggregated the data by combining some sectors, and some teams made use of subtotals from across all sectors. Any of these approaches was good, and regardless of which approach a team decided to use, there was an expectation that a team clearly state which sectors were used and clearly state their methodology for how they used the information.

Once a team decided which parts of the job market to examine, team members had to create a model that would approximate the trends that occur over time. A wide variety of models were employed; some examples of the

models used were linear, exponential, logistic, and compound interest functions. The judges did not place a priority on which type of model was best but focused instead on the teams' justification and analyses of their models. For question one, teams were expected to approximate what would happen in the near future, and because of the limited time span, a simple model was able to yield adequate results. For example, the choice of a linear approximation can be justified by simply noting the local linearity of the function over the relatively short time span examined, which was especially appropriate in light of the small changes seen from year to year.

Some teams noted that the provided data indicated different kinds of trends for the different cities. Ideally, a modeler would be expected to provide some insight into why such differences exist. In this case, however, teams had limited time and many other pressures. Due to the complexity of the differences between cities and the nature of the event, the judges recognized that requiring such details would be a significant burden. We felt that it was more important that teams recognize the differences, react in appropriate ways, and clearly state their observations.

With respect to the model itself, different teams interpreted the notion of a percentage in different ways. Some teams normalized their totals with respect to the overall population or, in some cases, the number of people in a given sector for each of the cities. Others normalized with respect to the number of positions. In either case, the teams had to develop models to project their respective totals in future years. As long as teams clearly stated how and why they needed such totals, the judges focused on whether or not their approach and subsequent calculations were consistent with the quantities being calculated.

Once a team decided on the general form of the overall model as well as the different parts of the model, the team had to use the data to determine approximations for the various coefficients within their model. Most teams made use of all the data and performed a regression technique. For example, teams assuming a linear model tended to make use of a linear regression model for the full data set. Some teams recognized that the start of the pandemic marked a significant departure from the years preceding the pandemic, which is an important insight when examining the data. In response, many teams simply truncated the data and only made use of the data observed before the pandemic.

Some teams tried to make use of a piecewise defined function by splitting the data into two parts, before and after the start of the pandemic. Such an approach revealed that a team had important insights into the problem, but at the same time, the small amount of data available after the start of the pandemic made it difficult to rely on the robustness of the results. From the judges' point of view, it was difficult to balance the two concerns. Recognizing the nature of the discontinuity at the start of the pandemic demonstrates the team has excellent insights into the nature of the problem, but at the same time, trying to construct an approximation based on so few data points introduces a good deal of uncertainty into the final approximation.

The uncertainty associated with the approximations is an important aspect of the problem. The different regression methods used to approximate the parameters in the models include different assumptions about the data. Teams that noted the assumptions and were mindful to provide warnings made it clear they recognized some of the limitations and potential problems with their calculations. Additionally, it is important to note the uncertainty associated with regression methods. Teams that noted confidence intervals or noted the uncertainty in other ways for their final calculations demonstrated that they recognized some of the limitations in the methods used.

Some of the difficulties in working with data manifest themselves in other ways as well. The data that was provided came from multiple sources. It was not immediately clear how the methodologies and assumptions differed for the collection efforts with different sources. Teams that noted the possible problems with the combined data made an impression and demonstrated an understanding of some of the issues when comparing calculations based on data coming from multiple sources.

Another issue associated with the use of this particular data set is the small changes in some of the sectors over the given time span. Many of the calculations for the slopes returned values close to zero for linear regression and growth rates close to zero for exponential regression. Very few teams made note of this, nor did they provide warnings about the difficulty in establishing the relationship between the two variables being compared. If a coefficient is close to zero the resulting confidence interval may include both positive and negative numbers, and the nature of the long-term behavior is difficult to predict.

## Question Two

To answer the second question, teams had to develop a model to decide whether a given person will be allowed to work from home and whether or not they will want to work from home. Most teams approached this by examining the two questions separately, and few teams examined the case where someone had to work from home regardless of their preferences due to a corporate decision. The task of combining the two questions was handled using a wide variety of methodologies.

One common approach was to determine a probability distribution for whether or not a person would want to work from home given information about the person and then repeating the process to separately determine the probability distribution for whether or not an employer would allow the person to work from home. It was not uncommon for teams to multiply the two probabilities to determine the probability that an individual would work from home. This is problematic in that it assumes that the two events, wanting to work from home and being allowed to work from home, are independent of one another. It is likely that some employees and employers arrive at their conclusions using similar processes which would mean the two decisions may not be independent.

Another issue arose when teams divided different considerations into different sets, determined probabilities associated with each set, and then added the probabilities. For example, a team might determine the probabilities associated with people based on aspects such as their number of children, their commute time, and their business sector. The sets associated with each grouping are not necessarily disjoint, and it may not be appropriate to add the probabilities to get a final estimate. Teams that recognized this problem and took steps to refine their groupings made it clear that they understood an important aspect of their approach and tended to make a more positive impression on the judges.

The issues associated with using sets that are <u>not</u> disjoint also arose for those teams that used Bayes theorem to construct their probability distributions given a wide variety of aspects such as number of children, commute time, and age. Bayes theorem assumes that the subsets examined do not overlap and that their union is the whole sample space. Even for those teams that did construct disjoint sets, the methodology to do so could be quite complicated. The explanations of the subsets could be difficult to understand, and teams had to be very careful and provide many details when describing how they partitioned their population of people. Teams that were able to describe their subsets and convey an understanding of the requirements associated with Bayes theorem tended to stand out.

The difficulties in describing the method to determine the probability distributions was compounded when the teams described their results. The number of aspects used to describe a specific person could be extensive, and as more details were examined, the number of results associated with the groupings increased. Between the large number of business sectors and extensive number of characteristics, clearly reporting the resulting probability distributions was a daunting task. Many teams included multiple tables, one for each business sector or other groupings of their choosing. Other teams included graphs or bar charts, and some provided lists of expressions. Simply providing a list of tables without commentary results in a report that is difficult to read for a judge seeing the paper for the first time. Some teams focused on one table and provided instructions on how their results were calculated, including a brief note about the other tables. As a rule of thumb, someone reading a report should have

enough information to reproduce the results. The papers that found a balance between being concise and making it clear how the results were determined tended to make the most positive impressions.

The challenges associated with reporting results was an acute issue for teams that made use of a machine learning algorithm. It was more common for teams to make use of machine learning algorithms this year compared to previous years and presenting the results of the algorithms posed the same challenges as discussed above. Moreover, providing the details for how a machine learning algorithm was implemented is a difficult task. Again, a reader should be given enough information so that the results can be reproduced, but at the same time too many details can overwhelm the reader. It is a difficult balancing act when the details about training sets, verification, and validation are vital. Another critical detail, which was rarely included, is how the cost function is defined and implemented for a machine learning algorithm. Additionally, these kinds of algorithms make use of many assumptions and requirements. The reader should be aware of them, and the team should make the limitations clear.

## Question Three

To answer the third question, teams had to bring together the results from the first and second questions and use them to predict the number of people who will work from home for each of the five cities. Explicitly requiring the integration of the first two questions was a departure from previous years. It is a challenge to bring together two different ideas into a new, coherent response to a different question. It was also a challenge for the teams to schedule and divide their efforts since they could not produce results on question three until after the first two questions were addressed.

When answering the first two questions, teams defined the business sectors as well as the characteristics of individuals to use in constructing the prediction. When bringing them together to address the third question, teams were required to review the previous results and assumptions and ask if they made sense. For example, a large number of teams assumed that having more children made it more likely that someone would want to work from home, but the difficulties for a parent to accomplish other tasks when children are present might imply the opposite conclusion. When applying the previous results to the third question some of the results may appear counter-intuitive to the reader, and it is vital that the team discuss the meaning of the results and justify why their conclusions make sense.

Another set of assumptions from the previous questions are the interpretation of what it means for a job to be "remote ready." Most teams assumed that technology and the understanding of which jobs could be done remotely would remain static. As demonstrated over the last couple years, though, both the technology available and the perception of what tasks could be done remotely has changed dramatically. Teams that recognized that trends might change demonstrated an important insight into the problem, and this aspect was more readily apparent when examining the results to question three.

Teams addressed the changes over time in a wide variety of ways. One common method was to include more categories of jobs, such as those that are currently remote ready, those that are partially ready, and those that are not ready. Over time, some of those jobs could move between the three states. A small number of teams recognized that some jobs could go from remote ready to not ready and noted that a negative experience associated with working from home in some sectors might push some employers to change their policies to make it less likely that their employees would work from home.

Finally, another consideration whose impact was not apparent until examining the results for the third question was the impact of commute time. The five cities have different contexts with respect to commuting between home and work. Many teams considered the divide between rural versus urban settings, but many did not consider suburban settings and the travel between the different settings. This could lead to surprising predictions, and

without commentary or discussion from the team, a judge had to decide if a team noticed surprising results or just had a different sense of intuition.

# General

Several general issues associated with modeling and the presentation of results are discussed in this section. These topics arise every year but are important things to keep in mind. The topics discussed here include the annotation of graphs, units and linear combinations, assumptions, and the discussion of results.

First, nearly every paper included tables, expressions, or graphs. Every table and graph should include a caption describing the table or graph, as well as a broad overview. All expressions should be numbered and part of a complete sentence with proper punctuation. A description of the various terms in an expression should be discussed, and it should be made clear why some terms are added, multiplied, or make use of some other function. Graphs should include labels and be properly annotated, and a complete description of each graph and table should be described within the narrative.

Another issue that arises each year is the use of models formed by using a linear combination of unrelated variables. For example, a model might be composed of the addition of terms in the form
$$= c_1\, D + c_2\, E + c_3\, F,$$
where $D$, $E$, and $F$ are different, unrelated variables, and the coefficients $c_1$, $c_2$, and $c_3$ are constants. When two variables are added they should have the same units. Adding meters to seconds does not yield a meaningful result. When multiple variables with different units are added each variable must be multiplied by a constant that converts all the terms to a common unit. In the example above $c_1$, $c_2$, and $c_3$ have units so that the linear combination of $D$, $E$, and $F$ makes sense. One common example from this year's event was to define a function like the following:
$$= c_1\, Age + c_2\, Education\ Level + c_3\, Number\ Children + c_4\, Income.$$

The units associated with the constants can then convert each term to the same unit, and one common example is to make each term the monetary value associated with the term. If the constants $c_1$, $c_2$, $c_3$, and $c_4$ were omitted then the expression would not make sense. It does not make sense to add a person's age to their education level. Also, the order of magnitudes associated with the constants should be consistent. For example, a person's income level is likely to be orders of magnitude greater than the person's number of children, and if all of the constants are of the same magnitude then the only variable in the resulting equation that matters would be income. Finally, it was not uncommon for a team to state the values of the constants without any citation or any discussion as to the origin of the values. The reader should not be left to assume how the values were determined.

Another issue, the assumption about independence between sets, has been discussed above in the comments about question two. This is a common topic that arises often. When constructing a mathematical model, the decision to add, multiply, or use some other operation to combine terms is determined based on the modeler's assumption about the relationship between the different factors. These assumptions should be explicitly stated and justified. In the M3 Challenge, it is reasonable to simply state that the time constraints make it difficult to explore something more complicated, but the team should make note of the potential issues and also provide a brief discussion of what might be done to address the problem in the future. Both the terms and the operations within every expression matter and should be justified.

Finally, I will discuss the most basic but important practice of stepping back and looking at results. When a model or algorithm is implemented it is not good enough to simply state the results. Asking whether the results make sense or seem reasonable is the very first step in the analysis of the results. For example, many teams predicted that the number of jobs that would be performed at home for Liverpool, England will decrease in the future, which is a surprising result. Teams that noted this and either provided an explanation or mentioned that it is

problematic tended to receive more positive responses compared to teams that made no mention of their prediction. Given the very short amount of time teams have to complete their work, it is unreasonable to expect complete answers and a complete analysis. However, it is important to be able to recognize potential problems and provide at least some insight into how the problem might be addressed.

# Conclusion

We all have experienced many changes these last two years, and we are gradually beginning to gain a better grasp of how we will interact in the future. One of the changes we have experienced is a sudden and significant change in how we view how we perform our work. In this year's M3 Challenge students were asked to predict how many people would be working from home over the next several years.

Because of the relatively short time span of the predictions, it is possible to provide a good argument for many different models and approaches, and often the simplest model that captures the behavior of a phenomena is the best option. One of the primary burdens for students was to provide a deep analysis and discussion of their approach as well as their results. For example, simply multiplying versus adding two terms requires modelers to recognize and discuss the underlying assumptions that led to their decisions. Additionally, sampling data and constructing regression calculations to approximate the values of parameters within a model leads to uncertainties about the results, and it is important to provide an estimate of the uncertainty and robustness of the results.

Finally, we cannot overstate how grateful we are to everybody taking part in M3 Challenge. We do not take for granted that students are the center of this event, and we recognize their determination to achieve and take part in the full experience of exploring, modeling, analyzing, and reporting their results on a topic they have not fully considered. Their determination and focus is amazing and brings us joy. We also recognize that it is possible because of the support the students receive, and the teacher-coaches and parents who support the students are deeply appreciated. Their efforts play a key role in providing a context in which students can interact and push themselves to achieve and grow.

# Acknowledgments