Judge's Commentary

MathWorks Math Modeling Challenge 2024

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

Introduction

Students were asked to examine the long-standing and interrelated crises of housing shortages and unhoused people in this year's MathWorks Math Modeling Challenge (M3 Challenge). The first task required the student teams to create a model to predict future trends in the housing supply in two of four given cities for the next 10, 20, and 50 years. The second task required them to predict future trends in the homeless population in those same cities for the same time periods. The third task asked teams to consider their results from the first two questions for at least one of the cities to create a model that would help a city determine a long-term plan to address homelessness.

The primary hurdle was developing a model that could be used over a very long time span when a relatively small amount of data is available. Most teams made use of relatively straightforward relationships in time, while other teams employed sophisticated time series techniques to construct an approximation of the potential future patterns. When comparing the papers, however, the bigger question was: how did a team analyze and evaluate the model they developed?

This commentary includes a discussion of the responses to each of the three tasks in separate sections, followed by a discussion of some general modeling issues. One of the main difficulties for the judges in this year's Challenge was to find a way to compare teams that made use of relatively simple models to those making use of more advanced time series methodologies. The fundamental struggle to balance the desire to gain basic insights from a simple model versus the desire to incorporate as many interactions as possible is a recurring theme throughout the discussion that follows.

As judges, we were fortunate to have the privilege to read the papers and provide feedback to the student teams. The effort the teams gave and insights they provided under intense circumstances is truly inspiring. We also recognize that much of the hard work is made possible, in many cases, by parents, guardians, and teachers, and we thank those individuals for their support and the impact they have made on the students. We are also grateful for MathWorks, the sponsor, and Society for Industrial and Applied Mathematics (SIAM), the organizer, of M3 Challenge, who provide the resources and structure that make it possible to bring us all together.

Task One

In the first question, teams were asked to predict the changes in the housing supply for two cities 10, 20, and 50 years in the future. The provided data included the total number of housing units over 13 years, and a link within the data included a breakdown for different kinds of housing. Team members had to decide what parts of the housing market to model and which factors impact the housing supply.

The majority of teams made use of a relatively simple model with coefficients estimated using regression techniques to approximate the total number of housing units as a function of time. Some teams determined the number of housing units as a function of economic variables and then found expressions to model those variables as functions of time. For example, some teams decided that the changes in the number of housing units depends on current employment rates, income, population size, and other measures of economic activity. These teams then found approximations of total housing with respect to these different factors. This approach allowed teams to make use of different models for the different variables. For example, income might grow exponentially while employment rates may experience seasonal changes.

In addition to deciding what variables impact the number of housing units, teams had to decide what types of housing units to model. Most teams developed models for the total number of housing units. A smaller number of teams divided the housing market into multiple components. For example, small rental units, condominiums, small single-family dwellings, and larger housing units were considered as separate variables. The motivation for examining different kinds of housing types is that they can be developed at different rates and the resources dedicated to their construction can vary depending on local economic circumstances. One advantage to this approach is that it provided greater flexibility in exploring different policy options when addressing the third question.

Once the variables were established, most teams made use of a regression method to construct a final approximation. The most common approach was to estimate the total number of housing units as a linear function of time through the use of a linear regression technique. Another popular model was a logistic model for the total number of housing units as a function of time. Teams that decided to make use of other variables besides time tended to use a multivariate linear regression technique. Some teams noted there is a periodic component in the data and added a trigonometric term to their model to attempt to accommodate the nature of the data. This was often done to better mimic the data, but it was rare for a team to provide a reason why an oscillation might be expected. Other models included exponential and logarithmic, as well as growth modeled as a square root of the change in time.

The task set forth in the first question is a difficult one. The long-term trends in the housing supply is a complex phenomenon, the data is limited, and the time span is quite long. In this situation, starting with a simple model is appropriate and can be just as effective as a more complex approach. Regardless of the modeling approach, the long time span for the approximation (50 years) requires a structured exploration of the final model. It is vital to inform the reader of the potential problems when trying to predict the long-term behavior of a complex phenomenon.

As a way to test their model, most teams reported a coefficient of determination, R^2. It is also important to provide a qualitative exploration of the residuals in the data. The existence of patterns in the residuals can be a good indication of the efficacy of a model, and they can provide insights into the potential deficiencies in a model. Given the short time span to respond to the questions, insights into the models are vital for identifying potential changes that should be explored in the future.

In addition to examining the residuals for a model, some teams examined the impacts that would occur for small changes in the model. Compared to previous years, more teams took part in a structured sensitivity analysis to investigate which aspects of the data or the model had the biggest impact on the final results. Teams used a variety of approaches, including making small changes in the data, removing data points, or making small changes to calculated parameters. The magnitude of the resulting changes provided direct insight into the reliability of the model. There is no preference in how a team performs this analysis. Given the short amount of time, it was not possible to perform a complete and comprehensive sensitivity analysis, and judges did not expect a complete battery of tests. If a team demonstrated an appropriate examination of the model, the efforts were likely to make a positive impression on the judges.

Several teams included a section in their report called "sensitivity analysis" but for some teams the section did not include an exploration of the sensitivity of their model. Given the continued growth in the number of teams taking part in at least one kind of exploration of the potential impacts that might occur in their results for subtle changes, it is important that these kinds of analyses be included in a team's evaluation of their model.

The kinds of sensitivity analyses and discussions of uncertainty that are appropriate tend to be more easily executed for the simpler models used by most teams. Some teams, however, made use of much more sophisticated models. For example, some teams treated the data as a time series and made use of methods such as AutoRegressive Integrated Moving Average (ARIMA) methods. Some teams noticed that the data had some periodic patterns and implemented a Seasonal ARIMA (SARIMA) method.

These kinds of time series methods require many assumptions and have several parameters that must be defined. When a team makes use of a specialized or advanced method it is expected that the method be adequately described and properly documented. For example, ARIMA methods require that the data be stationary, and the requirement and what it means should be explicitly described by a team. There is also an expectation that there is an adequate number of data points in the time series. A team should conduct appropriate tests to demonstrate the data meets the inherent assumptions associated with the techniques employed. For example, in the case of an ARIMA method it is expected that an Augmented Dickey-Fuller test be conducted to confirm the data is stationary. Additionally, there are other methods that should be used to identify the values of the three parameters associated with an ARIMA model. Using such techniques, a team should clearly identify the relevant parameters, their meaning, and how they are determined.

Several other more sophisticated modeling methods were also used by a small number of teams. For example, some teams made use of a Black-Scholes model to account for uncertainties in the data. Some teams made use of machine learning algorithms such as a random forest model. In this situation, though, the use of such methods may be problematic. The small number of data points, small number of related variables, and long time span makes this a difficult problem. A complicated model may overfit the data in an inappropriate way, while a simpler technique can be more appropriate as well as being open to a more straightforward analysis of the model itself.

Many of the more advanced modeling methods that teams used provided predictions based on a time series analysis of the data. While these methods may have provided good predictions, it is important that teams ensure that the results include caveats about the limitations of the methods and detailed quantification of the uncertainty associated with the predictions. The sources used by the team should also be carefully documented. Over the last several years the growing use of more advanced techniques has led to numerous discussions between judges, and there is not a consensus on how to interpret their use. The problem statements often ask for predictions, and they also ask that the predictions be made using mathematical models.

One of the key roles of a mathematical model is to provide insight into a problem and provide a tool to analyze and understand a complex system. The problems posed do not have a definitive answer, and every model has deficiencies that may be addressed in future modifications. There is always an implicit question, was the model used to facilitate a deeper understanding of the situation? It is expected that a model will be used to go beyond a prediction and be an effective tool to promote greater insight into the problem. Just like any other tool, it is also vital to understand and justify the use of the tool itself. The tool must be investigated and analyzed to recognize its inherent limitations as well as provide a pathway for future improvements.

Task Two

The task set forth in the second question was to predict changes in the homeless population for the next 10, 20, and 50 years. Teams tended to struggle with this part of the problem more than the other two questions. The straightforward statement of the question masked the difficult nature of the task. For the most part, the approach used by teams tended to mirror their approach in the first question.

The data provided for this part of the problem is similar to the data provided for the first question. The question and the data have the same limitations and potential problems as described in the previous section. Just as in the first question, a large majority of teams employed linear regression to construct a linear approximation as a function of time. Some teams found other variables, such as income, inflation, or other economic indicators and then used multivariate linear regression to construct an approximation.

Given the complex nature of homelessness and access to housing, this was a difficult problem to address. Combined with the 50-year span for the prediction and the brief time to construct an approximation, a simple model with a detailed analysis was an excellent first step in trying to better understand the potential modeling issues. The primary difference in the papers was the level of analysis of the model and its results. Teams that were able to explore the model and state results with appropriate caveats tended to make a better impression.

Just as in the first question, some teams made use of much more sophisticated time series methods, such as ARIMA and SARIMA. These methods require that several parameters be defined, and implementing the algorithm is a non-trivial task. For these reasons, teams should not assume the reader is familiar with the more sophisticated methods discussed, and the details of the implementation should be discussed. As a rule of thumb, a reader should be able to take the information in a report and reproduce the team’s results without reading their code. It should be clear what parameters are required in the algorithm, what they mean, and why the team chose particular values. It should not be assumed that the reader is familiar with more specialized techniques. Finally, the team should explore what happens when the parameters are changed or when small changes in the assumptions or data are present. The reader should have a good idea of the limitations and potential uncertainties that occur when implementing the team’s approach in a different situation.

The most common approach to the second question was to construct a model of the total number of unhoused people. Another common approach was to construct a model of the percentage of the unhoused population with respect to the total population. A small number of teams broke down the unhoused populations into different categories such as individuals, families, or transient populations, and constructed different models for each subgroup. This approach demonstrated an important insight into the problem and showed that a team recognized a key component of the complexity of the system.

With respect to transient populations, many teams struggled with the role of migration. Most teams did not address this or made assumptions that minimized the role of migration within a given community. Other teams relied on past trends or assumed a simple relationship based on the larger population of the city. A few teams attempted to build a relationship based on the larger economic context and incorporated their model as a component of their overall results. This was a difficult task especially given the time constraints and difficulty in obtaining additional data.

Finally, a small number of teams made use of Monte Carlo simulations to try and simulate the changes in the status of people over time. The methods and dependencies varied widely among teams. Every year we see several teams use Monte Carlo simulations, and presenting the results is a difficult task. There are many details required as part of an implementation of a Monte Carlo simulation, and teams must decide several different probability distributions and interactions. To clearly state the details and then clearly state the results and associated uncertainties of a stochastic simulation is extremely difficult. This is compounded by the short amount of time in which the teams have to make decisions, implement an algorithm, run the simulations, collect and analyze the data, and then create a coherent presentation of the results.

Task Three

The third question asked teams to use their results from the previous questions to construct a model that could be used to develop long-term policies to address homelessness. An additional part of the task was to consider the adaptability of their model to the potential impacts associated with unforeseen situations such as natural disasters, economic distress, or changes in migrant populations.

Many teams interpreted this to mean a particular policy should be explored. Teams determined their policies in several different ways. Most teams defined a policy and then constructed their model. In some cases. the team’s policy was not necessarily well defined and was implied using an extension of the models from the previous questions. Other teams took a different approach and used a model to combine the previous results. They then used their model to determine a policy that would have the greatest impact.

It is an impressive achievement to be able to integrate two different models and build on them in such a short amount of time. It is even more impressive when a team can combine multiple models into a general form and then gain insight into the situation as a way to determine a policy that will address a problem in an optimal way. Teams that were able to do all these tasks and then describe a coherent policy based on their analysis of their model showed how modeling is a process and demonstrated the value of introspection in examining and building on existing results.

One of the more common proposed plans was to promote the construction of new housing. Some teams proposed various incentives to increase housing while others proposed giving direct aid to people to allow them to purchase access to housing. Teams tended to rely on different kinds of policies depending on whether they examined cities in the United States or in the United Kingdom. Teams that examined the cities in the United States tended to propose indirect methods that would promote easier access to housing. Teams that examined the cities in the United Kingdom tended to offer policies for government actions offering more immediate and direct impacts to promote changes in housing facilities.

Regardless of the policies, a common hurdle for the teams was to decide what kind of new housing facilities should be made available. Most teams treated housing as a single type, but a few teams divided the housing supply into separate categories. These teams recognized that the creation of luxury housing, smaller single-family dwellings, and apartment buildings had different costs as well as different timescales with respect to development. Some teams included separate categories for mobile housing such as recreational vehicles, trailers, and even boats. The use of different categories allowed these teams to determine a more nuanced methodology to allocate resources as well as predict the time required for the different kinds of housing to become available.

Most teams used their approach to examine one city, while some teams extended their model to two cities. Since the problem statement asked teams to examine *at least* one city and given the difficult time constraints, there was no penalty associated with focusing on one city.

Finally, the third question also included a request for teams to examine the adaptability of their model if unforeseen circumstances occurred. A few teams examined at least one potential circumstance, but most teams did not provide detailed analysis for any particular event. This extra aspect may have been too much extra work given the constraints of the event. Every year the limited time and resources available to teams require them to make choices about what can be addressed, and the judges generally struggle with trying to find a way to balance the choices that are made. This year, however, it was clear where students struggled to make progress.

General Comments

One of the first decisions that arose for the teams in addressing the sequence of questions was deciding which country to focus on. The students had to choose between a pair of cities in the United Kingdom or a pair of cities in the United States. The trends, history, and underlying causes of homelessness differ between the two countries. It is appropriate to use different models for the two contexts, and the different models may appear to differ in complexity.

Elegance in Simplicity

It is vital to note that M3 Challenge is not about constructing the most complex or intricate model. The primary purpose of constructing a mathematical model is to gain insight and create a set of tools to better understand a given set of phenomena and explore the connections within a complex system. A more complex model, especially as a first step, can be detrimental to accomplishing this and to obtaining a deeper understanding.

This is a common issue, and it is often discussed in terms of the parsimony principle. Rather than use the most elaborate technique available, there is both utility and elegance in simplicity. The model that makes use of the smallest number of parameters and isolates the minimal number of interactions while also demonstrating the core behaviors of a situation can be the most illuminating. Such a model may be lacking important features, but it can help guide an investigator to understand the fundamental relationships between variables. After a deeper understanding of the relationships are revealed, the next steps in the modeling process are to slowly add variables and more interactions, and to employ more complex ways to approximate the interactions. We do not expect teams to be capable of developing a complete model. We expect a model to be presented that identifies a few of the most important relationships followed by introspection of the model itself.

The Logistic Model

With respect to the types of models used, one common model was a logistic model. Many of teams approximated the number of housing units as a logistic function based on the idea that there is limited space in a given city, and once a maximum density is reached the city would not be able to create more housing without resorting to extreme measures.

A logistic model is a reasonable choice in this situation. The main caveat is that a team should provide a good justification for their model. Once the model and the motivation are stated, the focus should be on the details of estimating the relevant parameters and the resulting analysis of the final model. One of the downsides with the logistic model is that it can be difficult to estimate the parameters. If the data is not over a long enough time period to get close to the long-term saturation level, then the resulting approximation can be quite sensitive to small changes in the data. It is important for the investigators to examine what might happen if there is a small change in the data or the assumptions.

Another challenge associated with a logistic model is making the reader aware of how the parameters were estimated. Some teams simply used the data and regression from a software library to approximate the parameters. Some teams used other means. For example, a few teams looked at other cities with very high population densities to establish an upper bound on the density of housing units, and they then used the area of the cities in question to estimate the highest possible housing density. Either approach is good, but it should be clear to the reader how the results were obtained.

Teams should be aware that for any calculation, other teams will also be faced with a similar challenge and will likely use a different method. A team should explicitly state how a calculation was made and not assume the reader will find their approach as intuitive as they themselves found it to be.

The larger issue of approximating parameters is a general problem that arises every year. It is not uncommon for equations to be stated including values for coefficients that are not justified or discussed. Given the small size of this year's dataset this was an acute problem. A judge is likely to have a more positive response to a paper when the values of parameters are clearly stated and the methods used to obtain the values are clearly stated and discussed.

The origin of the estimates of parameters can be more confusing when it is not clear what dataset a team used. Many teams made use of the dataset that was provided, but in many cases they did not state which data was used or cite the data source. Even if a team uses the data provided, the source should be clearly stated and a citation provided. This is explicitly noted in the data statement. When teams augment the data, it can be confusing for judges to keep track of what information a team is using; judges cannot make assumptions about what data a team is using.

Another issue arose this year with respect to the logistic model. A small number of teams decided that the carrying capacity should change in time. The reasoning provided was that changes in the maximum housing density depend on changes in technology, changes in zoning laws, and other factors that occur over a long time span. This is insightful and showed that a team thought deeply about the processes required to create new housing. At the same time, however, the logistic function itself is derived based on an assumption that the carrying capacity is a fixed quantity. I spoke with some of the judges about this, and the idea of a time varying carrying capacity was met with diverse opinions. Some thought it was clever, but others were troubled that if the carrying capacity changes over time, then that would impact the derivation itself and lead to a potentially very different function. I personally struggled with this and have not yet reconciled the conflict.

A Tale of Two Cities

Finally, an issue that was the subject of some debate between judges was the consistency of models for different cities. In the problem statement, the teams are asked to come up with a model (singular) that can be used to estimate the long-term trends in two very different cities. Some teams decided that the differences between the cities varied enough that two different models should be used, and often the teams provided a strong argument why this is the case. Since the problem statement explicitly stated that one model should be used, many judges interpreted this to mean a single general model should be developed that could be employed across a wide variety of circumstances. A team that used different models was not penalized, but a team that created a more general model for both cities was generally thought to be closer in line with the nature of the task.

Conclusions

The questions asked in this year's M3 Challenge centered around two closely related crises—the housing shortage and homelessness. These problems are quite complex and directly impact the lives of many, many people. Teams were asked to construct models of the housing supply and homelessness for two different cities and to use their results to provide a tool that could be used to devise a long-term plan to alleviate the impacts of the problems associated with these interconnected issues. .

Once again, the teams taking part in the event extended a tremendous effort, and we continue to be impressed by the dedication and talent of the students who participate. We are also grateful for the people who support the students and recognize that their efforts are vital for the success of the event. This is a rare opportunity for students to explore a complicated problem with no clear answer, do so within a team, and then share their results in a formal report. M3 Challenge is a formative experience made possible with a wide range of support from parents, teachers, sponsor, and organizer. Thank you all!

Acknowledgments

I am grateful for the direct aid provided by Kathleen LeBlanc of SIAM. Her careful editing and direction made this a better document. Once again, I submitted this document to her late in the process, but her professionalism and expertise shine through to greatly improve the result. Thank you!

I am also grateful to Karen Bliss, who, in her first year in the role of director, has taken on an enormous challenge with aplomb and has worked diligently to ensure that M3 Challenge continues to thrive. Her efforts are vital to making this event a success, and the positive impacts on the students is a testament to her dedication and efforts.

Finally, I wish to thank the people at MathWorks whose support, expertise, and resources make this event possible.