Fox-Like Transportation Modeling

Andrew Harris

CE-599 Data Science in Transportation

May 4, 2018

Introduction

Good science hinges on the accurate collection, representation and analysis of data. It is how observations become actionable predictions. In recent history, public and private institutions began compiling data on an unprecedented scale. Tapping this vast reservoir of open source data seems like a boon for modeling the objective world; however, without the careful direction of data collection preceded by a direct question, research, hypothesis and experiment this data has not carried the weight of more pointed observations in controlled experiments. Although, when predictions using this data are successful, it turns the scientific method paradigm on its head. Making use of large data sets, data scientists answer questions often unconnected to the data collection process. This tangential use of data compounds many of the pitfalls of statistical analysis e.g., bias, noise, generality. Transportation engineers use employment, income, population size, and many other variables in models. David Hartgen suggests that insufficient local data and model complexity hampers forecasting for untolled urban road facilities. If used well, these growing repositories of data indicate greater accuracy in travel demand forecasting for these facilities and other areas lacking sufficient data, if the knowledge within these datasets can be pinpointed.

The State of Transportation Forecasting

Estimating travel demand accurately results in a facility with an acceptable cost benefit ratio that has adequate capacity to accommodate traffic flow while having a manageable impact on the surrounding regions. The 4-step method, trip generation, distribution, mode split, and assignment, is still the primary forecast model. Its results are used to determine impacts and design features of a proposed facility. Although these factors are treated as deterministic, each is based on assumptions with variable factors. As a result, fundamental problems in forecasting emerge due to biases that often result in an overstatement of need and minimizing a project's cost. Tests of the 4-step model have a high range of uncertainty showing a great difference in actual and predicted traffic. A 50-year retrospective of travel demand forecasting models focused on accuracy in predicting costs and traffic flow while also testing how relevant the forecast was to public need. Overall, 20-year forecasts leaned toward overestimating traffic while underestimating construction costs with an average variance of ±30-40%.

Other countries are undertaking extensive research to test the validity of their travel demand forecast against the finished facilities. In 2013 The United Kingdom Department for transport found 90% of the 55 projects researched were within a 43% range of actual traffic. Australia studied 14 major toll roads and found traffic was overestimated by 40% for at least 5 of the roads which was attributed to various factors, e.g. less toll road capacity, time of operation, shorter road length, etc. Currently three lawsuits in Australian courts charge that the overly optimistic forecasts were deliberately misleading and threatened investor confidence.

While not as extensive as European reviews, the US is attempting to increase accuracy in travel demand modeling through review of previous forecasts. The FTA focused on 10 projects from 1971 to 1987 and found patronage overestimated by 257% in some cases and costs underestimated by 61%. A more recent review saw an accuracy improvement to 40% for cost

overrun and actual patronage of 75% of the forecasted amount. It seems that the FTA has many recommendations that allow forecasters to build more accurate predictions, but guidelines are rarely followed since there are no formal requirements or accuracy demands.
The Federal Highway Administration's Travel Model Reasonableness Checking Manual uses a predict and provide method of forecasting that focuses on models fitting results and not overall traffic demand accuracy as long as it is "reasonably" close, in this case 20-30%. Empirical studies of previous travel forecasts show overall bias and inaccuracy.

Model accuracy and forecast errors are a result of the surrounding circumstances, e.g. political, fiscal, geographic, institutional, and regulatory. The increasingly complex planning and regulation process results in unattainable goals that fail to challenge underlying assumptions and favor box-ticking over critical thought. O'Toole's 2008 review of 75 long range plans for US metro regions lacked sensitivity analysis and did not show reliable travel forecasts. The 4-step process is often used to conform to policy guidelines but factors like air quality are over-modeled while other factors like economic development and regional accessibility are under-modeled in long range plans.

Increasingly complex issues like carpool mandates, energy constraints, pedestrian-bike systems, climate change etc. are handled with more complex models. However, increased complexity does not lead to increased accuracy as there is usually no way to test model validity and there are few specialists that can successfully integrate the more complex forecasts. It is difficult to predict populations, a critical assumption needed for forecasts, in zones of declining size especially a small zone influenced by a planned route. Projects also seem to take on a uniformity in format to comply with regulation and secure funding, instead of specific sensitivity analysis or "reference class" reviews. Cross- sectional data used to build the 4-step model is an unrealistic snap shot of behavioral relationships used to make assumptions for forecasting. Panel data should be the standard used to track changes in behavior over time.

The 4-step modeling paradigm does not account for how drivers make traveling decisions and is based on unverified assumptions that are too malleable to validly forecast. Any major improvements have included increased complexity and cost but have not produced significant advances. Since the early 1970s there has been a widening gap between new model theory and implementation. Scarce expertise and high costs prevent application, especially in smaller regions. It is susceptible to many variables that could invalidate results like inadequate sampling, over sampling, misspecification, weak calibration, limited before and after testing, and models that are often irrelevant to the alternatives being considered.

The Signal from the Noise

In Nate Silver's book "The Signal and the Noise", he expounds on the two titled concepts elucidating the relationship between signal and noise, and their significance when making predictions. In doing so, he formulates a process to effectively use data that includes: thinking probabilistically, expertise in the subject matter, consensus seeking and trial-and-error. The signal and the noise are terms borrowed from electrical engineers that describe the bit carrying pertinent information and the random phenomena that propagate amongst the signal.

As applied to models of other messier systems the definitions of signal and noise are used a bit looser. An example of reducing the noise in a model occurred during a recent paradigm shift in baseball forecasting. Baseball has always compiled statistics to evaluate players. Scouts leaned heavily on runs batted (RBI) to evaluate the offensive capabilities of players. Runs batted in are simply a tally of the number of scoring runs that occur as a result of a player's at bat offensive production. This metric is flawed in purely examining one player's productivity. What it does instead is measure a player's productivity as a member of a specific lineup of players. A player's ability to drive in runs is largely contingent on his team members' that directly precede him in the line-up ability to get on base.

When analyzing a player's offensive ability, a far less noisy metric turns out to be the on base percentage (OBP). The on base percentage is a measure of hits, walks and times a player is hit by a pitch per at bats, walks, times hit by a pitch and sacrifice flies. On base percentage does not show the full range of offensive capabilities (e.g., stolen bases, penalizes sacrifice flies, and leaves out certain situational plays that do not directly attribute to production). However, it is a direct way to measure a player's ability to put themselves into a scoring position, which like runs batted in necessarily includes hitting ability.

So, in this case the system being modeled is an individual player's offensive productivity. The signal is hitting ability, situational awareness, and offensive productivity. The noise is the ability of other players in the lineup. On base percentage is a simplified tool that gives a truer representation of this feature of the system than runs batted in. Nate Silver suggests that the signal is the truth and the noise is anything that distracts from the truth. The noise however is a part of the objective world, and leaving it in gives a more complete understanding of where things fit into a system as it exists. The problem with noise is its random nature makes it unpredictable and in order to understand a system it is necessary to isolate and identify trending features in the system. That makes noise anything that obfuscates trends in these features.

Noise represents the random variability inherent in complex systems. For instance, there are 163 games in a Major League baseball season. Players often play multiple games a week, and when pontificating about the beauty of the game fans point to the frequency with which a player fails offensively (the best hitters barely surpass a batting average of .300), and the great players distinguish themselves by stepping up to the plate with the confidence that they can be successful even when they are going through a slump. Somehow even with all of this variability in play great players emerge because of their ability to contribute to team wins during the course of the entire season and their entire careers, which can be seen through various different collected statistics as compared to other players. Cleverly Silver points to this ability to move on to the task ahead of you with the confidence that the methodology used will increasingly produce the desired result is a key attribute for a data scientist.

Bayes's Theorem

Bayes's theorem is a case for probabilistic thinking. It is a simple mathematical phrase that states the probability of a hypothesis conditional on some newly acquired evidence is equal to the probability of that evidence given the hypothesis times the probability of the hypothesis over the prior probability of that evidence.

Often the probability of the evidence given the probability of the hypothesis is very high. For instance, if a person has the flu you are likely to develop the symptoms of the flu, but that is somewhat obvious and limited in its usefulness. Using the evidence alone to predict the hypothesis can be misleading. A runny nose, dry throat, cough and fever in concert might point to the flu, but it could also prove to have many other causes. Bayes's theorem is a conditional statement that weighs the likelihood that a result will occur given the evidence. So, if new evidence were to come to light that a person has these symptoms and their entire family has strep throat, then the likelihood that these symptoms point to the flu changes dramatically. When new pieces of pertinent information are introduced it forces data scientist to adjust the model showing a new more resolute understanding of the picture being shown. However, the reason that it is somewhat controversial is because an initial probability must be assumed to arrive at an answer and getting that probability right can be tricky.

Early and often in science and statistical education students and researchers are warned about correlation versus causation. By coddiwompling towards objective truth in this manner it forces the analyst to state their initial understanding of the system they are studying, thus exposing their personal biases i.e., showing what evidence they find pertinent and to what extent.

Fisherian Frequentism runs counter to Bayesian thought in that they define probability differently. The frequentist defines it as the frequency that a repeated event will occur based on prior measurements of that event and designates deviations from that event as error based on the limitations of the information. Frequentists believe that the data varies and they make assumptions about models, which can be seen by the use of standardized normal and t-distributions. Whereas, Bayesians leave room for the model of the data to shift based on increased understanding of what the data intimates. Therefore, instead of zeroing in on the most likely outcomes (taking a narrow view of the data) it assigns probability of likelihoods based on the evidence presented at the time, given the prior probability that it would occur.

According to Silver, the frequentist implication is that a certain amount of data is both necessary and sufficient in reducing error to so that it approaches zero. A Bayesian would suggest that as long as there is bias in a model no matter how much information a frequentist has they will compute incorrectly. They also suggest that the output of the probabilistic prediction using the Bayesian model is a more applicable interpretation of uncertainty that applies directly to the parameter being studied rather than our ability to capture that interval in a given sample.

Bias

There were a fair number of academics and industry people that accurately assessed the housing bubbles in the United States as early as the year 2001, and described the impending collapse. And it is apparent that this view was well known and understood even by laymen. Investment banks had been taking advantage of mortgage backed securities for many years, and there is some evidence that the investment banking industry (which was not demonstrably separate from loan banking) had a hand in driving the housing market toward reckless investment to capitalize on these financial instruments.

Credit rating agencies like Moody's were meant to be a moderating force in this industry, but they are also a for-profit institution, paid by the bank for every financial instrument it is given to rate. The rating is meant to assign the risk the instrument poses to investors. The investment arms of the banks also repackaged the riskier mortgages into collateralized debt obligations (CDOs), which layered them in a way that made it tricky to assess. Whether from fraud or willful ignorance, it was the rating agencies duty to assess the risk of these financial instruments in good faith. They failed to understand the interconnectedness of the system they were measuring even suggesting that they were aware of a housing bubble, but did not think it was a big deal.

It appears—due to the conflict of interest—both the banks and the rating agencies ignored the signal in the data for a more desired if untrue result. The repercussions of the biases resulting from not acknowledging what these institutions did not know were felt throughout the world in the years after the housing collapse in 2008.

Another more innocuous example of bias is in weather forecasting. Weather forecasting is a success story for probabilistic forecasting. It uses a model that breaks the sky into manageable sections. The model operates on the principal that though the causes of weather events on a granular level are unpredictable weather in adjoining areas behave in a similar manner. In a public private partnership an enormous amount of computations are done on an enormous amount of radar data from these sectors and to track weather systems producing increasingly accurate and precise outputs of temperature and storm system paths.

Although these measurements bear out, the weather channel is afraid to run a forecast with the appropriate probability. They fear that their consumer will not understand the probabilistic forecast. If they give a relatively small chance of rain, on the off chance it does rain, they worry the consumer's perception of the weather channel's accuracy will be tainted. The weather channel then purposely over predicts rain biasing their analysis. The difference with this example is that it is clear that the weather forecast have an awareness of what factors they understand and what factors they are ignorant of and do not appear to risk public endangerment with these actions.

Politics Involved

One of Silver's stated principles includes looking at consensus in the professional community. The purpose is not to default to their position, but rather to form a baseline from which to start by examining both the successes and failures of the people that are looking to solve the same types of problems.

The Highway Act of 1964 requires 20-year forecasts of travel demand for highway designs with periodic adjustments to maintain validity. State and local governments submit these forecasts to obtain federal funds for major transit projects. These analyses must justify the need for a chosen project and its cost effectiveness. Public policy requires neutral and objective forecasts, which are evaluated to choose the project with the best cost benefit ratio. The reality of forecasting is its technical complexity and so called objective scientific analysis is often used to advocate for a plan of an interest group or public official possibly to the detriment of the public it serves. The contractors make a great profit and the politicians gain prestige and political influence at the expense of the taxpayer. The following are examples of forecasts that proved to be misleading:

- Ex 1: Forecasts for nuclear power plant construction showed growth in demand for electric power, little environmental impact, low cost of production, and a rise in the cost of fossil fuels and coal. Corporate managers used more optimistic forecasts to mislead investors and government officials while any internal dissent was silenced.
- Ex 2: Forecasts for the C5-A military transport plane used unrealistically low-cost estimates to mislead the federal government into continuing the project.
- Ex 3: Urban Rail Transit Systems- A US federal study found forecasts used to get ten major transit projects approved in ten major cities overestimated patronage and the actual construction cost was 4-5 times higher. As a result, the cost per passenger was much higher than forecasted especially when compared to less expensive bus transit. The study found the lowest cost per passenger was at least twice the amount forecasted with the taxpayer picking up the difference.

The current climate in public policy leads forecasters into ethical quandaries. Agencies with the most optimistic forecasts obtain grants, contracts, promotions and other rewards for their interest groups, leaving many forecasters susceptible to pressure to provide more optimistic predictions than originally obtained. For example, a planner adjusted patronage figures up, and lowered cost figures at the request of a local official for a federal grant. Later, the politician passed the blame for the unrealistic estimates to that planner. "The politician, however, was given credit for getting the rail line built, was labeled skillful at 'cutting the red tape,' and escaped any blame whatsoever for having ordered the planner to falsify the forecasts." [3] Construction costs are routinely underestimated and actual patronage overestimated to secure funds for the chosen project instead of the impartial information the public expects. Even after cost overrun, money is usually found to complete a project underway, usually at taxpayer expense. This misdirection is well established and garners little attention.

Forecasts are often not the result of technical errors or inadequate methods but deliberate misinformation. Planners, engineers, and economists "revise" their results to please the client or risk losing business or their reputation. If the purpose of a forecast is to get a project greenlit, then inaccuracy and bias becomes irrelevant if the forecast is effective in attaining this goal. Also, helping a major project move forward may lead to continued design or engineering work which may provide more income to a consultant than just the forecast fee.

Forecasts are inherently difficult to critique especially by the public. Their accuracy cannot be verified until the projects are underway, and, even if forecasts are proven inaccurate, once projects begin they are frequently completed. Forecasts use intricate data bases and complex mathematics to come up with the need for and cost of large public projects. They are prepared by highly trained consultants or technical staff over long periods of time for sometimes millions of dollars. Untrained citizens or interest groups rarely have the time, training, or budgets to accurately review, replicate, and critique forecasts for these large public projects. Even if experts can review reports, there is often not enough data to accurately critique the forecast and bias is thus hard to prove. In Ethics and Advocacy, Wachs conveys his twenty year opposition to the LA rail transit system, because of its unnecessary cost and because the bus system was a cheaper alternative that would provide better service. Officials combatted his critiques by saying other cities' excessive costs and incorrect assumptions would not be repeated in Los Angeles. Los Angeles had learned from previous mistakes made in other cities and officials took a wait and see approach. The project had millions in cost overruns and probably will not have half of the ridership predicted.

Although forecasts use complex methodology, all predictions are based on critical assumptions about the future. These assumptions, e.g. demographic changes, employment, lifestyle, are often inaccurate making any modeling or mathematical analysis invalid. Professionals, who aren't really trained in making assumptions about the future, are susceptible to pressure to adjust findings to suit a client's need since changing any one critical assumption about the future would skew the results of the forecast in their favor.

The consultant has an overall duty to use his expertise to provide a forecast that uses criteria set by law, regulation, or professional convention to help the public choose between alternatives. Unfortunately, the client regularly sets the result and then the consultant must provide a forecast consistent with client demands even if it is not the most cost-effective choice. The forecaster often rationalizes biased reports by passing any consequence or harm caused on to the client. In the role of advisor, the consultant is just providing information while the client is ultimately responsible for making the final decision.

Robert A. Caro wrote about New York City master planner Robert Moses, "Furnishing misleading information about it was justified; so was underestimating costs. Misleading and underestimating, in fact, might be the only way to get a project started. If they refused to give you the rest of the money, what they had given you would be wasted, and that would make them look bad in the eyes of the public… If they had been misled, that would mean they hadn't investigated the projects thoroughly, and had therefore been derelict in their own duty." Caro believes Robert Moses' numerous major projects later contributed to New York's financial crisis decades later.

Conclusion

If the totality of the profession is using data to advocate for projects rather than attempting to extract knowledge from which to build a foundation for accurate modeling of transportation projects, then there is little hope of progress in accurately forecasting the cost or travel demand in transportation. Furthermore, the proliferation of data in this area will only help in building spurious relationships that allow projects to be steered by the noise. However, with increased GTFS data and GPS data there will soon be access to data in areas that were lacking, that can potentially illuminate trends. Data scientists cannot simply plug in numbers and accept the outputs at face value because they appear favorable. If they limit themselves to quantitative analysis they risk opening themselves up to being steered by special interests. The bulk of a data scientist's effort it seems ought to be in burning away misconceived notions by following leads, crossing off failures, and attaining a thorough understanding of the subject matter so that they can make qualitative judgements on the significance of data they are inputting into their models and the feasibility of the stories they are telling about their transportation projects. This ability is a fine line between the chaos of overfitting data and the stringency of underfitting data. According to Silver, it takes applying data to models in good faith and learning how to tweak the probabilities based on gaining an understanding of the uncertainty about the system and information.

It is difficult to look at the results of transportation demand forecasting, and not sense some cynicism in the ability and necessity for increased accuracy in forecasts. The data shows a range of consistent underestimating of project costs and overestimating of project demand. These results point to a bias towards ensuring that projects are dressed up during the design and production of these facilities. The consistency appears to point to a feedback loop caused by a conflict of interest; furthermore, the range of inaccuracy in missed estimates points to a failure to recognize the uncertainty surrounding transportation projects.

The purpose of travel demand modeling and forecasting is to provide the public with the most cost-efficient use of their tax dollars when making policy decisions and facility investments. Hartgen's suggests institutional changes to the industry for increased reliability in forecasting. The most fox like identifying areas needing research and what knowledge is missing in travel demand forecasting reports, and coordinating the industry around the common goal of tackling those areas using standardized metrics for understanding uncertainty, and ensuring transparency of outcomes to promote self-policing to reduce positive bias.

References

1. Silver, Nate. The Signal and the Noise: Why So Many Predictions Fail--but Some Don't. 1 edition. New York: Penguin Books, 2015.

2. Hartgen, David T. "Hubris or Humility? Accuracy Issues for the next 50 Years of Travel Demand Modeling." Transportation 40, no. 6 (2013): 1133–57.

3. Wachs, Martin. "Ethics and Advocacy in Forecasting for Public Policy." Business and Professional Ethics Journal 9, no. 1 & 2 (1990): 141–57.