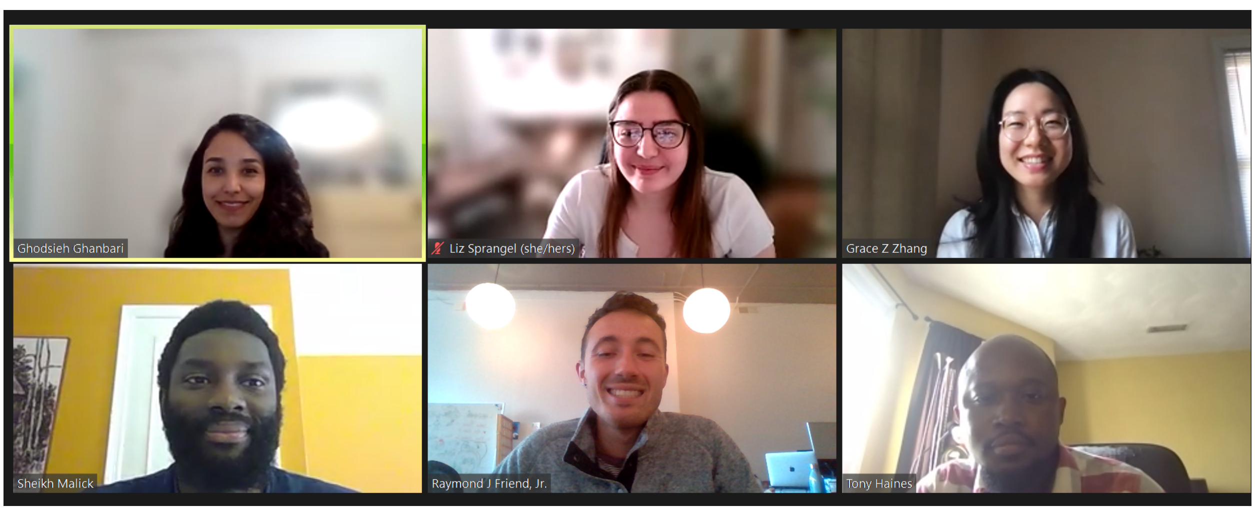
C.H. Robinson | IMA Bootcamp VI | July 2021

Modelling Trucking Cost

A new approach to cost predictions at C.H. Robinson.

Raymond Friend, Ghodsieh Ghanbari, Tony Haines, Malick Kebe, Liz Sprangel, Grace Zhang

Mentors: Kaisa Taipale, Ryan Bieber, Brady Thompson, Cora Brown



Executive Summary

One type of transaction offered by C.H. Robinson (CHR) is a **contractual price**. This is a **long term agreement** with a client (e.g. 3, 6, 12 months) with a **fixed shipping price** agreed upon for the duration of the contract[1]. Although the contractual **price** paid by the client to CHR is fixed, the **cost** that CHR must incur in fulfilling the shipments will fluctuate over time, according to market changes in surface transit services.

According to the CHR 2021 Q2 Earnings Conference Call, over two-thirds of the volume of CHR's newly awarded contracts in surface transit is covered by contracts lasting 12 months. With a pandemic straining the supply of drivers and leading to the highest cost per mile (CPM) on record which is expected to last at least until the end of the year, **it is imperative that CHR improve its contractual pricing predictions**.

Our role as a technical team is to provide the most accurate possible forecast of shipping cost in any given lane over a period of 3-12 months.

To this end, we have invested resources along two promising directions:
1. Refining the purely technical aspects of our quantitative modelling techniques;
2. Developing an innovative method of forecasting which hybridizes technical modelling and expert knowledge about the state of the consumer and producer markets.

Along both avenues, we have encountered preliminary successes. Incorporating external regressors, such as diesel fuel prices, into our models improves their forecasting success. Exploring sophisticated machine learning methods such as LSTM neural networks has led to further opportunities for improvement. Investigating a case study of the lettuce market and its relationship to trucking costs in the lettuce production powerhouse region of Yuma, Arizona has generated a promising route for bridging the gap between market knowledge and technical methods.

In addition to promising preliminary results, we have also encountered numerous challenges to overcome in the future. The contents of this white paper will further outline our findings, successes, limitations, and future directions.

[1] While prices can be renegotiated with clients, it often falls within the strategic interest of CHR to honor the contract and make up for any differences.

Table of Contents

Abstract

C.H. Robinson (CHR) is a third party logistics & supply chain company with a significant hand in North American surface transit. In the midst of a rapidly changing market, lasting pandemic effects, and increasing unpredictability in weather events, it is in the interest of CHR to refine and restructure its approach to predicting the cost of shipping goods over time; specifically, we consider 3-12 month contracts. New models should account for acute spikes or dips in supply or demand, geographic effects along routes, competition between markets, and expert opinions across industries and locations. We present our approach to designing a new system for price modelling.

What We'll Cover in This White Paper

This white paper serves to summarize our group's contributions to pricing trucking prices at CHR. Previous efforts to predict futures for long-term contracts have been criticized for their smoothness (possibly missing anomalies and inherently brief markets), as well as for their poor modeling assumptions (possibly contradicting the better sense of industry experts).

We structure our new approach to modeling trucking prices at CHR as follows:

0. Establish a Baseline
1. Enhance Technical Methods
2. Incorporate Expert Knowledge

0 - Establish a Baseline

In order to get a grasp on this general problem, our group was provided data on a few lanes, the most explored of which included the route starting in Fresno, CA to Chicago, IL using refrigerated trucks (*reefers*) between 2017 and the present. To begin, we replicated many models from a standard toolbox: some univariate and others multivariate.

Univariate Models

We first implemented simple **univariate models**, or models which do not allow for external regressors. The models we implemented were selected from a standard toolbox of commonly used univariate time series models. Each model, when provided historical shipping cost data along a particular lane, produces a forecast for the future cost of shipping along that lane.

To assess model accuracy, we implemented a **cross validation routine with a rolling origin and 12-week forecasts** to backtest forecasts against actual cost values. Accuracy statistics are shown in the table below for one specific lane. The subsequent table contains a summary **description** of each univariate model and a brief **assessment** of its utility in the context of our problem

	SNaive	**STL**	**SARIMA**	**TBATS**	**Prophet**
Pre-pandemic	14.45	5.13	10.33	5.72	7.52
Pandemic	17.65	8.42	8.80	13.12	9.29

Table: MAPE (Mean Average Percent Error) values for univariate models on the *Fresno to Chicago, Reefe*r lane. MAPE values are obtained via cross-validation with rolling origin and 12-week forecasts. The *pre-pandemic* row uses forecast windows contained in the year 2019 only. The *pandemic* row uses forecast windows from the beginning of 2020 up until the present.

	Description	**Assessment**
SNaive **(Seasonal Naive)**	**Copy-and-paste** method: the previous iteration of a season is projected unchanged as the forecast for next season.	Provides a **benchmark** level of forecasting success. Since this is a **highly naive** model, we expect **poor performance**. Indeed all other models in this list perform better (lower cross-validation MAPE value).
STL **(Season-**	**Decomposes** data and applies separate models to the resulting **seasonal and**	**Best performing** among all univariate models we considered.

Trend Loess)	**non-seasonal** components. We used SNaive on the seasonal part, and **exponential smoothing (ETS)** on the seasonally-adjusted part.	Effective at modelling seasonality, **without large risk of overfitting**.
SARIMA **(Seasonal Auto-regressive Integrated Moving Average)**	One of the most commonly used time series models. Focuses on **autocorrelation** and accounts for seasonality.	Did not perform as well as STL on our data, and had the **widest confidence bands** of all models considered. **Risk of overfitting** when training data covers only a short period of time. However, a major advantage of SARIMA is its ability to incorporate **external regressors** (see next section).
TBATS **(Trig seasonality, Box-Cox transform, ARMA errors, Trend and Seasonal)**	Entirely automatic model which accounts for **multiple seasonality and non-integer seasonality**.	Our data only has **yearly seasonality**. If it also had weekly or monthly seasonality, then TBATS may have been more appropriate. As is, TBATS produces **too smooth** a forecast on our data.
Prophet	A model developed at Facebook and is **easy to implement** without specialized knowledge.	**Simple to implement**. Can easily adjust for **holidays** and **changepoints**. (Although we did not take advantage of holiday and changepoint adjustments in our analysis.) This model tends to perform best on short-term forecasts, and relatively poorly on our desired 12-week forecasts.

Table: Summary of univariate models we implemented, including a brief description of each model and assessment of its utility.

Multivariate Models

Next we implemented a few **multivariate models**, or models which accept external regressors. For this part of the analysis we selected three external regressors to incorporate:

1. Maximum **temperature** from the origin region.
2. **Precipitation** from the origin region.
3. National average **diesel fuel prices**.

	SARIMA with external regressors	**Prophet** with external regressors
Pre-pandemic	7.75	6.03
Pandemic	7.94	11.01

Table: MAPE (Mean Average Percent Error) values for multivariate models on the *Fresno to Chicago, Reefe*r lane. MAPE values are obtained via cross-validation with rolling origin and 12-week forecasts. The *pre-pandemic* row uses forecast windows contained in the year 2019 only. The *pandemic* row uses forecast windows from the beginning of 2020 up until the present.

	Description	**Assessment**
SARIMA with external regressors	One of the most commonly used time series models. Focuses on **autocorrelation** and accounts for seasonality. External regressors are modeled as linear regression terms.	Accuracy improved when external regressors were added, both for pandemic and pre-pandemic forecasts. Better performing multivariate model out of the two discussed.
Prophet with external regressors	A model developed at Facebook and is **easy to implement** without specialized knowledge.	This model tends to perform best on simple short-term forecasts, but risks overfitting in other contexts. For instance, pandemic forecast accuracy actually became worse when external regressors were

		added (compared to univariate Prophet model).

Table: Summary of multivariate models we implemented, including a brief description of each model and assessment of its utility.

1 - Enhance Technical Methods

Neural Networks

Time series prediction problems are sometimes a difficult type of predictive modeling problem. As such, one important recurrent neural network that was employed in our forecast was the **Long Short-Term Memory network or LSTM network**. It has the ability to incorporate large architectures that can be successfully trained. We considered a Multivariate Multi-step LSTM model with configurations as follows.

1. One neuron with five features in the input layer comprising the external regressors
    a. Maximum **temperature** from the origin region.
    b. **Precipitation** from the origin region.
    c. National average **diesel fuel prices**.
    d. Trucking **volumes** from the origin region.
2. Fifty neurons in the first hidden layer
3. One neuron in the output layer for predicting cost

The data was first converted to supervised learning and divided into a train and test sets. We trained the model on the first two years while keeping the remaining data for testing and prediction purposes. The Mean Absolute Error (MAE) loss function was plotted to analyze the effectiveness of the model during training, and the efficient Adam version of stochastic gradient descent for optimization. The model was then fitted for 1000 training epochs with a batch size of 72. Using a Vector Output Model, the last three time steps were used as input to forecast the next two time steps. The goal was to determine multiple forecasts using past observations.

Several iterations were performed while keeping the parameters constant and the average root-mean-square error, along with the mean-absolute-percent error were determined. This can be seen in the following figures.

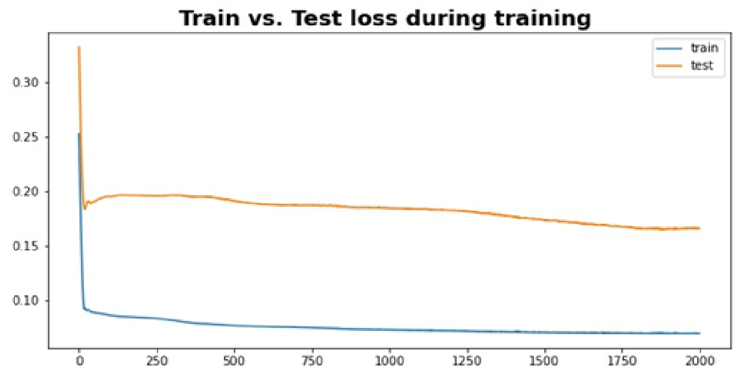


Figure: Evaluation during training. This shows that the model is not overfitting the data. Train loss function in blue; test loss function in yellow.

	RMSE	MAPE
Average error	37.2	13.0

Table. Average error after five iterations.

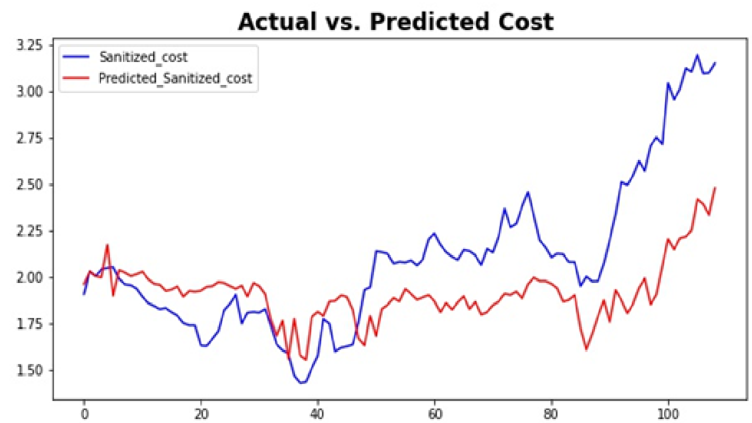


Figure: A graph of **actual cost (blue)** vs. **predicted cost (red)** with a three-week forecast.

Alternative Neural Networks

Many times, neural networks are used for classification, but they can also be used for predictions. We looked at the **NNetar** model in Rob Hyndman's forecast package as well as the **MLP (Multi-Layer Perceptron)** and **ELM (Extreme Learning Machine)** models from the nnfor package. We trained the data from 2017 Week 13 until 2020 week 52. We then tested the data on the first week of 2021 through the 20^{th} week of 2021. The models we looked at were compared based on the testing set.

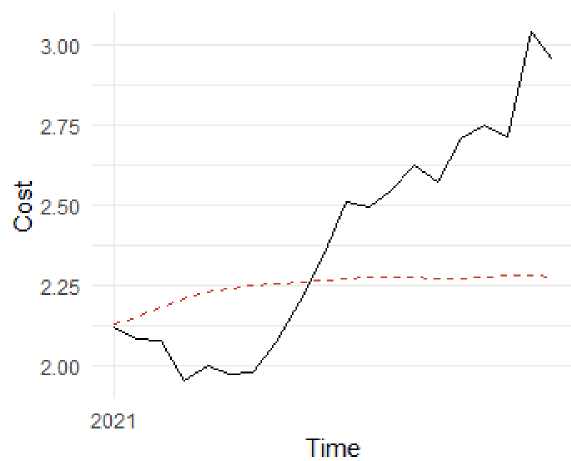


Figure: NNETAR Forecast.

MLP (Multilayer perceptron)

Many people refer to the multilayer perceptron as the hello world of deep learning. It is a good starting point, and in terms of models it can be thought of as a **naïve** model- a good place to start, as well as a **benchmark** for more advanced models. It is a **feed forward** neural network. As stated earlier, we used the MLP model from the nnfor package. The default settings of 5 hidden layers were kept. The visual below shows how the model performed.

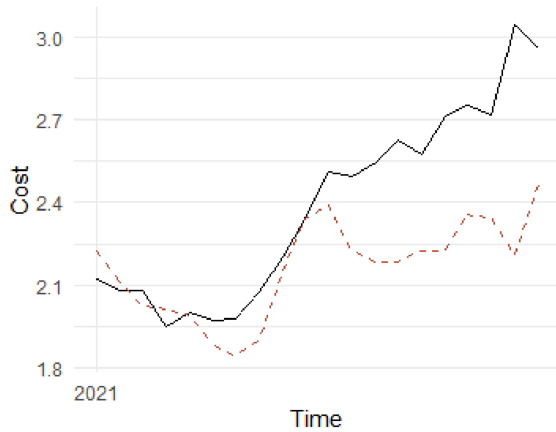


Figure: MLP forecast. See later subsection "Comparing in quantified manner" for numerical comparison to NNETAR forecast accuracy.

ELM (Extreme machine learning)

ELM is another relatively straight forward **feed forward** neural network. Weights don't need to be adjusted. Backers of this technique believe that ELM can outperform algorithms that use back propagation. The visual below shows how ELM performed.

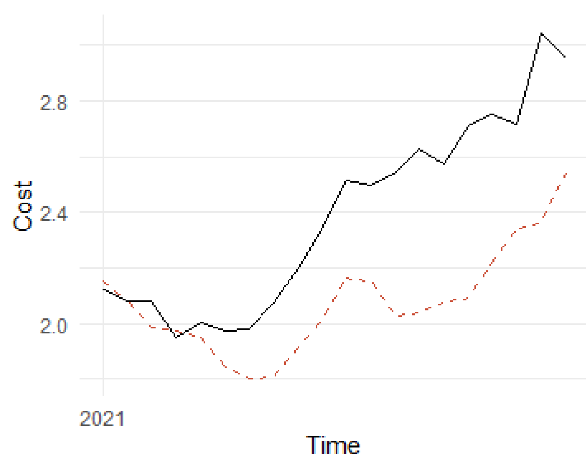


Figure: ELM forecast. See later subsection "Comparing in quantified manner" for numerical comparison to NNETAR forecast accuracy.

Comparing in quantified manner

We decided to use **MAE (Mean average error)** as a comparison metric since it serves as an easy-to-understand, quantifiable measurement; the MAE takes the average of the difference between predicted and actual values for every point in the training set. MAE values for one specific train test split are shown in the table below.

	NNETAR	MLP	ELM
MAE	0.284	0.228	0.316

Table. MAE (mean average error) for three neural networks on one specific train test split (train 2017 Week 13 - 2020 week 52, test 2021 Week 01 - 2021 Week 20).

As we can see, **the models have much room for improvement**.

Fine Tuning

We then looked to see how we could modify these neural networks. A BoxCox transformation was applied to the nnetar model and the model was run again. It resulted in an MAE of 0.325, slightly worse than before. The graph is presented below.

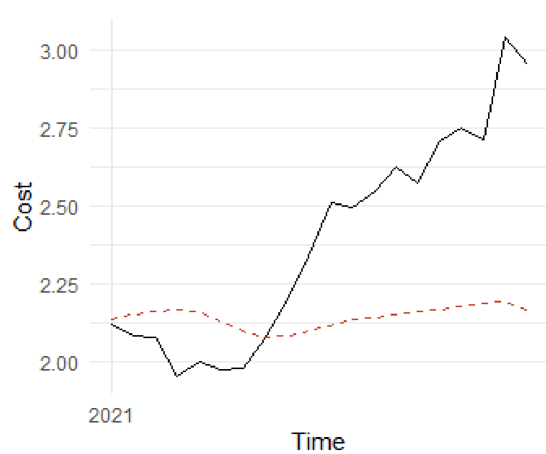


Figure: NNETAR with BoxCox forecast.

For the MLP model (nnfor package), we took advantage of a method that allows us to instruct the computer to find the optimal number of hidden layers in lieu of the default of 5. Upon employing that method, we were able to obtain an MAE of 0.196.

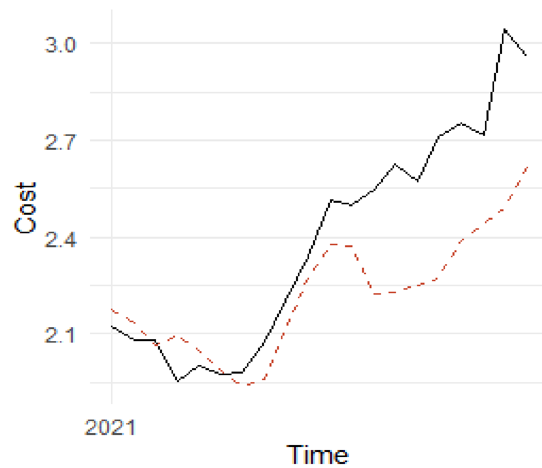


Figure: MLP with Optional Hidden Nodes forecast.

Next, we attempted to modify the ELM model but did not come up with anything obvious. Thus, the MAE did not change.

Summary

We believe that using more sophisticated models could improve accuracy even if they are more costly. For example, although it took longer to run, the MLP model resulted in a better MAE when the **number of hidden nodes** was optimized. Creatively adding **pertinent external regressors** would also positively impact model accuracy. **More data** from which to train the models could also prove fruitful.

However, neural nets are not a panacea. Therefore, it is important to not singularly focus on neural nets. Perhaps they are not the best tool to approach this problem.

2 - Incorporate Expert Knowledge

One criticism of technical forecasting methods is that they typically cannot accommodate the influence of expert opinion. While a domain expert may have knowledge about **market cycles**, **strategic inten**t, the **effects of current events**, and so on, traditional time series modelling techniques do not easily allow one to adjust one's predictions in response to such knowledge.

At CHR, expert knowledge is currently incorporated into the pricing process in a **manual** way - before offering a deal to a client, a **pricing expert** personally adjusts the algorithmically generated suggested price.

Our team is taking preliminary steps toward a different, **hybrid** approach of incorporating expert knowledge into technical forecasts[2]. In our preliminary work, a **representation of domain knowledge** is **automatically generated** based on data about the state of consumer and producer markets for shipped goods. Afterwards, that generated knowledge becomes an input into a traditional forecasting model.

Because of the complexity of national freight and shipped goods markets, we begin by restricting our attention to a relatively simple **case study** - that of the lettuce market in Yuma, Arizona.

Case Study Introduction - Yuma Lettuce

The dominant export out of the Yuma region in Southwestern Arizona is lettuce. The corresponding freight market which processes shipping volume of Yuma lettuce exports is Phoenix, Arizona. Because lettuce forms a significant portion of the shipping volume, **forecasts of trucking costs out of the Phoenix freight market should reflect knowledge about the state of the lettuce market**. For instance, market shocks may be triggered by events such as E. coli outbreaks, extreme weather events, and the success or failure of

[2] This method is **not** suggested as an alternative or replacement to pricing experts. Rather, it is a complementary approach which can potentially make the automated portion of price estimation better.

competitor markets in California; and these events ought to reverberate through trucking costs in the region.

Market Scenario Clustering

How do we hybridize knowledge about the current and historical state of the lettuce market with standard time series forecasting tools? This section outlines a rudimentary, preliminary approach used in the context of the Yuma lettuce case study.

1. First, we captured the state of the lettuce market via weekly time series data of lettuce prices[3]. We took both producer prices and retail prices for three categories of lettuce (iceberg, red leaf, and romaine[4]) for a total of six series. We performed PCA (Principal Component Analysis) to reduce the dimension to four principal components. Next, we performed K-Means Clustering with three clusters to categorize each week of the lettuce market into one of three possible **scenarios**. We coined these clusters or scenarios: "Market shock," "Transition," and "Normal market."

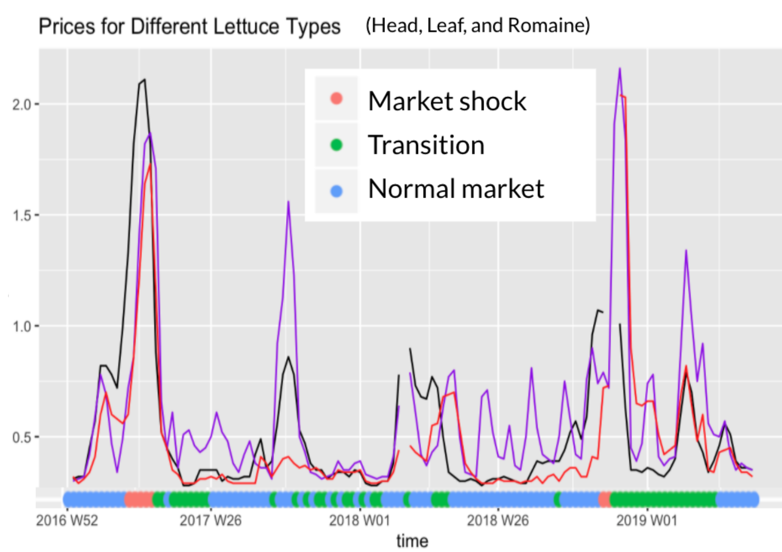


Figure: Producer prices (i.e. prices paid to farmers) and market scenarios. Prices for *romaine* lettuce shown in black; *red leaf* lettuce in red; *iceberg* lettuce in purple. The results of K-means clustering are shown in the multicolored bar below the plot, with corresponding cluster names shown in the legend. Note

[3] Data from the USDA and Western Growers.
[4] Green leaf lettuce data is also available from the USDA, but not used in our analysis due to too many missing values.

that although the line plots portray 3 series of producer prices, clustering was actually performed on a different but related 4-dimensional data set. (3 producer prices plus 3 retail prices not shown, then all dimension-reduced to 4 components via PCA.)

2. Afterwards, the **scenario classification becomes an input into the forecasting model**. To do this, each scenario was transformed into a **dummy variable** time series[5] - the value of the dummy series is equal to 1 for weeks belonging to that market scenario and equal to 0 for other weeks. The dummy variables are then input as **external regressors to a SARIMA model**. A comparison against univariate SARIMA is shown in the next figure.

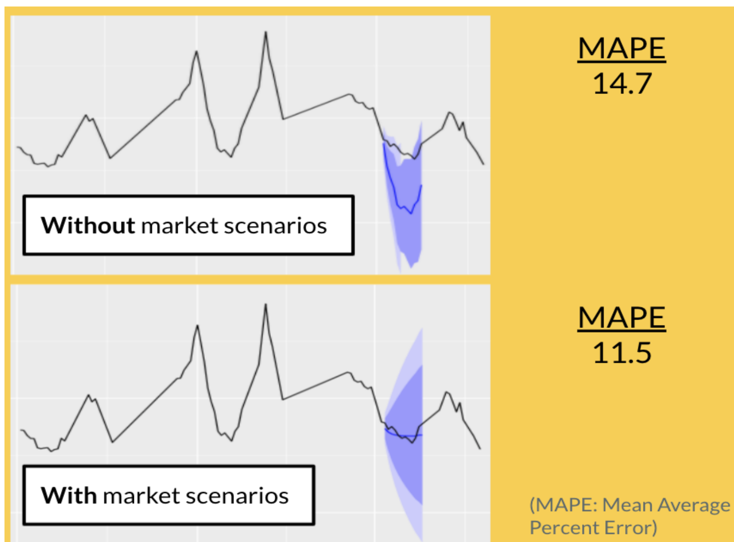


Figure: Trucking cost along the *Phoenix to Chicago, Reefer* lane between 2017 and mid-May of 2019[6]. Black curves portray actual costs, and blue curves depict simulated forecasts. Blue regions represent 80 and 95 percent confidence intervals.

Upper plot shows univariate SARIMA, and lower plot shows SARIMA with market scenarios added as dummy regressors. MAPE values are computed via a cross-validation with rolling origin and 12-week forecasts, with the earliest forecast window beginning in 2019 Week 1. Blue curves specifically show only one example of a forecast window.

When dummy regressors are added, MAPE is reduced by about 3, which demonstrates that the overall approach may be promising.

[5] Note that creating only two of three dummy variables is actually necessary.
[6] We remark that the reason for stopping analysis in mid-2019 was that our lettuce price data was only easy to obtain up until then. Lettuce price data is also available up until the present moment; however it needs to be scraped from USDA Market Reports before it can be used.

However, the resulting MAPE value of 11.5 is still too large to be satisfactory. We believe that the inaccuracy of our model is largely due to the poor quality of data used in the analysis - a) the time frame is too short. SARIMA does not perform well when trained on fewer than 3 seasons, and b) there are numerous missing values in our trucking cost data, which are likely affecting the accuracy of the model.[7] These are all data quality issues that can be addressed with further work.

<u>**Discussion**</u>: Our market clustering approach described thus far is **merely a preliminary step** in the direction of building innovative models that **hybridize expert knowledge with traditional time series models**. There are numerous future challenges which should be addressed in order to increase the potential power of such a hybrid approach. First, we should **expand from this case study** of lettuce markets in the Phoenix freight market to other regions. Significant challenges will be presented when those regions have no single dominant export, and/or have many competitive or codependent markets. Second, steps that should be taken towards a **tipping point analysis**. Such a direction is very exciting, because it has the promise to anticipate market shocks before they occur, and to anticipate the length of transition period after a shock. One difficulty that will present itself is that of distinguishing between seasonal shocks (e.g., summer vs. winter - Yuma actually exports lettuce primarily in the winter, but a variety of other crops in the summer.) Lastly, there are many **technical improvements** which should be made in order to improve forecasting quality. These include gathering higher quality data, and improving the clustering and and forecasting algorithms. One possibility towards the latter is to use Bayesian time series methods (e.g. use market scenarios as Bayesian priors).

Scenario Analysis

One aspect of modelling is to discern the quality of possible models; but the goal of a scenario analysis is to <u>**leverage a quality model to probe the outcomes across a range of future scenarios**</u>. The designated model must be multivariate so that it may be fed artificial values for its exogenous regressors. An example scenario might be, "Drought

[7] We remark also that we recently received better trucking cost data, but have not had time to use it, at the time of writing this white paper. For the purposes of this paper, we have imputed missing values through linear interpolation.

conditions and a 10% spike in diesel prices over the next three months." Our basic method for performing a scenario analysis was as follows:

1. **Construct** candidate multivariate models and compute a summary statistic (e.g., `MAPE`) for each using cross validation.
2. **Identify** the model with the best summary statistic as the designated model M.
3. **Construct baseline futures** for each exogenous variable univariately.
   a. For clustered variables, you may perform a univariate model on each component of the original quantitative data and then categorize each using the same clustering as in training.
4. **Establish a baseline prediction** for cost of trucking by feeding the baseline futures for all exogenous variables into M.
5. **Test other scenarios** than baseline using scaled, translated, or otherwise transformed versions of the exogenous baseline futures into M.
6. **Compare** cost futures under various scenarios.

In the final step, by plotting the future of trucking cost under a range of plausible scenarios can provide us an **explainable alternative** to error bars on M: we can get a grasp of the center and spread of possible futures for trucking cost. Explainability comes from the fact that, trusting our model's fidelity, we can **look backwards** to discern a basic explanation for having followed a certain path in trucking cost.

Codependent Markets

Markets do not exist in isolation: competing, mutually reliant, or otherwise **codependent markets complicate the task of predicting trucking price**.

For instance, pockets of Arizona and California markets compete over the same commodity: lettuce. If one were to perform an analysis of trucking prices out of Phoenix by only considering factors local to Phoenix, one would miss out on explaining the effects that the California lettuce market has on that of Phoenix. In October of 2017, leaf lettuce underwent a restricting of supply due to a significant heatwave in California. Lettuce in

Phoenix subsequently became more valuable and sought after, increasing the demand for trucks out of Phoenix.

This kind of event has the potential to alter trucking costs, and so can any codependency. Our approach to incorporating knowledge of codependencies is to feed our model information local to the markets with which there are high amounts of overlap or reliance. For our lettuce example, knowing the recent weather history in California could have helped CHR to predict a shock to trucking costs in the Phoenix market in October, 2017.

Conclusion

Throughout our three weeks of work, our group has developed a new approach to sophisticating the trucking price prediction methods currently employed at C.H. Robinson. In response to common critiques of the current methods, we have laid out plans to improve the technical aspects of modelling as well as to incorporate expert knowledge at the price prediction level. Our solutions have the potential to drastically improve long-term predictions for trucking price across various lanes.

Appendix

All code and datasets can be found at the following link:

https://github.com/kaitai/IMA_2021

Contact Kaisa Taipale for access.

Our final presentation can be found at the following link:

https://bit.ly/3BXKSdB