Forecasting short term trucking rates

Xiwen Bai

Advised by Dr. Chris Caplice

22 May 2018

MIT Center for
Transportation & Logistics

Introduction

MOTIVATION

Transportation

$1.5 trillion spent in 2015 on logistics and transportation

8% of US GDP

Up to 50% of total logistics costs are transportation cost

Emphasize on lean production and inventory minimization

Trucking

Lifeblood of the US economy

Earned $726.4 billion revenue in 2015

Represents 81.5% of US freight transportation revenue

Forecasting

Transportation budget planning

Economics order quantity

Inventory replenishment

Facility location

Introduction

OBJECTIVE & SCOPE

➢ Objective:

- To develop a forecasting model that predicts both contract and spot rates for dry van on individual lanes for the next seven days

➢ Scope:

- One high-volume truckload (TL) lane as a sample lane for forecasts

➢ Contract vs Spot market:

- Contract: stability for shippers, less flexibility
- Spot: more flexibility, higher rate volatility

Trends in truck rate forecasting research

- Contract rates more frequently studied than spot rates
- Methodology evolvement

Approximated rate function

TL: distance
LTL: distance weight

Regression based method

Market-
Lane-specific variable:
Origin/destination
Tender rejection
Economics of scale
Freight class

Machine learning

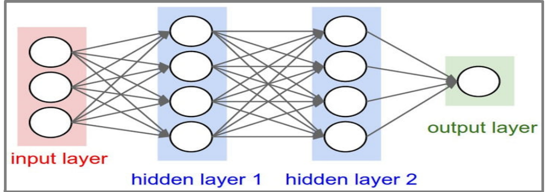
Not often used in trucking
Except Budak, et al. (2017)

Research Gaps

- Most studies using linear regression methods
- Spot market rates have been less studied
- No study considering <u>interactions</u> between <u>spot and contract rates</u>, <u>between rates for a particular lane and its adjacent routes</u>, or <u>between rates and volumes</u>.

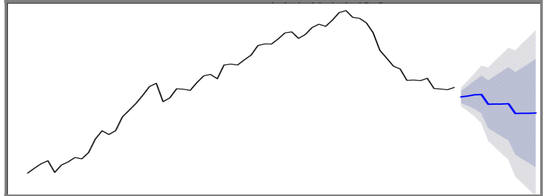
Methodology

Neural Network



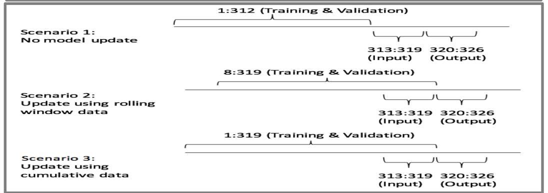
- **NAR**: Nonlinear autoregressive (NAR)
- **NARX**: Nonlinear autoregressive with eXogenous inputs

Time Series Models



- **ARIMA**: Autoregressive integrated moving average

Model update



- When to update model with new information

Methodology

FORECASTING MODELS: NEURAL NETWORK & ARIMA

Model Structure

- **NAR & NARX models**
  - Hybrid of neural network and time series models

$$y(t+1) = f\left[y(t), \ldots y\left(t-d_y+1\right); u(t), u(t-1), \ldots u(t-d_u+1)\right]$$

$u(t)$ is exogenous input; dy and du are memory delays

- **ARIMA model**

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$$

Neural Network Advantage

- Ability to find complex and nonlinear associations between the parameters
- Higher tolerance for errors, robust to noise
- Less concerned with multicollinearity issue
- No need for error distribution assumption

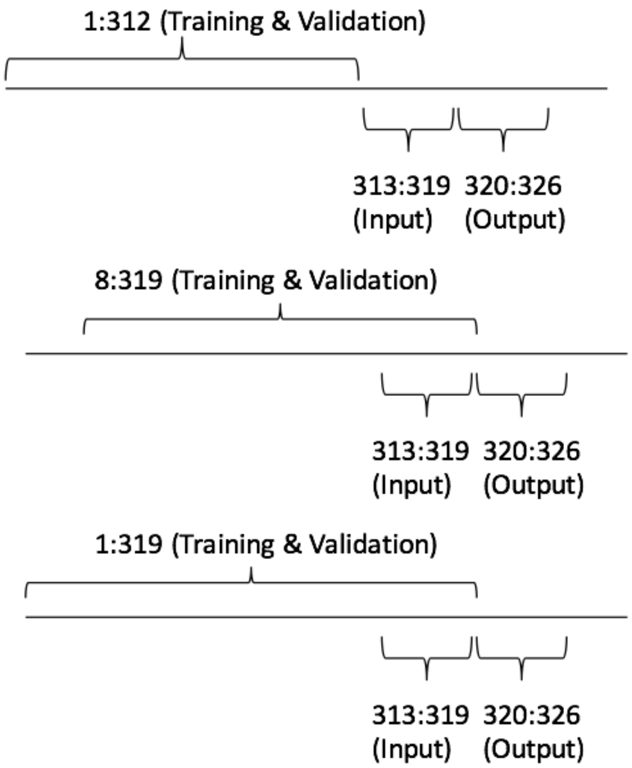
Methodology

MODEL UPDATE WITH NEW INFORMATION

Scenario 1:
No model update

Scenario 2:
Update using rolling window data

Scenario 3:
Update using cumulative data

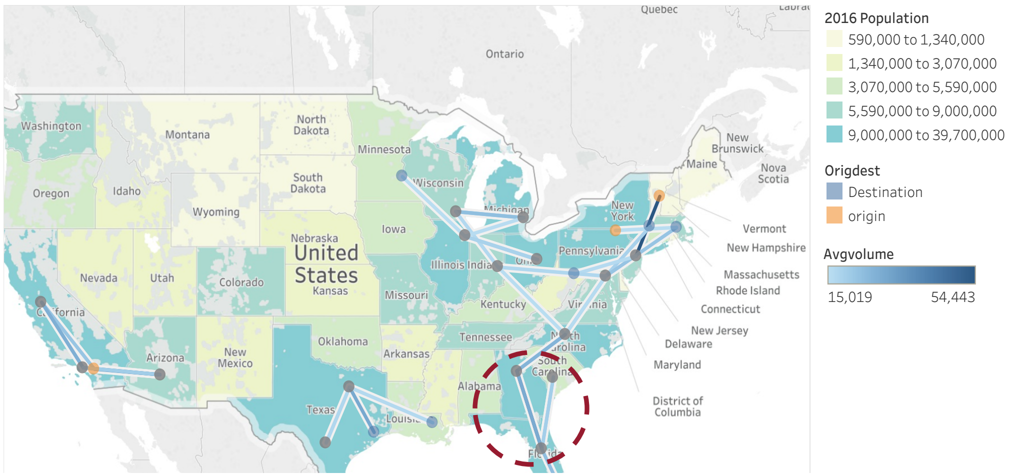


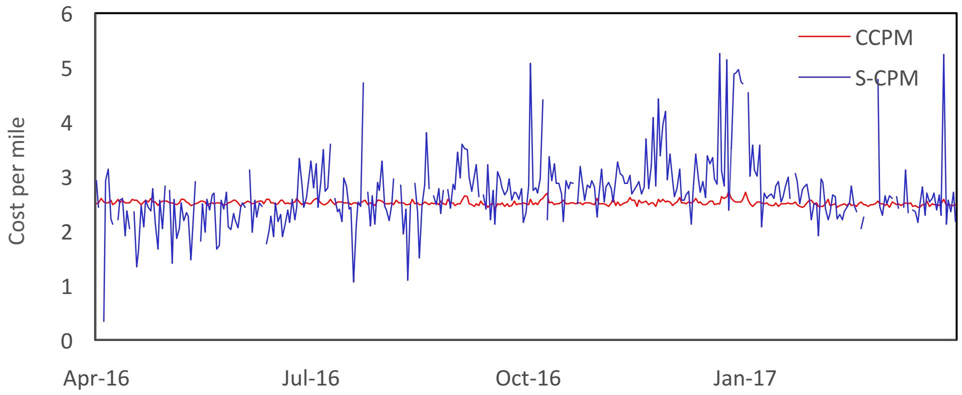
Methodology

DATA

- Georgia Central (GA_C) to Florida Central (FL_C) daily spot and contract cost per mile over a one-year period (1 Apr 2016 to 31 Mar 2017);

- Training: Validation: Test set = 70:15:15





Methodology

INPUT DECISION VARIABLES ANALYSIS

Autocorrelation and cross-correlations between GA_C to FL_C **contract rates** and other variables

Number of lags	[Contract, Contract]	[Contract, Spot]	[Contract, Contract GA_C|FL_N]	[Contract, Contract GA_C|FL_S]	[Contract, Contract GA_C|SC_C]	[Contract, Volume]
0	1	0.25	0.05	0.5	0.08	-0.41
1	0.38	0.11	-0.11	0.23	0.12	-0.03
2	0.22	0.1	-0.05	0.16	0.13	0.23
3	0.04	0.13	-0.07	0.13	0.03	0.33
4	0.01	0.17	0.03	0.1	0.16	0.15
5	0.12	0.19	0.12	0.22	0.08	-0.05
6	0.21	0.16	0.02	0.26	0.07	-0.32
7	0.42	0.14	0.05	0.4	0.06	-0.34

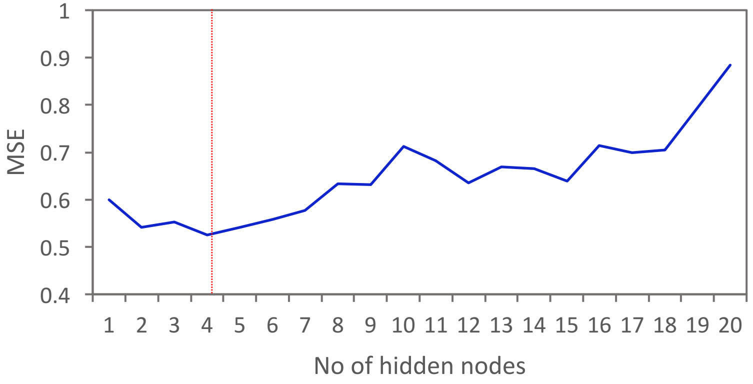
Autocorrelation and cross-correlations between GA_C to FL_C **spot rates** and other variables

Number of lags	[Spot, Spot]	[Spot, Contract]	[Spot, Volume]
0	1	0.25	-0.18
1	0.3	0.21	-0.09
2	0.28	0.23	-0.01
3	0.3	0.11	0.04
4	0.23	0.13	0.02
5	0.27	0.06	0.05
6	0.26	0.12	-0.08
7	0.23	0.14	-0.16

Results and discussions

NAR RESULTS FOR SPOT RATES

MSE for the validation set with different numbers of hidden nodes (*Nh*) (*dy*=7)



NAR model results for spot rates with different feedback delays (*dy*)

Feedback delays (d_y)	Hidden nodes (N_h)	MSE for validation set
7	4	0.494
8	2	0.507
9	5	0.543
10	2	0.527
11	3	0.526
12	4	0.539
13	2	0.558
14	2	0.544

The same model for each *du* and *Nh* is run 10 times to get stable results. MSEs are the average values of 10 runs.

Results and discussions

NARX RESULTS FOR SPOT RATES WITH CONTRACT RATES AS INPUTS

Input delays (d_u)	Hidden nodes (N_h)	MSE for validation set
1	6	0.544
2	5	0.513
3	4	0.511
4	2	0.552
5	3	0.501
6	2	0.530
7	2	0.559

Note: *dy*=7. The same model for each *du* and *Nh* is run 10 times to get stable results. MSEs are the average values of 10 runs.

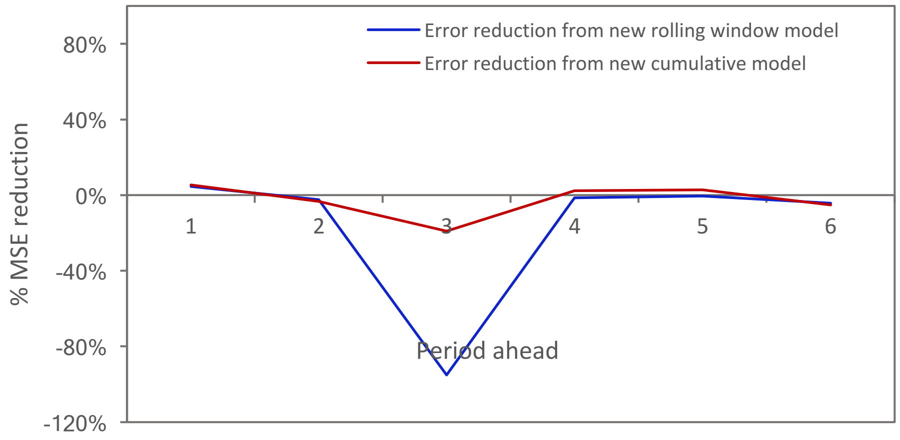
Results and discussions

MODEL UPDATE WITH NEW INFORMATION FOR SPOT RATES

- Updated models do not perform better than the original model in most cases.

- **Reason**: parameters (numbers of hidden nodes and feedback delays) trained in the original model not stable due to high level of volatility and noise in the original data

MSE reductions by updating models with new information for spot rates



Results and discussions

NAR AND NARX RESULTS FOR CONTRACT RATES

Input variables	Input delay (d_u)	Hidden nodes (N_h)	MSE for validation set
CR	7	3	0.00239
CR, SR	7	3	0.00235
CR, GA_C|FL_S	7	1	0.00205
CR, Volume	7	2	0.00192
CR, SR, GA_C|FL_S	7	2	0.00226
CR, SR, Volume	7	1	0.00210
CR, GA_C|FL_S, Volume	7	4	0.00211
CR, SR, GA_C|FL_S, Volume	7	1	0.00200

Note: Feedback delay (*dy*) is set to 7 for all cases. **CR** stands for spot rates, **SR** for spot rates, GA_C|FL_S for contract rates on GA_C|FL_S. The same model for each *du* and *Nh* is run 10 times to get stable results. MSEs are the averaged values of 10 runs.

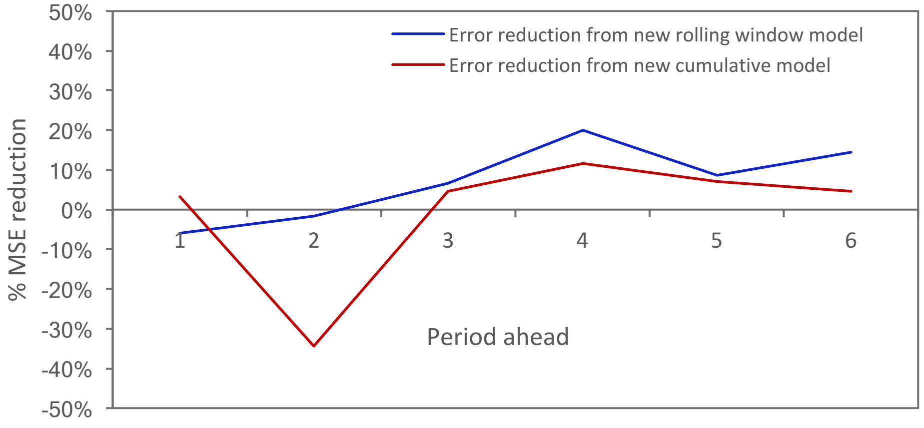
Results and discussions

MODEL UPDATE WITH NEW INFORMATION FOR CONTRACT RATES

- Updated model performs better than the original model from 3-period ahead onwards

- Rolling window model performs better than the cumulative model

MSE reductions by updating models with new information for contract rates



Results and discussions

RMSE COMPARISON BETWEEN NAR(X) AND ARIMA

Rate type	Model type	Best model	RMSE for 7 days (rolling forecast)
Spot	NAR(X)	NAR with d_y=7	0.56511
	ARIMA	ARIMA (6,0,2)	0.60167
	% Difference		-6%
Contract	NAR(X)	NARX with d_y=7, d_u=7	0.03286
	ARIMA	ARIMA (7,0,1)	0.03082
	% Difference		7%

Note: % Difference is calculated as (RMSENAR(X)- RMSEARIMA)/RMSEARIMA , as a way to measure relative performance of two types of models.

Conclusion

KEY FINDINGS

1. **Spot rate**: NAR ✓ vs ARIMA ✗ **Contract rate**: NAR ✓ vs ARIMA ✓

2. **Spot rate: Additional information**, such as past values of contract rates
Model update with new information does not improve forecasting accuracy

3. **Contract rate**: Adding **volume, rates on adjacent routes**,
retraining of the model increases the model's performance

4. Much higher accuracy and less forecasting variability for contract than spot rates

5. No short-term information transmission between spot and contract rates

Conclusion

CONTRIBUTION

1 Aid supply chain decision making

- Transportation budget planning
- Economics order quantity
- Vehicle routing
- Facility location

2 Forecasting guideline for practitioners

- how to select input variables
- what model to use
- when to update model with new information
- what forecasting error expected from model

Thank You!