Portland State UNIVERSITY

Predicting Average Daily Traffic on Local Roads: Parametric or Nonparametric Methods?

Northwest Transportation Conference
March 11, 2020
Session 7B: Traffic Probe and Vehicle Data

Jason C. Anderson
Portland State University
Presenter

Rohan Sirupa
Portland State University

Avinash Unnikrishnan
Portland State University

Miguel Figliozzi
Portland State University

Maseeh College of Engineering and Computer Science
PORTLAND STATE UNIVERSITY

TTP Lab

Motivation

- MAP-21 and FAST Act Both Emphasize Importance of Data Collection and Applications for Prioritizing Investments
  - Traffic Counts
- In Oregon, This is Accomplished on Major Facilities Through Permanent or Short-Term Counts
  - For Minor or Local Facilities, These Counts Can be Expensive and Labor Intensive
- Short-Term Traffic Counts That Cover All Links is Not Practical Nor Economically Viable
- There is a Need to Develop Predictive Methods to Estimate Traffic Volumes on Minor and/or Local Roadway Segments
  - **The Premise for Prediction is to Classify Local Roadway Segments Into Specific Traffic Volume Ranges**

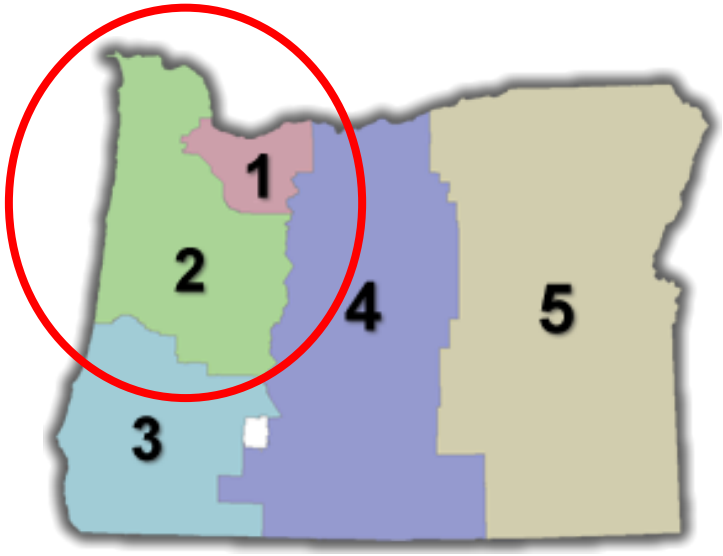
Motivation

- **Why Predict Volumes Ranges Over Continuous Values of Traffic Volume?**
- Classify Local Facilities Based on Specific Traffic Volume Ranges Based on a Variety of Factors
- Prediction of Traffic Volume Ranges Was More Intuitive and Informative to ODOT Practitioners
- Predicting Traffic Volume as a Continuous Variable Through Linear-Based Approaches Has Room to Be Improved
    - Results in High Errors and Low Prediction Accuracy
- Multiclassification and Binary Classification Considered

Data

- Average Daily Traffic (ADT) Counts for Local Roads Was Obtained Through an ODOT Database as Part of SPR-804
- At Time of Study, Most Recent Traffic Counts Were for 2016
- Of Particular Interest Were ADT Counts in ODOT Region 2

Clatsop County	Lincoln County
Columbia County	Benton County
Tillamook County	Linn County
Yamhill County	Lane County
Polk County	Washington County (Western Portion)
Marion County	



Data

- 426 Data Samples Were Collected

Variable	Example
Lane-Use Characteristics	• Primary ADT Generator Adjacent to Roadway • Primary Dominant Land-Use Adjacent to Roadway • Presence of Single-House Residences • Sub-Region (e.g., Coastal, Mountain, Valley, and MPO)
Type of Median	• Undivided • One-Way • Single Lane
Intersection Characteristics	• Presence of Right-Turn Lane • Presence of Left-Turn Lane • Presence of Traffic Signal
Roadway Characteristics	• Pavement Type • Presence of Horizontal Pavement Markings • Presence of Shoulder

Data

- 426 Data Samples Were Collected

Variable	Example
Roadway Characteristics (Cont.)	• Presence of Crosswalk • Presence of Sidewalk • Presence of Bike Lane • Presence of Bus Stop • Presence of Parking Lot • Presence of Calming Device
Signage Characteristics	• Presence of Stop Sign on the Crossroad • Presence of Stop Sign on Main Corridor • Presence of Sign, Other Than Stop Sign on Main Corridor
Demographic Characteristics	• Population Within 1,000 ft. of Main Corridor • Employment Within 1,000 ft. of Main Corridor

Prediction Methodology

- Multiclassification
- Traffic Volumes Were Classified Into Four Distinct Categories Based on Interest From ODOT
  - These Ranges Were the Dependent Variable in the Prediction Models

Summary of Traffic Volume Classifications				
	ADT Range			
ADT	0 to 250	250 to 500	500 to 1,000	Greater Than 1,000
Training	112	92	61	33
Test	49	38	27	14

Prediction Methodology

- Binary Classification
- Binary Classification Based on Interest From ODOT
  - These Ranges Were the Dependent Variable in the Prediction Models

Summary of Binary Traffic Volume Classification				
Variable	Mean	Std. Dev.	Frequency in Training Dataset	Frequency in Testing Dataset
1 if ADT Between 0 and 250, 0 Otherwise	0.38	0.49	112	49

Prediction Methodology

- Multiclassification
  - Parametric: Ordered Probit Model
  - Nonparametric: Multiclassification Support Vector Machine (SVM) Learning
    - One-Against-One Approach
- Binary Classification
  - Parametric: Binary Ordered Probit Model
  - Nonparametric: SVM
- Pros/Cons
  - Parametric: Interpretability of Results/Lower Prediction Accuracy
  - Nonparametric: Prediction Accuracy/Further Analysis Required to Interpret Important Predictors

Prediction Methodology

- Prediction Evaluation

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Sensitivty} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

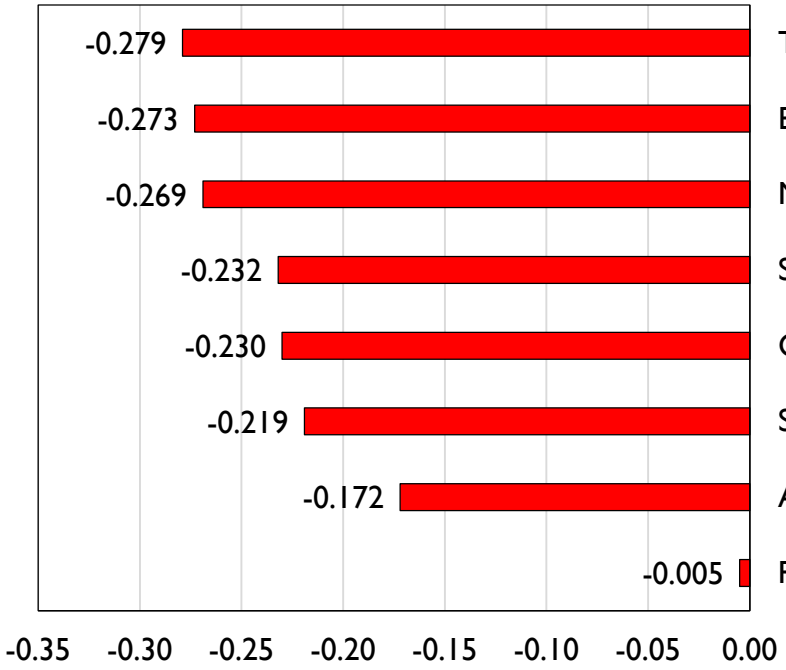
TP (True Positive)	**Samples Predicted as Positive and Are Truly Positive**
FP (False Positive)	**Samples Predicted as Negative, But are Truly Positive**
TN (True Negative)	**Samples Predicted as Negative and Are Truly Negative**
FN (False Negative)	**Samples Predicted as Negative, But Are Truly Positive**

Results (Multiclassification)

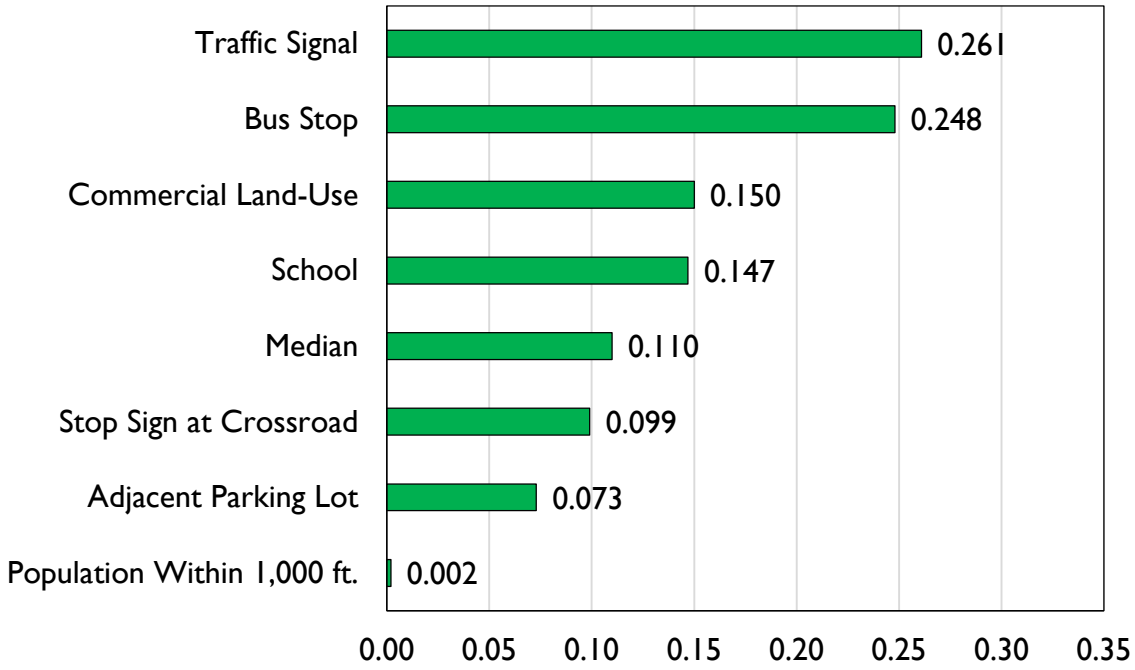
- Variable Effects Identified in the Ordered Probit Model

Ordered Probit Marginal Effects				
Variable	ADT Range			
	0 to 250	250 to 500	500 to 1,000	Greater Than 1,000
Generators				
1 if School, 0 Otherwise	-0.219	-0.055	0.128	0.147
Roadway Characteristics				
1 if Median Present, 0 Otherwise	-0.269	0.008	0.151	0.110
1 if Bus Stop Present, 0 Otherwise	-0.273	-0.119	0.145	0.248
1 if There is Adjacent Parking Lot, 0 Otherwise	-0.172	0.180	0.099	0.073
Intersection Characteristics				
1 if Traffic Signal Present, 0 Otherwise	-0.279	-0.127	0.145	0.261
Signage				
1 if Stop Sign at Crossroad, 0 Otherwise	-0.232	0.001	0.133	0.099
Land-Use				
1 if Commercial, 0 Otherwise	-0.230	-0.054	0.134	0.150
Demographic Characteristics				
Population Within 1,000 ft	0.005	0.000	0.003	0.002

Results (Multiclassification)



Negative Marginal Effects
(ADT Range: 0 to 250)



Positive Marginal Effects
(ADT Range: > 1,000)

Maseeh College of Engineering and Computer Science
PORTLAND STATE UNIVERSITY

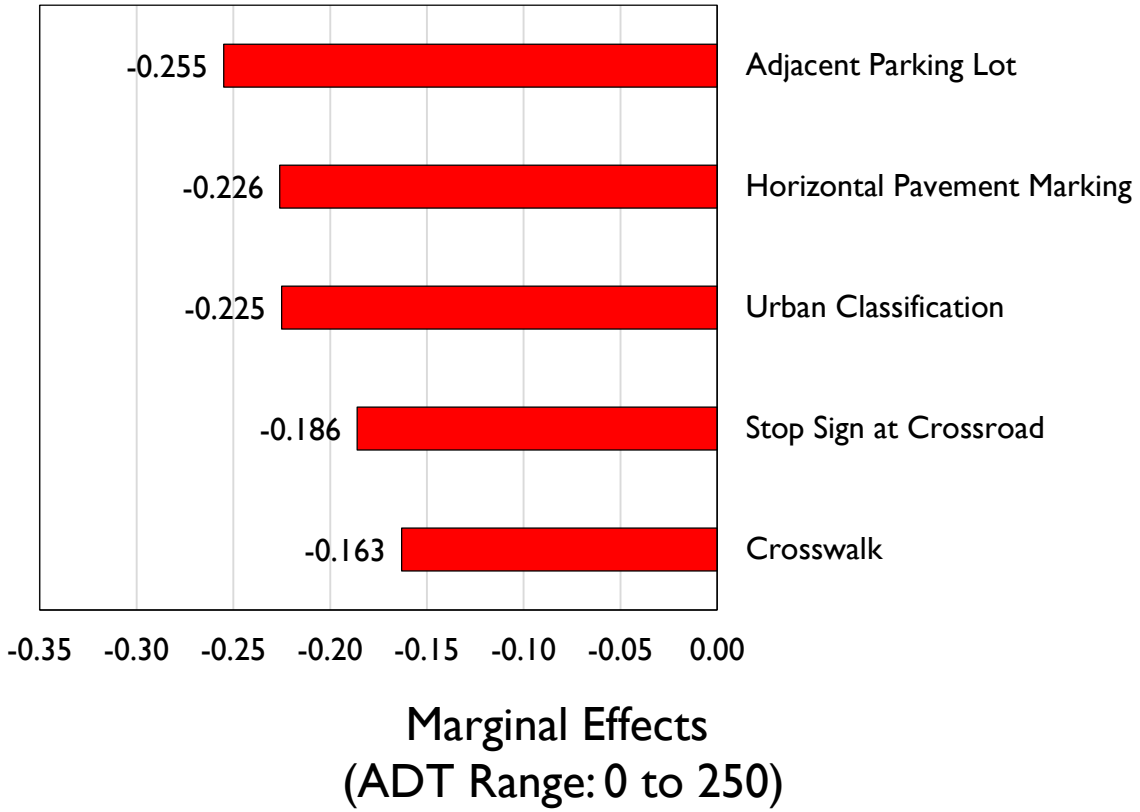
TTP Lab

Results (Binary Classification)

- Significant Variables Identified in the Binary Probit Model

Binary Probit Model Specifications (Training Dataset)					
Variable	Coefficient	Std. Error	*t*-statistic	*p*-value	Marginal Effects
Constant	0.95	0.21	4.47	0.00	
Functional Classification					
1 if Urban, 0 Otherwise	-0.71	0.20	-3.52	0.00	-0.225
Roadway Characteristics					
1 if There is Adjacent Parking Lot, 0 Otherwise	-0.84	0.20	-4.08	0.00	-0.255
1 if Horizontal Pavement Marking, 0 Otherwise	-0.78	0.22	-3.59	0.00	-0.226
1 if Presence of Crosswalk, 0 Otherwise	-0.56	0.28	-1.99	0.05	-0.163
Signage					
1 if Stop Sign at Crossroad, 0 Otherwise	-0.59	0.17	-3.40	0.00	-0.186
McFadden Pseudo R-Squared	0.18				

Results (Binary Classification)



Results (Prediction Performance)

Comparison of Prediction Metrics for Multiclass and Binary Class Models			
Multiclassification			
Model	Accuracy	Sensitivity	Specificity
Ordered Probit	63.4%	45.2%	70.7%
SVM (Radial)	62.7%	35.4%	71.0%
SVM (Polynomial)	63.2%	40.3%	70.7%
Binary Classification			
Model	Accuracy	Sensitivity	Specificity
Binary Probit	60.9%	68.3%	48.9%
SVM (Radial)	71.1%	87.3%	44.9%
SVM (Polynomial)	71.1%	87.3%	44.9%

Results (Prediction Performance)

Comparison of Prediction Metrics for Multiclass and Binary Class Models			
Multiclassification			
Model	Accuracy	Sensitivity	Specificity
Ordered Probit	**63.4%**	**45.2%**	70.7%
SVM (Radial)	62.7%	35.4%	**71.0%**
SVM (Polynomial)	63.2%	40.3%	70.7%

Results (Prediction Performance)

Comparison of Prediction Metrics for Multiclass and Binary Class Models			
Binary Classification			
Model	Accuracy	Sensitivity	Specificity
Binary Probit	60.9%	68.3%	**48.9%**
SVM (Radial)	**71.1%**	**87.3%**	44.9%
SVM (Polynomial)	**71.1%**	**87.3%**	44.9%

Summary

- All Significant Variables Decreased The Probability of Observing Traffic Volume in the 0 to 250 Range
- For Multiclassification, the Parametric Model (Ordered Probit) Performed Better in Overall Accuracy and Sensitivity
  - SVM Performed Better in Specificity
- For Binary Classification, the Nonparametric Model (SVM) Performed Better in Overall Accuracy and Sensitivity
  - Probit Model Performed Better in Specificity
- In This Context, the Parametric Method is Preferred

Moving Forward

- Apply Comparisons to a Larger Sample Size
  - Attempt Varying Test/Training Proportions
  - Include an Advanced Linear-Based Approach
- Utilize Alternate Parametric Approach to Interpret "Middle" ADT Ranges (Multinomial Logit)
- Alternate Machine Learning Approaches for Multiclass Problem
  - Neural Networks
- Comparison Using Alternate Data Sources
  - Cellphone Probes

Acknowledgements

- Tony Knudson
- TAC for SPR-804
  - Nick Fortey
  - Christina McDaniel-Watson
  - Zahidul Siddique
  - Jennifer Campbell
  - Don Crownover
- Data Collection
  - Mostafa Khademi
  - Sara Urbina

Results (Multiclassification)

- Significant Variables Identified in the Ordered Probit Model

Ordered Probit Model Specifications (Training Dataset)				
Variable	Coefficient	Std. Error	*t*-statistic	*p*-value
Constant	-0.73	0.15	-4.77	0.00
Generators				
1 if School, 0 Otherwise	0.72	0.30	2.42	0.02
Roadway Characteristics				
1 if Median Present, 0 Otherwise	0.76	0.16	4.78	0.00
1 if Bus Stop Present, 0 Otherwise	1.03	0.46	2.26	0.02
1 if There is Adjacent Parking Lot, 0 Otherwise	0.49	0.17	2.95	0.00
Intersection Characteristics				
1 if Traffic Signal Present, 0 Otherwise	1.07	0.45	2.36	0.02
Signage				
1 if Stop Sign at Crossroad, 0 Otherwise	0.66	0.13	4.92	0.00
Land-Use				
1 if Commercial, 0 Otherwise	0.75	0.25	3.00	0.00
Demographic Characteristics				
Population Within 1,000 ft.	0.01	0.00	3.89	0.00