Exh. DLT-9
Docket UG-240008
Witness: Daniel L. Tillis

BEFORE THE
WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION, Complainant, v. CASCADE NATURAL GAS CORPORATION, Respondent.	DOCKET UG-240008

CASCADE NATURAL GAS CORPORATION

EIGHTH EXHIBIT TO THE
DIRECT TESTIMONY OF DANIEL L. TILLIS

March 29, 2024

Cascade Washington:
Low Income Program Participation Propensity Analysis

Cascade Natural Gas
Energy Assistance Advisory Group: July 19, 2023

Mark E. Thompson
Forefront Economics Inc

Review Meeting Agenda

- Introductions
- Project Overview
  - Objectives & Approach
  - Results
  - Deliverables and Uses
- Wrap-up and Next Steps Discussion

Summary of Project Objectives and Approach

- Project Objectives
  - Develop premise level residential database for better understanding characteristics of low-income program participants
  - Use the enhanced residential database to identify residential prospects for low-income programs
  - Use results to target best prospects as a means of cost effectively driving low-income program participation rates higher
- Approach
  - Combine Cascade Natural Gas (Cascade) customer information with secondary data
  - Profile and contrast low-income participants with other residential customers
  - Model low-income program participation as a function of customer attributes
  - Apply model to “score” customers for program targeting

Developing the Data

- Combine Cascade CIS Data with Secondary Data
- CIS Data
  - Billing (usage, dollars)
  - L-I program participation
  - Payment data
    - Late payments
    - Arrearage balances
- Secondary Data – Household Level
  - Household income
  - Premise size, age and market value
- Secondary Data – Census Tract
  - Energy burden data
  - Concentration (percentage) of low-income households in Census Tract

<u>CIS Data</u>

- Energy Use
- Low-Income Program Participation
- Late Payment and Arrearage History

<u>Secondary Data</u>

- Premise Level
  - HH Income
  - Size of Home
  - Age of Home
- Census Tract
  - Energy Burden

Types of Cascade Data

- Premise Records
  - Service address and related location information
  - Serves as the basic unit of analysis
- Energy Bill Assistance Program History
  - The basis of dependent variable in propensity models
- Billing Records
  - Annual therms and dollars billed
- Payment history
  - Number of late payments
  - Arrearage balance
  - Non-payment related disconnects

Energy Bill Assistance Program Data (Cascade)

- Energy bill assistance program participation history obtained for the past five years.
- Participant counts jumped over the last two years from special pandemic relief assistance (Big Heart).
- This history is the basis for establishing dependent variable in low-income program participation propensity models.

--- Washington ---					
Program Name	**2018**	**2019**	**2020**	**2021**	**2022**
LIHEAP Washington	1,375	1,355	1,229	1,549	1,564
WA Big Heart Grant	0	0	0	6,690	3,697
WA Energy Assistance Fund (WEAF)	2,206	2,146	2,639	2,326	2,186
Winter Help	478	371	878	419	397
Total	**4,059**	**3,872**	**4,746**	**10,984**	**7,844**

- Geocode Cascade and Secondary Addresses
  - Standardizes addresses for improved match rates
  - Appends Census Tract numbers (2010 and 2020)
  - Appends latitude & longitude for GIS applications
- Match Cascade Records to Secondary Household Data
  - Run data enhancement routines, data cleaning, and reduction:
    - Calculate therms per square foot
    - Combine common fields (e.g. address fields)
- Match Cascade Records to US DOE Energy Burden Estimates
  - Source: Low-Income Energy Assistance Data (LEAD)
  - Census Tract level data
- Result
  - An information rich and site-specific data set for residential customers

Geocode and Match Results

- High percentage of records geocoded in both datasets
  - Indicates street addresses

--- Washington ---			
	Records	Geocoded	Percent Geocoded
CNG Premises	213,407	200,555	94%
Household Data (Secondary)	371,783	364,165	98%

  - Nearly 104,000 premises (53% of geocoded premises) matched to household data
  - Household data restricted due to business rules designed to prevent mismatch between occupant and attribute data
    - Example: If site record shows occupant moved and record has not been updated for a new occupant, the record was omitted.
  - Plenty of matched premises for statistical modeling

Comparing Attributes

- Combined data allow for comparison of premises receiving energy bill assistance (EBA) with those that did not
  - Have lower gas bills
  - Higher Cascade account turnover
  - More late payments and much higher arrearage balances
  - Lower household incomes
  - Live in smaller, older, less valuable homes
  - Use significantly more gas per square foot.

--- Washington ---	EBA Participant	
	No	Yes
Cascade Data	**N=183,120**	**N=10,903**
Annual bill (2022)	$871	$817
Annual therm usage (2022)	619	576
Account turnover at premise	9%	16%
Avg monthly arrearage balance (2018-2022)	$6	$55
Avg annual late payments (2018-2022)	0.5	1.9
Secondary Household Data	**N=115,232**	**N=4,161**
Household income	$112,068	$77,695
Age of home (years)	42.21	55.19
Market value of home	$404,062	$284,783
Size of home (square feet)	2,455.62	1,962.61
Therms/Sq ft (CNG & household data)	0.289	0.354
Low-income Energy Assistance Data (LEAD)	**N=201,405**	**N=11,952**
Mean household income	$89,617	$80,030
Pct of gas heated homes < 150% FPL	12%	16%
Energy burden - Total	1.8%	2.1%
Energy burden - Natural Gas	0.7%	0.8%
Energy burden - Electric	1.1%	1.3%

Propensity Models

- Statistical models used to explain and predict the probability of a given event or outcome
  - Models relate the "outcome" (e.g. participation in low-income programs) to explanatory variables ("drivers")
- Propensity models used extensively in:
  - Medical Research
    - What is the probability that a patient will develop lung cancer?
    - Driver variables: years smoking, years since last cigarette, sex, age, income
  - Social Research
    - What is the probability a student will graduate from college?
    - Driver variables: income, parents education, parents occupation, SAT/ACT scores
  - Economic Research
    - What is the probability a consumer will purchase a product or service?
    - Driver variables: price, price of competing and complimentary products, income

Propensity Models (cont'd)

- Results of logistic regression models are evaluated using many criteria
  - Experience table: number of "true" positives vs. false positives/negatives
  - "Lift" (preferred for model selection)
  - Actual experience using model (experience is the best teacher)
- Model results used to "score" other data beyond the sample used to estimate the model
  - Score: the estimated probability of event (e.g., low-income program participation) for a single observation (e.g., Cascade premise)
  - Sorting by estimated probability shows relative probability
    - Decile assignments based on sort ordered
    - More meaningful than absolute probability
  - A model with a poor experience table may still provide useful relative probability estimates

Sample of Premises for Propensity Modeling

- Select random sample of 5,000 premises that have been occupied by a low-income bill assistance program participant within the last 5 years (2018-2022)
- Select random sample 5,000 premises that have not been occupied by a low-income bill assistance program participant within the last 5 years (2018-2022)

--- Washington ---	
Program Participation Status	**Premises**
Billing Assistance Program Participants	5,000
Non-Participants	5,000
Total Number of Premises in Sample	**10,000**

Model Estimation

$$prob = (b_1 \bullet X_1) + (b_2 \bullet X_2) + \ldots$$

Propensity Model Results

- Models using all Cascade variables, Census variables and variables from purchased household data
- Models using only variables from Cascade and Census

- Statistically significant and correctly signed
- Overall model performance deteriorates significantly if variable is removed

- Area under ROC varies from 0.5 to 1.0
  - No set rules but values between 0.8 and 0.9 are generally considered excellent
- ROC curves show
  - both models have excellent prediction accuracy
  - neither model stands out from the other as a better predictor.

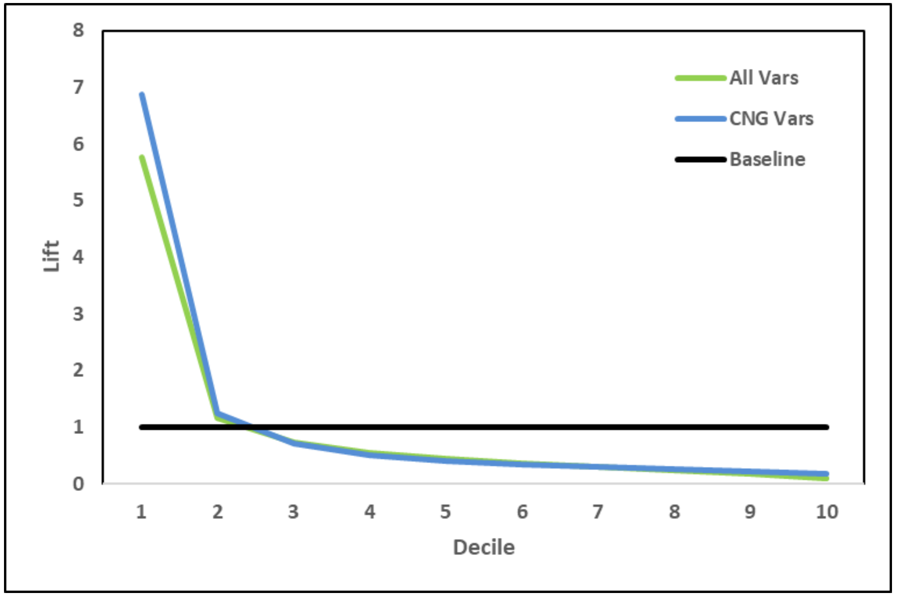
--- Washington ---		
Model	**Premises**	**ROC**
Best All Variables Model	4,558	0.863
Best CNG and Census Variables Model	9,136	0.854

Variables in Final Models

--- Washington ---						
Variable	Best Model - All Data Sources			Best Model - CNG and Census Only		
	Impact on Probability	Statistical Significance	Influence	Impact on Probability	Statistical Significance	Influence
Household Income	-	High	High			
Market Value of Home	-	High	High			
Age of Home	+	Moderate	High			
Therms per Square Foot	+	High	Moderate			
Average Monthly Arrearage	+	High	High	+	High	High
Late Payments	+	High	High	+	High	High
Late Payments - Moratorium	+	High	High	+	High	High
Account turnover	+	High	High	+	High	High
Energy Burden - Total	+	Moderate	Weak	+	High	Weak
Percent homes < 150 % FPL	+	High	Weak	+	High	Weak
Annual CNG Bill - Dollars				-	High	High
Premise Type - Multifamily				+	High	High
Premise Type - Manuf. Home				+	High	Moderate
Non-Payment Disconnect				+	Weak	High

Model Results - Lift

- Lift – how well a model identifies high likelihood prospects relative to average participation rate
  - Allows comparison of models ability to identify likely program prospects
- Lift shows the ratio of model predicted probability to average probability
  - Higher lift means better low-income program prospects
  - Lift of 1.0 means model no better than average of current participation
- Results in chart sorted from most likely to participate in low income programs (decile 1) to least likely (decile 10)
- In terms of ability to predict program participation, the best models from each category are not meaningfully different.
- Both provide excellent in first decile (10% of the premises predicted by model).
- Model using only Cascade and Census variables (CNG Vars)
  - Has better coverage (almost all Cascade premises can be predicted with model)
  - Does a slightly better job of identifying premises with the highest probability of program participation (Decile 1)



Customer Profiles by Decile

- Shows how decile groups compare across variables in the analysis
  - Use to contrast top prospects for program participation (decile 1 and decile 2) to other customers.
  - Top three deciles shown separately and remaining deciles grouped to better illustrate differences in analysis variables between groups of customers.

--- Washington ---					
Variable	Variable Means by Decile				
	1	2	3	4-5-6	7-8-9-10
Household Income	$81,193	$83,378	$88,886	$102,519	$131,410
Market Value of Home	$277,662	$300,460	$324,157	$383,848	$465,737
Age of Home	58	56	54	42	35
Home Square Footage	1,981	2,019	2,100	2,289	2,765
Therms per Square Foot	0.331	0.314	0.304	0.281	0.284
Energy Burden - Total	2.2%	2.2%	2.1%	1.8%	1.6%
Percent homes < 150 % FPL	18.3%	17.5%	15.2%	12.5%	8.6%
Average Monthly Arrearage	$64	$11	$5	$2	$0
Late Payments	2.7	1.3	0.7	0.3	0.1
Late Payments - Moratorium	46	10	5	2	1
Account turnover	24%	23%	18%	11%	3%
Non-Payment Disconnect	16%	4%	1%	0%	0%
Annual CNG Bill - Dollars	$780	$752	$772	$800	$993

Scoring All Customer Premises

- Scoring refers to using model to predict low-income program participation probability for Cascade customers
- Final model based on Cascade and Census variables used to “score” all premises

Interpreting Results

- Uses (relative probability)
  - Identifies premises with high probability of program participation relative to other premises
  - Identifies Census Tracts with high number of premises with high probability of program participation relative to other premises
- Limitations (absolute probability)
  - Can not use probability estimates as absolute estimates. Examples of absolute probability uses include:
    - Which premises will participate in low-income programs next year
    - How many premises will participate in low-income programs next year

Application of Results

- Drive program participation higher through targeted outreach
  - Example: Contact top 20%-25% of prospects
  - Example: Neighborhood events (top 20 Census Tracts shown in table)
- Other Possible Uses
  - Targeting of other customer service (e.g. DSM programs and services)

Service County	Census Tract (2020)	Number of Decile 1 and 2 Premises
Yakima	000700	917
Yakima	001202	771
Mason	960900	647
Kitsap	080200	643
Yakima	002003	500
Whatcom	000700	496
Whatcom	001000	490
Franklin	020504	458
Whatcom	010411	427
Yakima	000500	423
Kitsap	080600	402
Yakima	000600	401
Franklin	020605	400
Walla Walla	920600	391
Franklin	020606	382
Skagit	952401	377
Yakima	940006	375
Whatcom	001203	374
Adams	950500	370
Skagit	952500	355

Wrap-Up and Discussion of Next Steps

Deliverables

- Document Files
  - PowerPoint documenting approach and findings
  - Technical notes
    - Variable list, labels and coded values
- Programs and Tables
  - Excel workbooks with premise data and propensity model scores
    - Excel workbook with separate sheets for decile 1 and decile 2 prospects: "****_Final_Decile_1and2.xlsx"
    - Excel workbook with consolidated county and Cascade residential data: "****_Final_All.xlsx"
  - Score Code
    - Provides for easy updating of scores as underlying data changes
      - Consult with us before using to avoid misapplication