



*“Working together to  
fund Washington’s future”*

# Household Tax Burden Model

Kris Bitney, February 2020

Research & Fiscal Analysis

# Model Overview

**Goal: Estimate average household tax burden by income and region to assess the fairness of Washington's tax system for households**

- Multiple excise tax categories
- Property tax
- Estimate tax burdens under current law or alternative policies

## **Two Models**

- Excise Tax Model
- Property Tax Model

# Major Household Taxes in Washington

Current Washington Taxes	Possible Alternative Taxes
Retail sales tax	Extend retail sales tax to services
Alcohol	Repeal sales tax exemption on food
Cigarettes and other tobacco products	Capital gains tax
Insurance premiums	Personal income tax
Gasoline	
Public utilities	
Property tax	
Real estate excise tax	

## Tax Burden on Households (2014)

### Major State and Local Taxes

**Current Law**

**Household Income**

	\$0	\$15,000	\$25,000	\$35,000	\$45,000	\$55,000	\$70,000	\$85,000	\$105,000	over
	\$15,000	\$25,000	\$35,000	\$45,000	\$55,000	\$70,000	\$85,000	\$105,000	\$140,000	\$140,000+

Retail Sales Tax	\$905	\$1,170	\$1,453	\$1,690	\$1,988	\$2,340	\$2,729	\$3,217	\$3,832	\$5,908
Alcoholic Beverages Taxes	\$55	\$63	\$78	\$93	\$98	\$119	\$125	\$143	\$167	\$239
Cigarette & Tobacco Taxes	\$156	\$177	\$194	\$197	\$204	\$211	\$198	\$193	\$166	\$119
Insurance Premiums Tax	\$22	\$36	\$47	\$55	\$67	\$74	\$83	\$94	\$107	\$138
Gasoline Tax	\$132	\$184	\$233	\$273	\$311	\$349	\$392	\$427	\$467	\$498
Public Utility Taxes	\$113	\$140	\$157	\$171	\$186	\$199	\$217	\$236	\$258	\$324
Property Tax	\$804	\$1,089	\$1,307	\$1,524	\$1,896	\$2,230	\$2,678	\$3,102	\$3,824	\$6,130
<b>Total Tax</b>	<b>\$2,187</b>	<b>\$2,859</b>	<b>\$3,469</b>	<b>\$4,003</b>	<b>\$4,749</b>	<b>\$5,522</b>	<b>\$6,422</b>	<b>\$7,413</b>	<b>\$8,821</b>	<b>\$13,354</b>
<b>Tax as % of Income</b>	<b>26.5%</b>	<b>14.2%</b>	<b>11.5%</b>	<b>10.0%</b>	<b>9.5%</b>	<b>8.9%</b>	<b>8.3%</b>	<b>7.8%</b>	<b>7.3%</b>	<b>5.9%</b>

# Data Sources

## **IRS Individual Income Tax Data**

- All federal tax returns for Washingtonians in 2017 calendar year

## **Consumer Expenditure Survey (CES)**

- Produced by federal Bureau of Labor Statistics
- Nationally representative survey of American households on consumer expenditures, income, and demographic characteristics

## **County Property Tax Rolls**

- Provided to Department of Revenue by all Washington counties
- Has property values of all land parcels in Washington

## **American Community Survey (ACS)**

- Produced by US Census Bureau
- Nationally representative survey of American households on many topics

# Excise Tax Model

1. Prepare data
2. Develop statistical model to estimate each of 575 expenditure categories in CES data
3. Use statistical model to assign 575 expenditure estimates to each Washington household
4. Estimate tax burden associated with each expenditure for each Washington household
5. Aggregate tax burdens by income decile and geographic region

# IRS Individual Income Tax Data

## **Data point for every federal tax return filed by Washingtonians**

- Information about income, family composition, and place of residence

## **Generally representative of Washington**

- Some low-income Washingtonians are not required to file
- Number of dependents limited to four per tax return

## **Unit of observation is tax return (“tax unit”)**

- Mean number of persons in tax unit is 2.03
- Mean household size in ACS is 2.5

## **We collapse tax returns on address to make households**

- Mean number of persons in “household” is 2.64
- Assumes anyone living at the same address is part of household

# Representativeness of IRS Individual Income Tax Data

Data	Population (Persons)	Number of Households	Mean Family Size	Total Wage Income	Mean Household Wage Income
ACS household 2017	7,090,000	3,100,000	2.5	\$202b	\$71,000
OFM 2017	7,310,000	3,080,000	-	-	-
IRS Individual Income Tax data (collapsed) 2017	6,760,000	2,560,000	2.64	\$202b	\$79,000
IRS Individual Income Tax Data (tax units) 2017	6,760,000	3,340,000	2.03	\$202b	\$61,000



# Consumer Expenditure Survey (CES)

## Two nationally representative surveys of American households

- Two distinct samples

### Interview survey

- Respondents contacted each quarter for four quarters
- Asked about consumption during prior three months
- Asked about income during prior year

### Diary survey

- Respondents contacted each week for two weeks
- Asked about consumption during prior week
- Asked about income during prior year

# Consumer Expenditure Survey: Missing Data

## **Some households participate in fewer surveys than BLS intended**

- i.e. Less than 4 for Interview survey, or less than 2 for Diary survey

## **2020 model aggregates quarterly/weekly responses to annual estimates**

- Within-household mean imputation for missing quarters/weeks
- Possibly multiple imputation

## **2002 model matched CES Interview quarterly responses to ACS annual interviews**

- E.g., one WA household can match to multiple different CES household-quarters
- Pro: It's okay if households lack four quarters of data
- Con: doesn't account for seasonal/annual purchasing patterns
- Con: quarterly expenditures less correlated with annual income

# Excise Tax Microsimulation Model

1. Prepare data
2. **Develop statistical model to estimate each of 575 expenditure categories in CES data**
3. **Use statistical model to assign 575 expenditure estimates to each Washington household**
4. Estimate tax burden associated with each expenditure for each Washington household
5. Aggregate tax burdens by income decile and geographic region

# Statistical Model Candidates

- Baseline model (as used in 2016 or 2002)
- K-Nearest Neighbors (KNN)
- Ordinary Least Squares (OLS)
- Lasso with polynomial features
- Gradient Boosted regression ensemble
- Ensemble of Regressor Chains (ERC) with Lasso base estimator

# Statistical Model Selection

## **Empirical performance measurement**

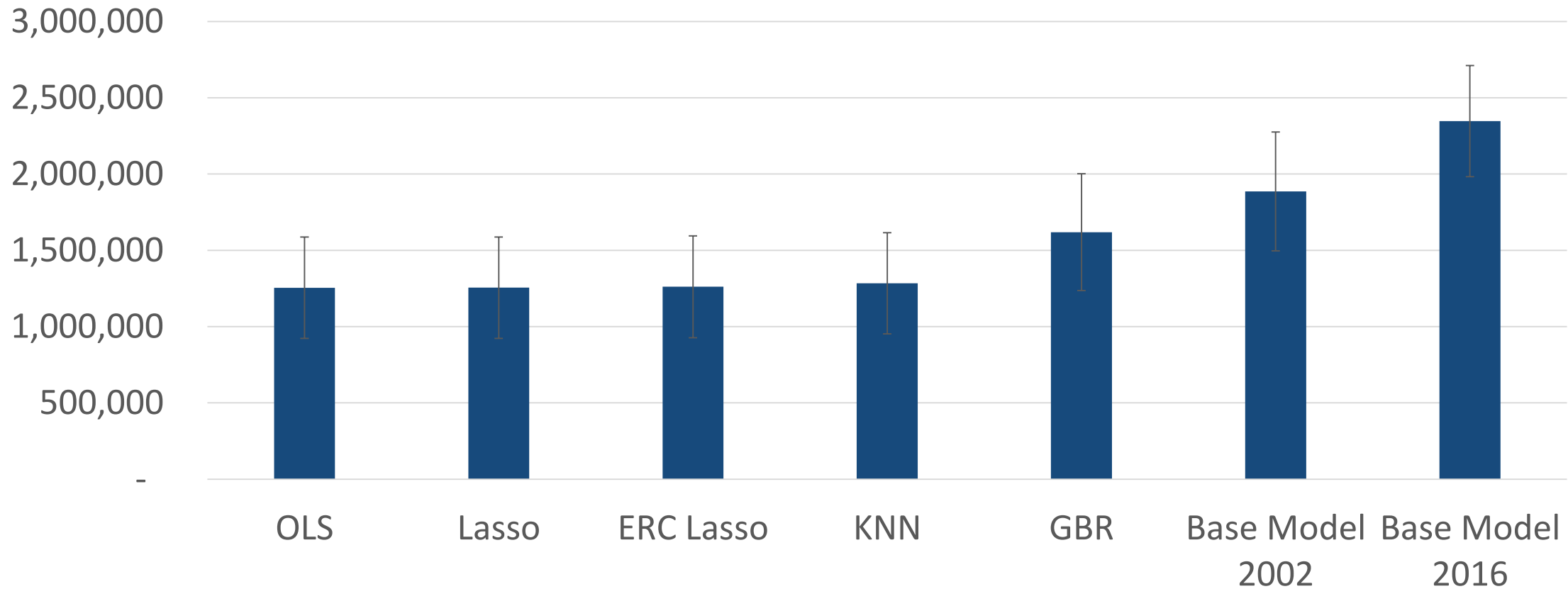
- How well will the statistical model estimate expenditures for households in the IRS Individual Income Tax Data?
- Statistical models perform better on training samples than other samples

## **K-Fold Cross-Validation**

- Measures out-of-sample performance (“expected prediction error”)
- Can compare performance of very different statistical models

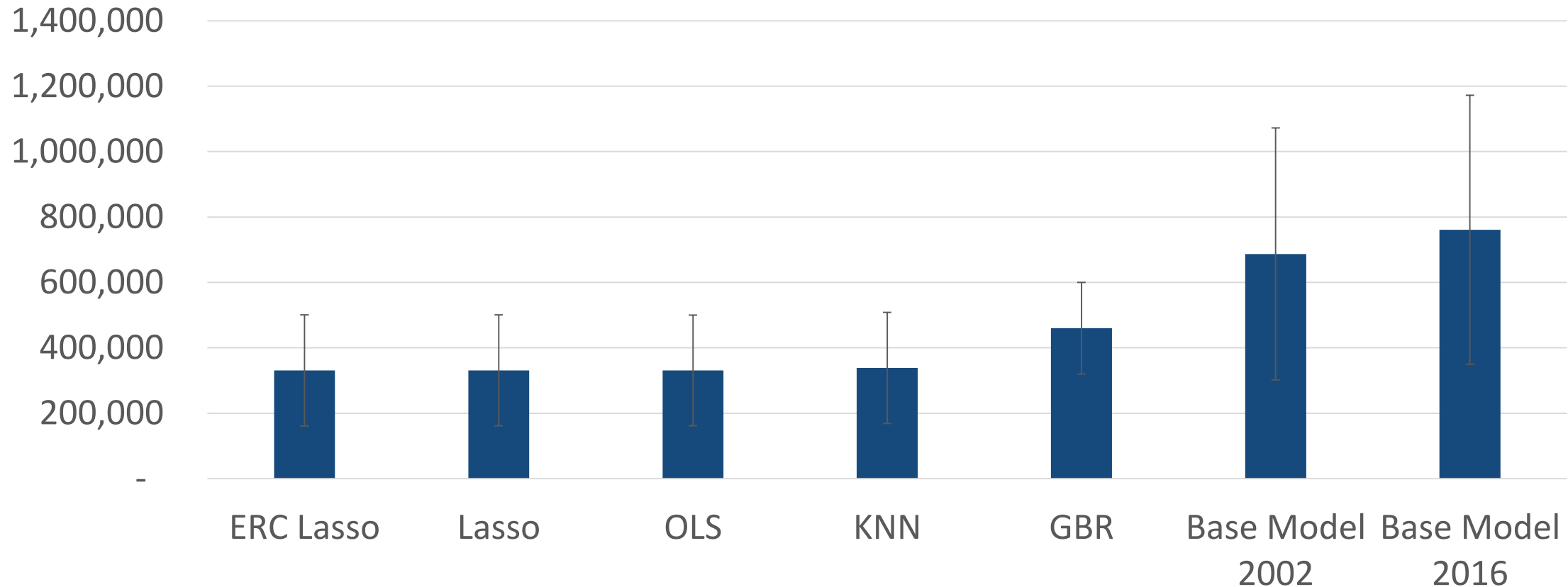
# Cross-Validation Results for CES Interview

Mean Square Error by Model Type (lower is better)



# Cross-Validation Results for CES Diary

Mean Square Error by Model Type (lower is better)



# Excise Tax Microsimulation Model

1. Prepare data
2. Develop statistical model to estimate each of 575 expenditure categories in CES data
3. Use statistical model to assign 575 expenditure estimates to each Washington household
4. **Estimate tax burden associated with each expenditure for each Washington household**
5. **Aggregate tax burdens by income decile and geographic region**



# Adjustments and Misreporting of Consumption

## **In survey responses, people tend to:**

- Overestimate healthy and socially desirable consumption
- Underestimate unhealthy and socially stigmatized consumption

## **How can we adjust consumption?**

- For misreporting
- For underrepresented populations
- To match known tax revenue totals

# Forecasting Tax Estimates

## **Model is based on Calendar Year 2017**

- We are directed to produce estimates for 2017-2019 Biennium
- We forecast CY2017 estimates to future years

## **Economic & Revenue Forecast Council**

- Publishes revenue estimates for major Washington taxes

## **IHS Markit**

- Publishes national consumption estimates for many types goods
- Used to forecast consumption of goods not currently taxed in Washington, and gasoline

# Regional Excise Tax Estimates

## **Geocoding IRS Individual Income Tax Data**

- Street addresses reported on tax returns

## **Geographic aggregation levels**

- State legislative and congressional district
- County
- ACS Region

## **Limitations**

- IRS disclosure requirements
- Larger samples produce better estimates

# Assumptions about Economic Behavior

## **Results aren't influenced by between-state differences in cost of living or economic policy**

- e.g., different rent prices, different tobacco taxes

## **Changes in taxation do not lead to changes in consumer behavior**

- e.g., introducing a tax on food items changing food purchase patterns

## **Tax burdens imposed on households are paid by households (taxes not shifted)**

- e.g., industry responding with price change

# Property Tax Model

1. Join IRS Individual Income Tax Data with County Property Tax Rolls data using physical and mailing addresses
2. Randomly assign home ownership based on conditional probability distribution
  - a. We often do not know if residents own their property
  - b. We estimate distribution of home ownership, conditional on income and geographic region
  - c. Data source: American Community Survey
3. Estimate property tax burden for each Washington household
4. Aggregate tax burdens by income decile and geographic region

# County Property Tax Rolls Data

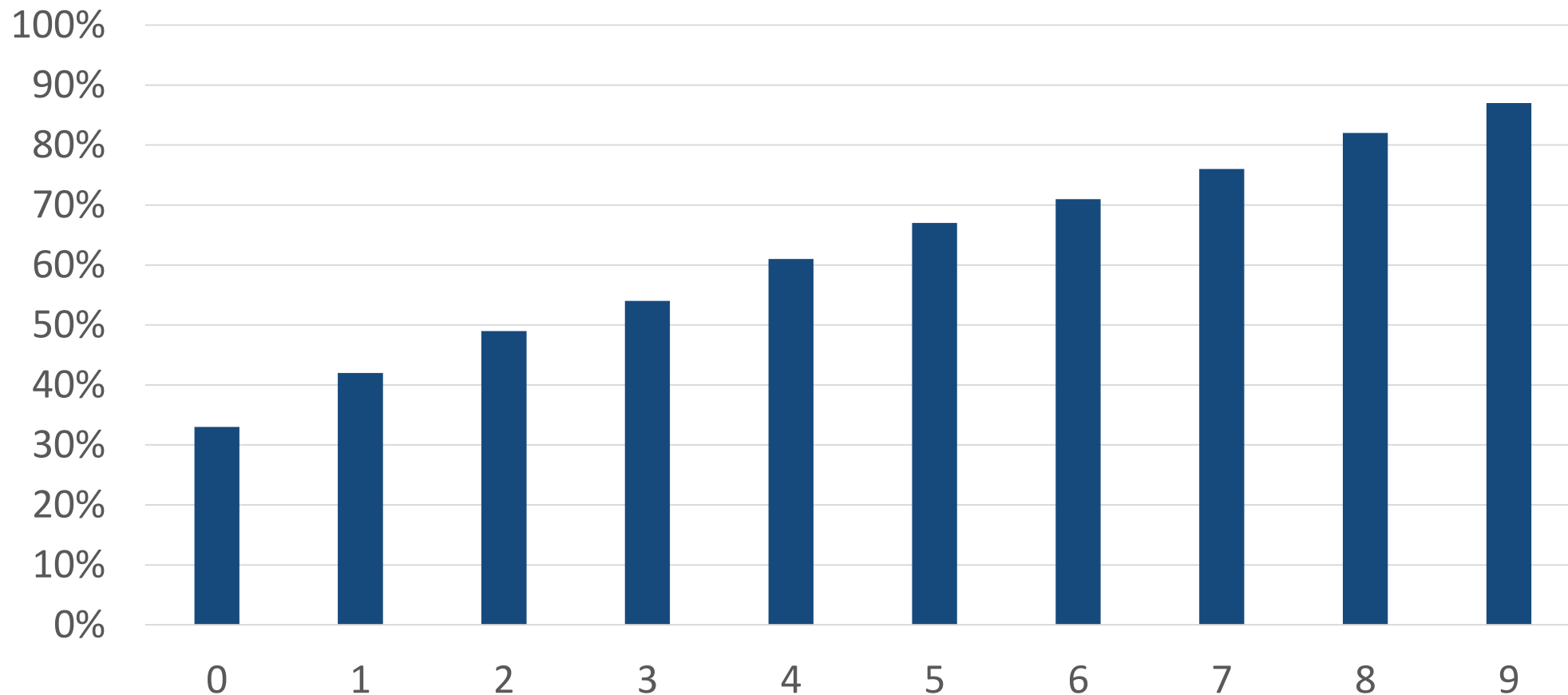
## **Data point for every land parcel in Washington**

- Information about property values, buildings, locations

## **We focus on single-family units**

- Includes residential condominiums and vacation homes
- Single-family units might be owned by households
- Multi-family units are presumed to be rentals

## Statewide home ownership rate by (net) income decile



# Questions?





# Questions for Technical Advisory Group

1. Survey respondents tend to overestimate their healthy consumption habits and underestimate their consumption habits that are unhealthy or socially undesirable. **How can we best measure misreporting and implement adjustments?**
2. In the 2002 model, we adjusted results so that aggregated revenue estimates matched known totals estimated by internal sources. This is an important adjustment. **How can we best make these adjustments?**
3. Our model assumes the population of federal income tax filers in Washington is representative of Washington's population of households. The population count in the IRS Individual Income Tax Data is about 8% lower than the April 1 Official Population Estimates reported by the Office of Financial Management. Likewise, the average income in the IRS Individual Income Tax Data is greater than that reported in the American Community Survey, suggesting the IRS Individual Income Tax Data is less representative of households with below-average income. **How can we account for the difference in representation and its potential to affect our results?**
4. The BLS conducts the CES Interview survey quarterly, but they ask respondents about their income during the prior year at each interview. For example, an interview that occurs in January 2017 will be part of the 2017 CES dataset but the household's responses will regard their income during the prior 12 months and expenditures during the prior 3 months—i.e. 2016 income and expenditures occurring between October 2016 and January 2017. The IRS Individual Income Tax Data represents income accrued during the 2017 tax year, which for most households is the 2017 calendar year. The alignment of income and expenditures can affect the correlation between the two, and therefore model performance. **Which CES Interview quarters would you suggest using in the sample?**

## Variables available in both IRS Individual Income Tax Data and CES data

- Family size
- Number of persons under age 18 in household
- Number of persons over age 64 in household
- Income
- State of residence

# K-Fold Cross Validation

1. Split sample into K equally-sized groups
2. Repeat K times:
  - i. Estimate model using K-1 groups
  - ii. Use model to predict outcomes for Kth group
  - iii. Calculate prediction error for Kth group
3. Take mean of prediction error estimates for all groups

Example: 3-Fold Cross Validation

Iteration	Group 1	Group 2	Group 3
1	V	T	T
2	T	V	T
3	T	T	V

# Cross-Validation Results

	CES Interview Survey		CES Diary Survey	
Model	Mean Square Error	Standard Deviation	Mean Square Error	Standard Deviation
OLS	1,255,000	332,000	331,000	169,000
Lasso	1,256,000	332,000	331,000	170,000
ERC Lasso	1,261,000	334,000	331,000	170,000
KNN	1,285,000	331,000	339,000	170,000
GBR	1,619,000	383,000	460,000	140,000
Base Model 2002	1,886,000	390,000	687,000	385,000
Base Model 2016	2,347,000	364,000	761,000	412,000

## Statewide home ownership rate, by income decile

Income Decile	Mean Household Income	Home Ownership Rate
0	\$8,000	33%
1	\$22,000	42%
2	\$34,000	49%
3	\$46,000	54%
4	\$59,000	61%
5	\$74,000	67%
6	\$91,000	71%
7	\$114,000	76%
8	\$149,000	82%
9	\$287,000	87%