Off to the Races:
A Comparison of Machine Learning and Alternative Data for Predicting Economic Indicators

Jeffrey C. Chen, Abe Dunn, **Kyle Hood**,
Alex Driessen and Andrea Batch

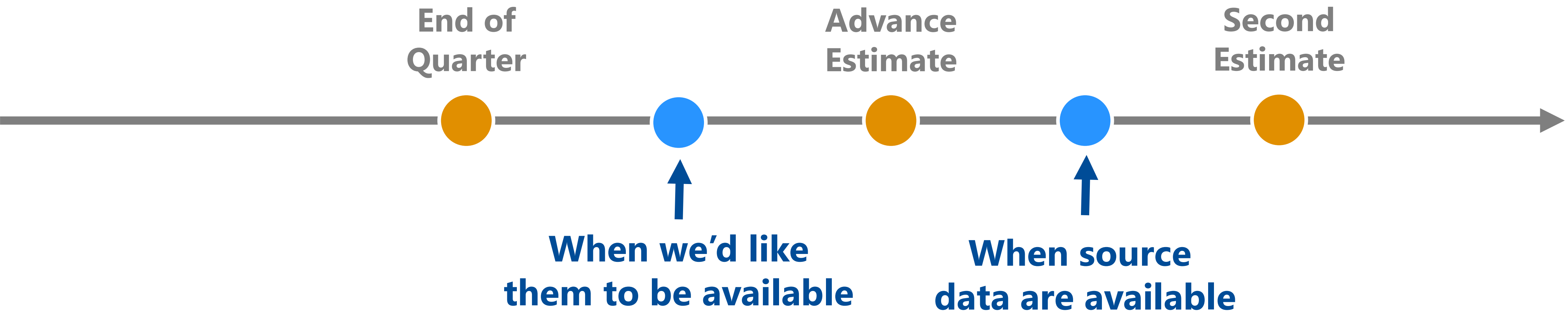
Federal Economic Statistics Advisory Committee Meeting
June 14, 2019

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

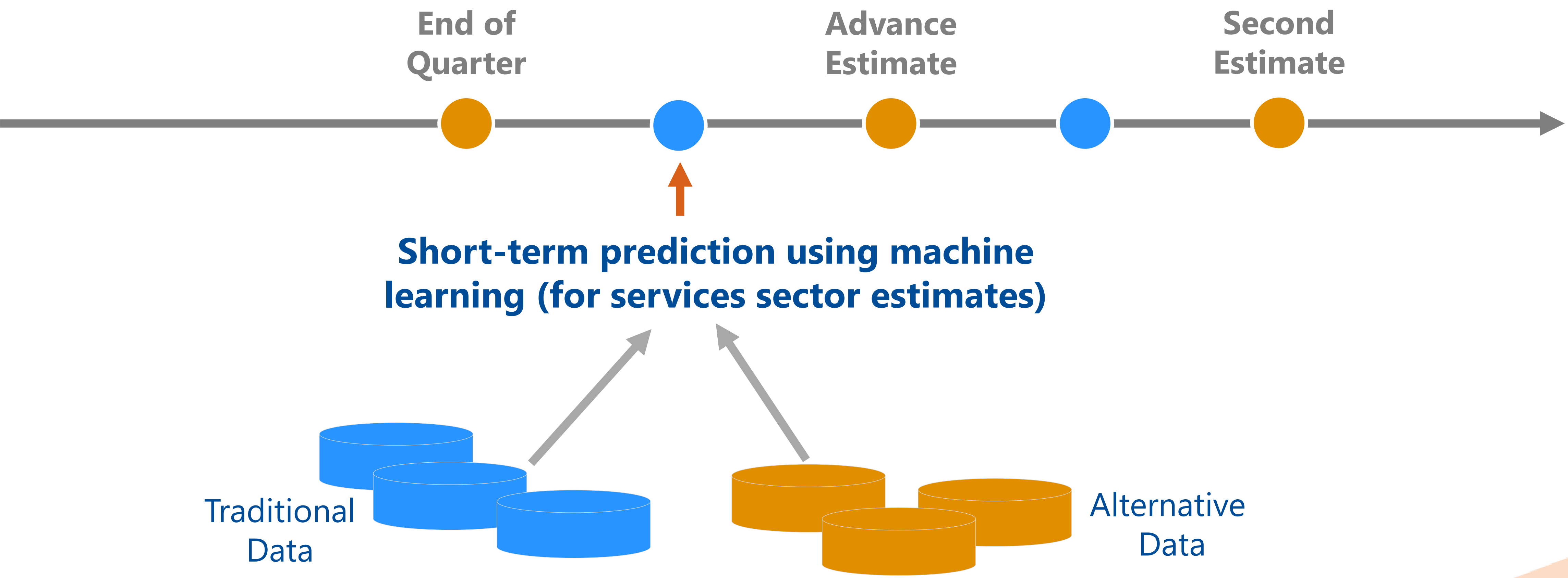
Roadmap

1. Motivation
2. Approach
3. Results
4. Implications

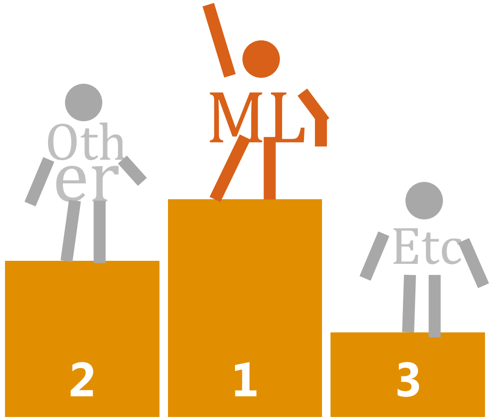
Timing of GDP Estimates



Timing of GDP Estimates



Objectives: **ML for National Economic Accounts**



- Identify which modeling considerations (e.g. algorithm, data, feature selection) are associated with accuracy gains for **PCE services component** of GDP.

M1 vs. M2

- Develop a framework to determine where predictions can be reliably applied to reduce revisions given sample size constraints.

There's more variables than records.

Issue

Traditional statistical methods have trouble with k > n

Id	Y	X1	X2	X3	X4	X5	. . .	x999
1								
2								
3								
.								
.								
29								

Which variables to choose?!

Solution

Many ML methods can efficiently sift through inputs that maximize predictive accuracy.

Id	Y	X1	X2	X3	X4	X5	. . .	x999
1								
2								
3								
.								
.								
29								
		3	1	2				

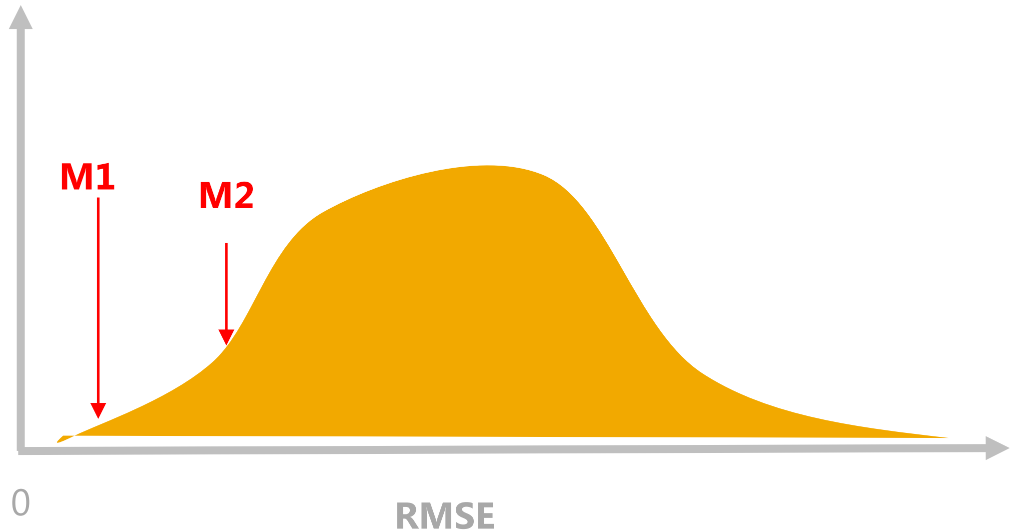
Ranked

Motivation

Small samples call for different strategies.

Issue

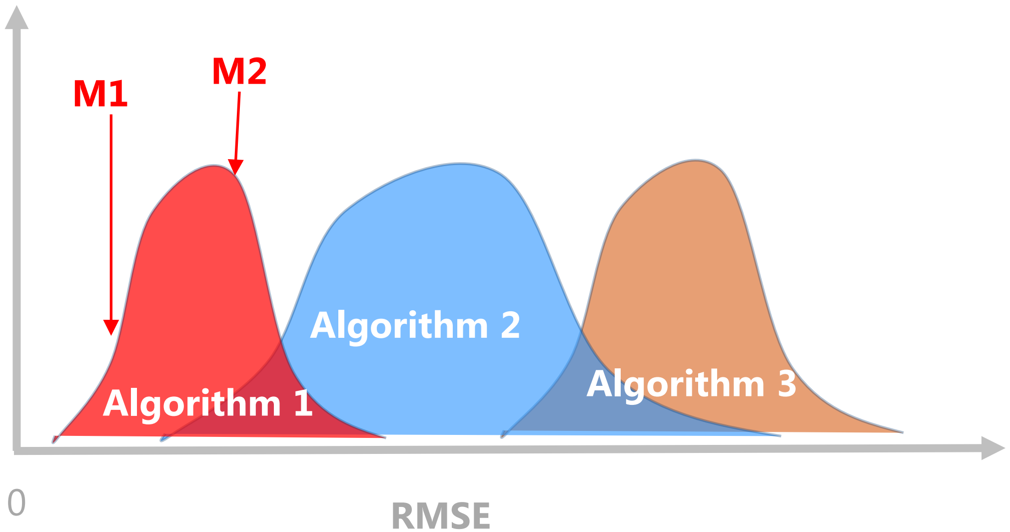
Typical goal of prediction is to crown a definitive winner among all tested models.



(While M1 is better than M2, in small samples there is effectively no difference)

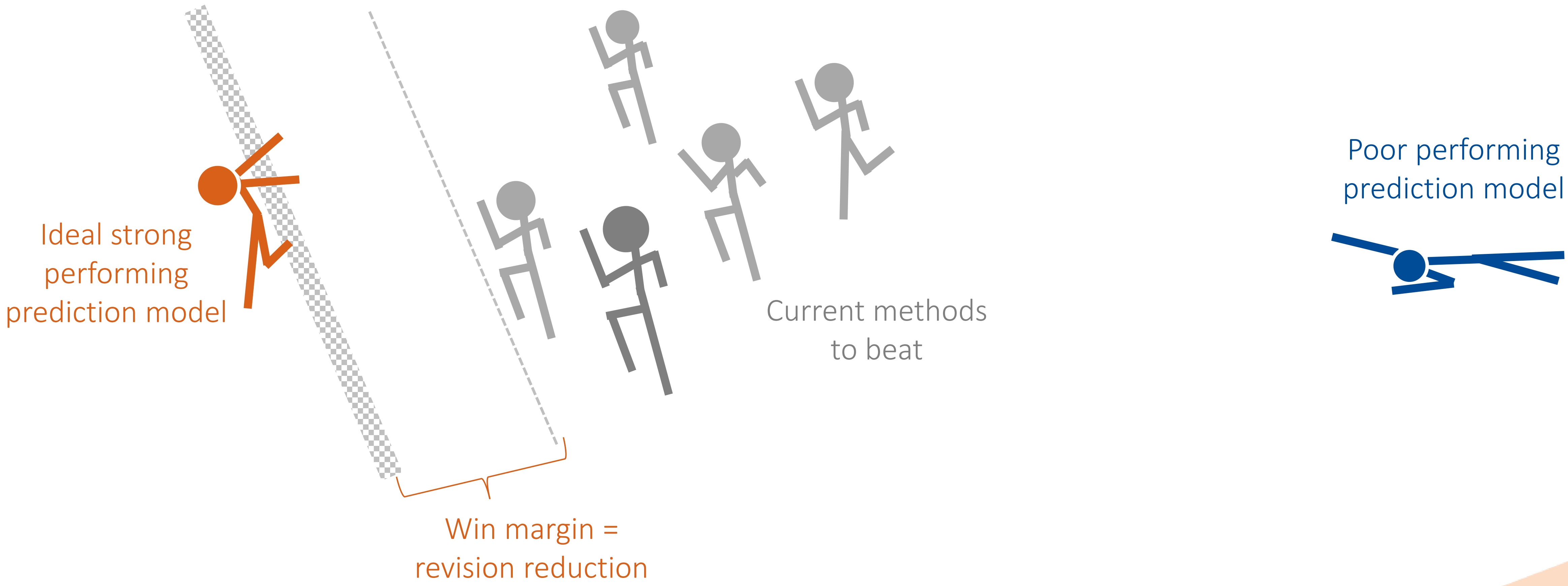
Solution

For national accounts, the ideal is to find a general set of approaches that will consistently yield accuracy gains.



(If M1 and M2 are derived from the same algorithm but with different inputs, we can form a strategy around a class of algorithm)

Predictions must beat current methods.



A Prediction Horse Race

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

1 **Prediction Horse Race**

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p},)]$$

2 **Evaluate Absolute Performance**

3 **Identify Best Relative Reductions**

Predict the Quarterly Services Survey (QSS).

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p})]$$

"Predict quarterly industry growth $\boldsymbol{y_{it}}$ using a large number of combinations of algorithms, data, and variable selection methods"

Step 1: **Data in Horse Race**

Draw on a broad range of potential source data to compare traditional sources and alternative sources.

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p})]$$

Quarterly Services Survey
U.S. Census Bureau

188 industry series
n = 31 quarters

Source data for significant proportion of PCE Services

Credit Card Transactions
First Data – Palantir/ Fed Board Revised Series
192 industries

Search Queries
Google Trends
230 associated searches

Current Employment Survey
BLS
140 industries

Consumer Price Index
BLS
600+ indexes

Lagged QSS
U.S. Census Bureau

188 industry codes lagged for t-4 to t-1

Step 1: **Algorithms in Horse Race**

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p})]$$

4Q Moving Average	Ridge Regression	Extreme Gradient Boosting
Stepwise Regression	CART	Support Vector Machines
LASSO Regression	Random Forest	Multi-Adaptive Regression Splines

Step 1: **Algorithms in Horse Race**

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p})]$$

4Q Moving Average	Ridge Regression	Extreme Gradient Boosting
Stepwise Regression	CART	Support Vector Machines
LASSO Regression	Random Forest	Multi-Adaptive Regression Splines

Type of Method

Univariate

Multivariate Regression

Non-Linear or Non-Parametric

Step 1: **Algorithms in Horse Race**

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p})]$$

4Q Moving Average	Ridge Regression	Extreme Gradient Boosting
Stepwise Regression	CART	Support Vector Machines
LASSO Regression	Random Forest	Multi-Adaptive Regression Splines

Single or Ensemble (many in one)

Single

Ensemble

Step 1: **Variable Selection Procedures in Horse Race**

$$y_{it} = f_m[g_k(X_t, Y_{i,t-p})]$$

Cherry Picking

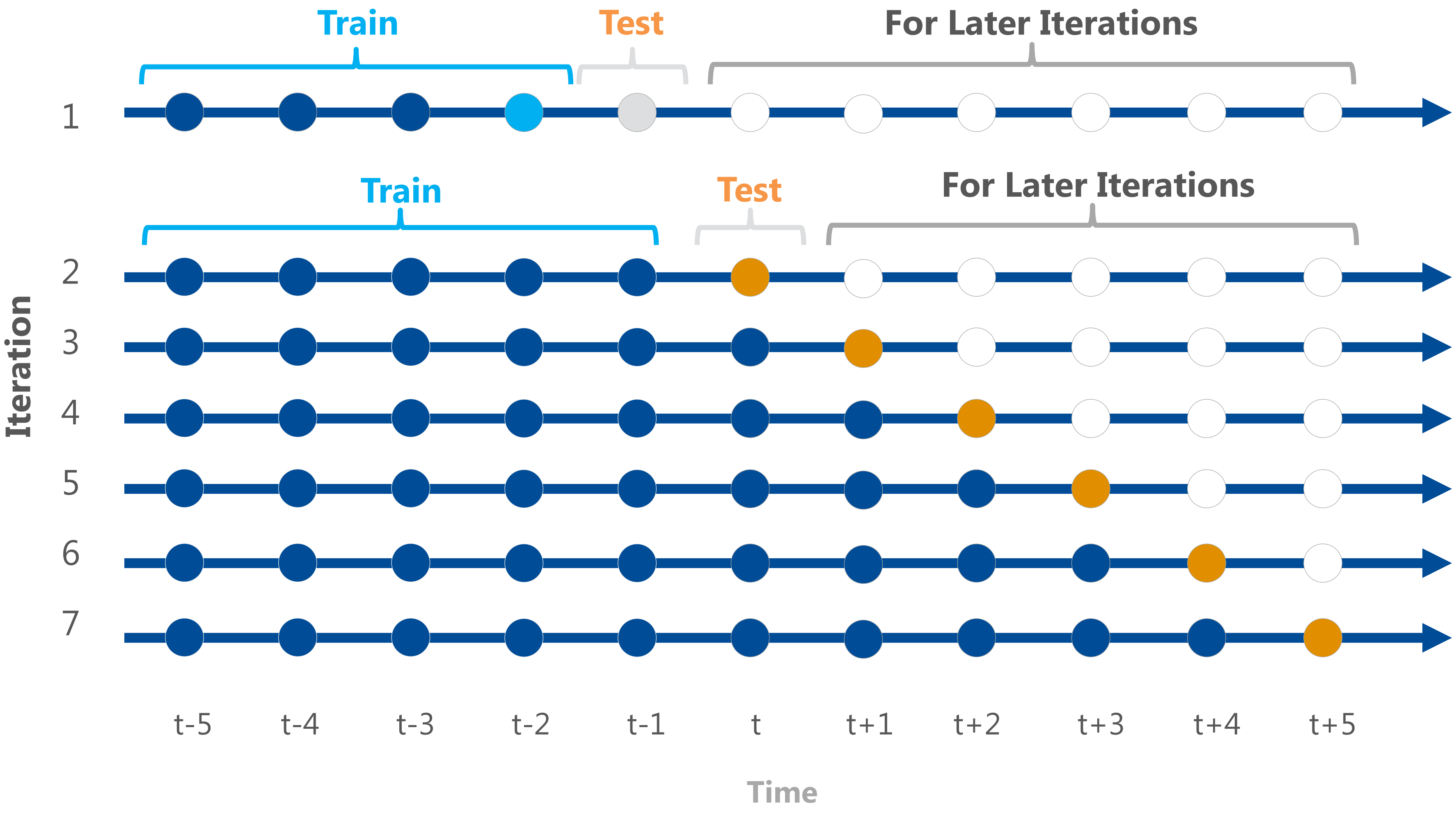
Include only conceptually similar variables.

Kitchen Sink

All-in.

25 data set combinations

Methods: **One-Step Ahead**



Methods: **A Prediction Horse Race**

For this study **886,608** models were trained, based on the combinations of

industry

x

data sets

X

algorithm

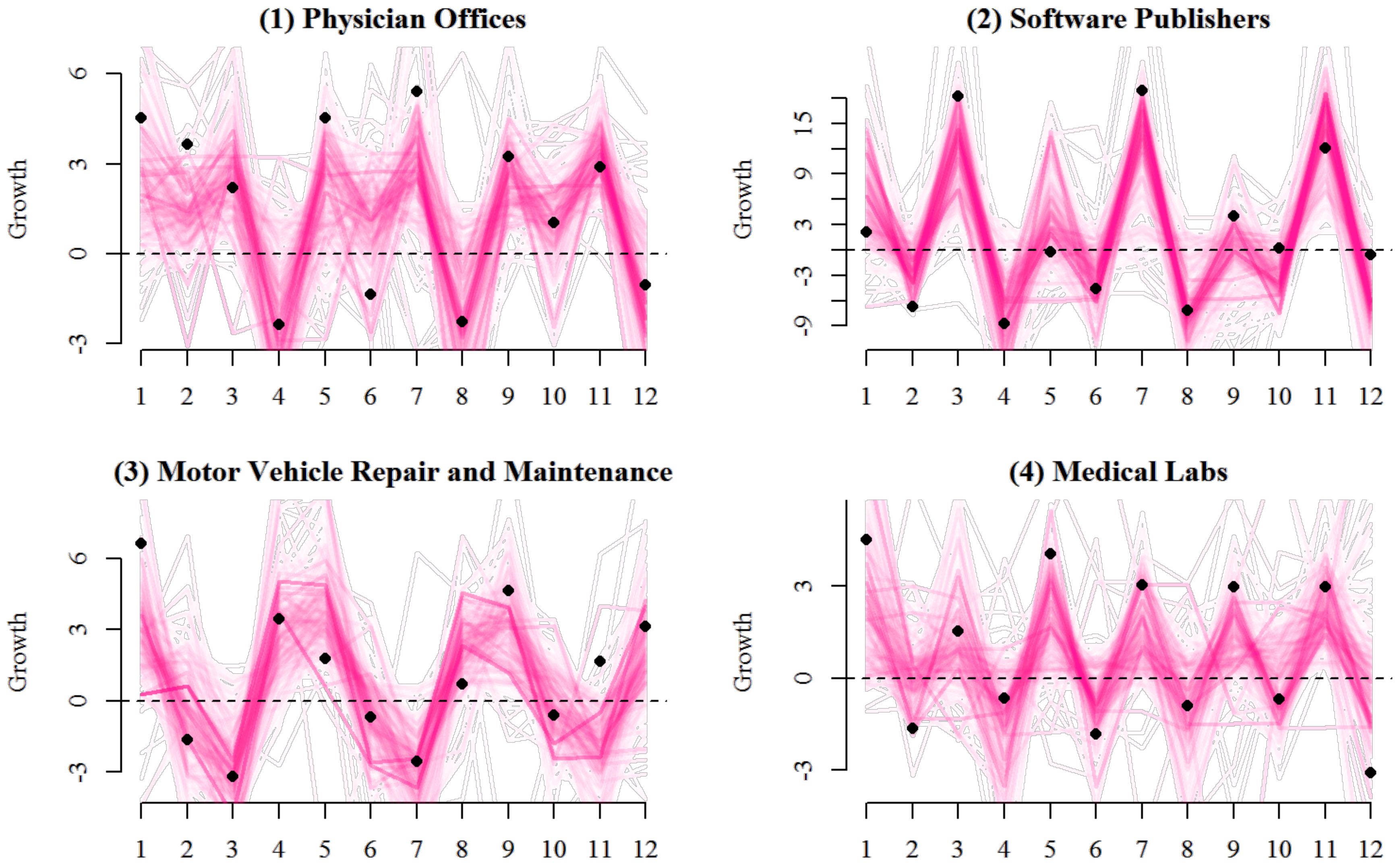
X

variable selection

X

time period

Prediction tracks show agreement and [disagreement] in growth patterns.



Approach (Part 2): **Evaluating Absolute Performance**

1. **Prediction Horse Race**
   $y_{it} = f_m[s_k(X_t, Y_{i,t-p},)]$

2. **Evaluate Absolute Performance**

3. **Identify Best Relative Reductions**

Measure what generally leads to an accuracy increase in the QSS

Absolute Performance

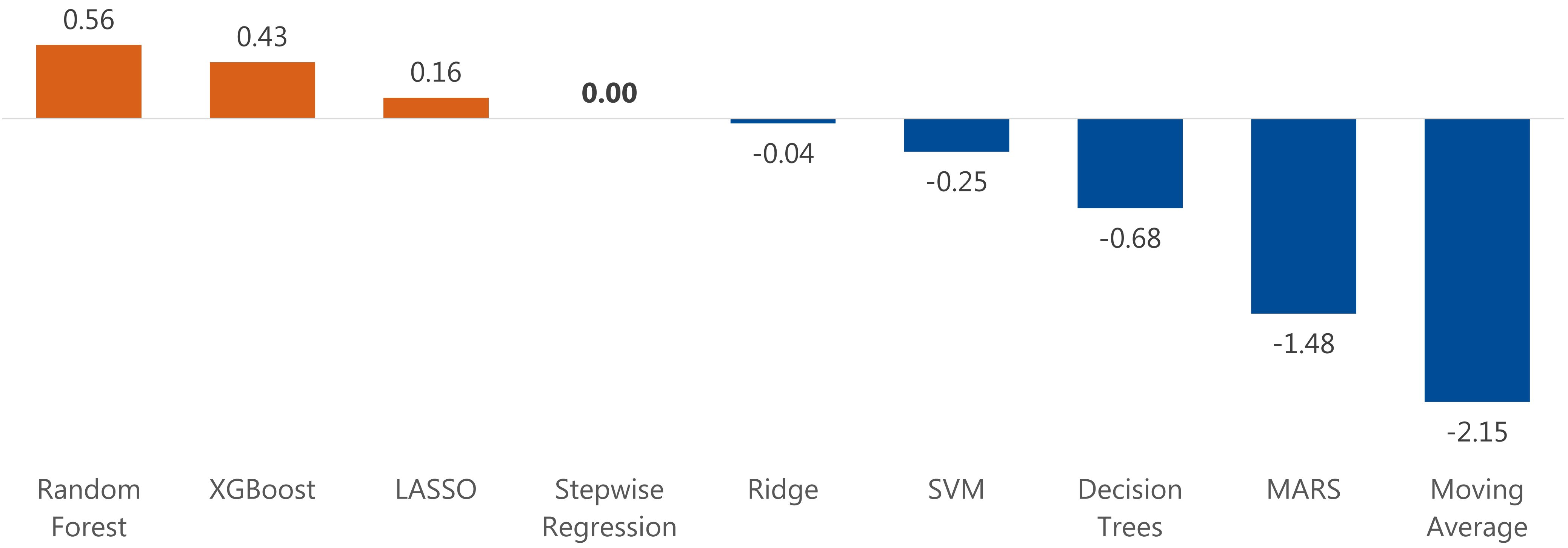
Average Absolute Accuracy

$$RMSE_{i,k,m} = \beta + \alpha_i + \gamma_m + \xi_k + \varepsilon_{i,k,m}$$

Estimate a **fixed-effects regression** to parse out the average accuracy gain associated with each algorithm, data set, etc.

Results: **Average RMSE Improvement (Relative to Stepwise)**

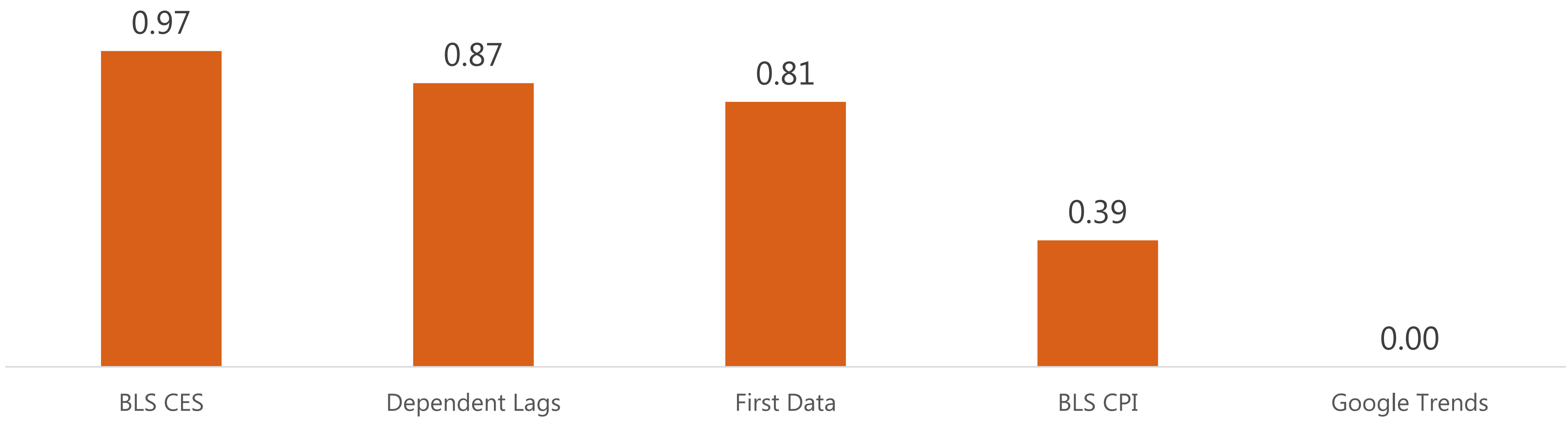
Takeaway: On average, ensemble methods improve accuracy the most.



Absolute Performance

Average RMSE Improvement (Relative to Google Trends)

Takeaway: Measures of consumption and employment help the most. Also, the processes are strongly seasonal.



Absolute Performance

More data might not better, and cherry picking does not help.

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

Cherry Picking vs. Kitchen Sink

-0.28 Cherry Picking *adds* error to predictions.

Number of Data Sets (Need to be considered in conjunction with dataset parameter estimates)

-0.31 ***Two data sets*** add some additional error, but can be offset depending on the datasets that are combined.

-0.8 ***Three data sets*** add a disproportionate amount of error, but no three data set combination is better than a two data set combination.

Revision Impacts

1 **Prediction Horse Race**

$y_{it} = f_m[g_k(X_t, Y_{i,t-p},)]$

2 **Evaluate Absolute Performance**

3 **Identify Best Relative Reductions**

Convert QSS into PCE and find sure-fire improvements compared with current

Calculate Sustainable Improvements

1. Convert QSS into PCE services components

$$\hat{C}_m = g_c(\hat{y}_{it})$$

2. Calculate **Percent Improved Periods (PIP)**

3. Calculate **Mean Revision Reduction Probability (MRRP)**

Mean Revision Reduction Probability

1. Calculate the Root Mean Squared Revision for each model ***m*** and ***current*** BEA methods.

$$\mathrm{RMSR}_{\mathrm{current}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathrm{C}}_{\mathrm{current}} - \mathrm{C}_{third})^2} \qquad \mathrm{RMSR}_{m} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathrm{C}}_{m} - \mathrm{C}_{third})^2}$$

2. Calculate revision reduction for model ***m***

$$\Delta\mathrm{RMSR}_m = \mathrm{RMSR}_m - \mathrm{RMSR}_{current}$$

3. Estimate probability that any model will result in revision reduction for component **C**

$$MRRP_c = \frac{1}{M}\sum_{m=1}^{M}(\Delta\mathrm{RMSR}_m < 0)$$

Percent Improved Periods (PIP)

How *often* do models offer an improvement?

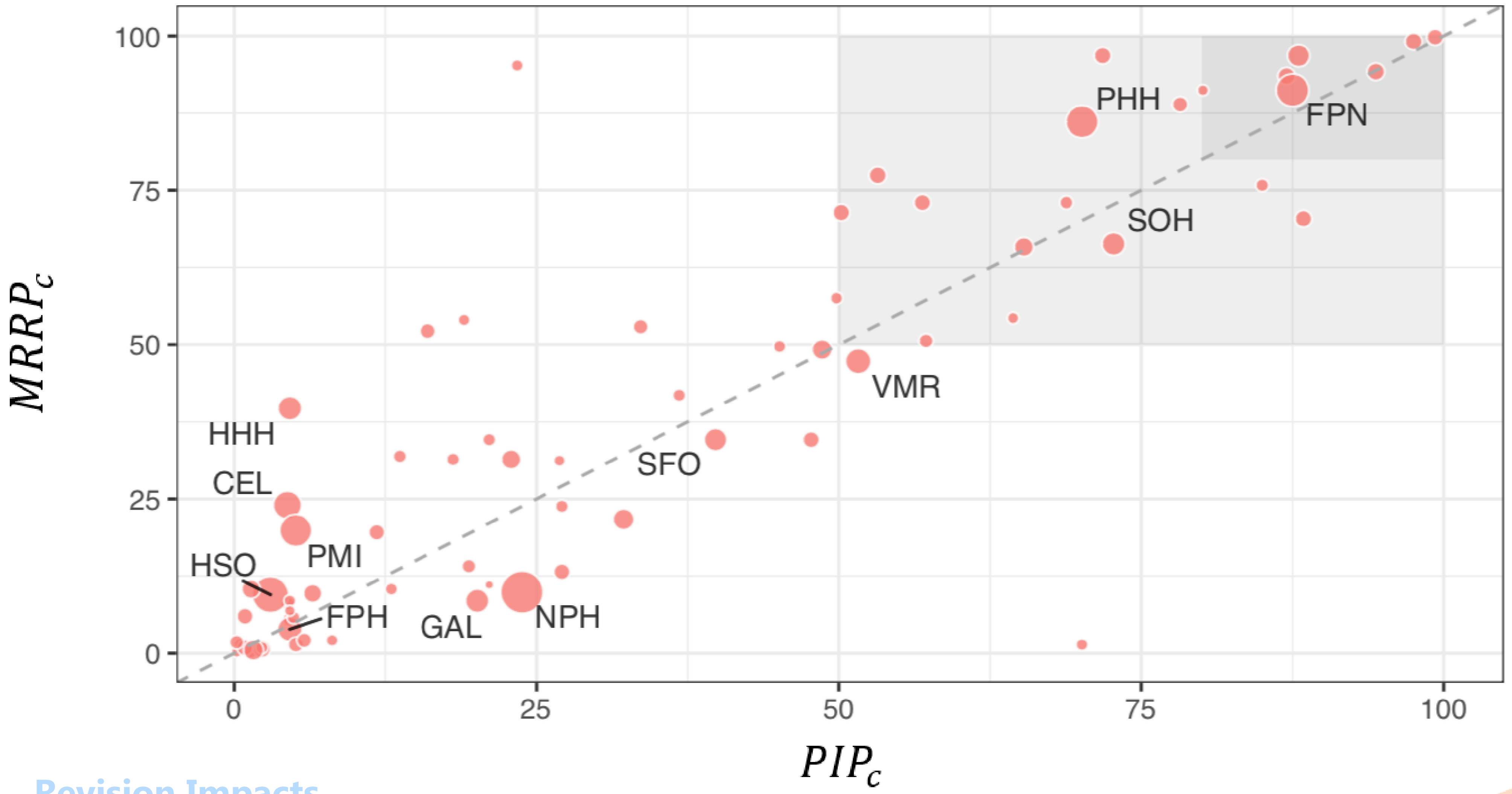
1. Calculate the Root Mean Squared Revision for each model ***m*** and ***current*** BEA methods.

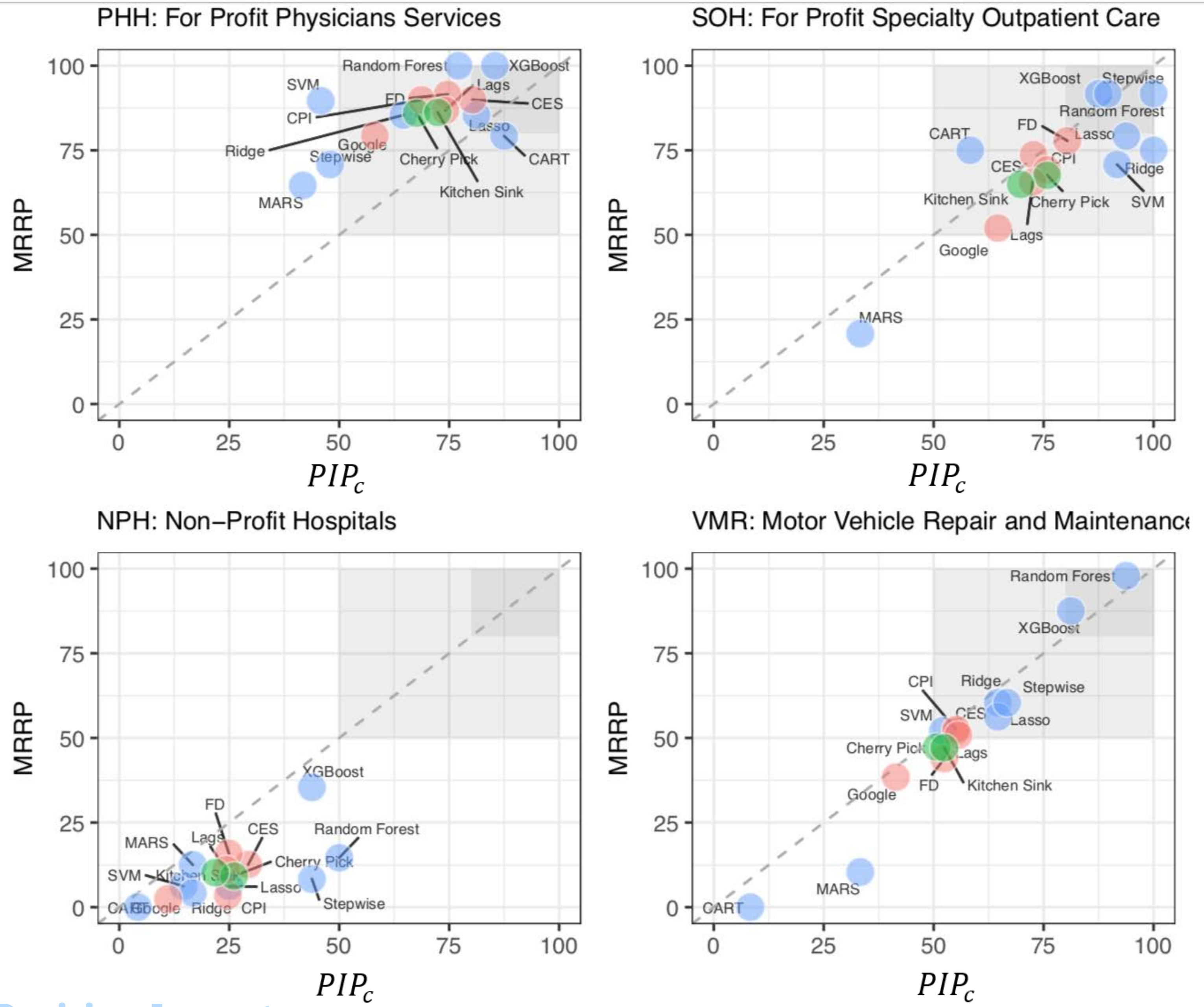
$$PIP_m = \frac{1}{T}\sum_{i=1}^{T}(|\hat{C}_{mt} - C_{third,t}| < |\hat{C}_{current,t} - C_{third,t}|)$$

2. Calculate average revision reduction using model ***m***

$$PIP_c = \frac{1}{M}\sum_{m=1}^{M}(PIP_m > 0.5)$$

Identifying predictable series comparing MRRP and PIP





Given the methods and data, some algorithms are far less predictable than others.

Mean Revision Impacts for Random Forest models

Component	Percent				Levels ($Mil)		Direction	
	10th	Mean	Median	90th	Mean	Median	ML	Current
PCE	5.59	12.17	13.11	18.33	2054.75	2213.61	100	100
..PCE Services	0.2	10.3	11.78	19.72	1552.69	1775.76	100	100
....Health Care	2.23	11.27	12.64	18.99	1442.62	1618	100	100
....Transportation	2.91	25.57	26.7	43.86	1100.38	1149.29	75	67
....Recreation	4.28	8.47	8.28	12.75	349.73	341.88	92	83
....Education	1.74	3.25	3.11	5.16	17.6	16.83	100	100
....Professional and Other	1.38	4.2	3.72	7.02	77.84	68.89	75	67
....Personal Care and Clothing	21.8	27.37	28.24	31.03	513.85	530.18	92	83
....Social Services and Religious	10.29	14.21	14.7	17.82	155.06	160.42	83	83
....Household Maintenance	-24.25	10.94	16.71	34.38	45.49	69.49	100	92
....GO NP Social Services	0.07	0.43	0.47	0.74	9.37	10.2	33	33
....GO NP Prof Advocacy	26.24	36.99	41.03	47.8	235.12	260.79	100	100

Nowcasting PCE services with traditional methods

- In a separate project, Baoline Chen and Kyle Hood use a more traditional nowcasting approach
- Nowcast PCE services directly (3rd estimate)
- Two algorithms:
  - (1) Model selection
  - (2) Model averaging
- Two models:
  - (1) General bridge equations (BE)
  - (2) Bridging with factors (BF)
- We look at **model** root mean square revisions (RMSR) relative to **actual** RMSR, in percent terms, on a held-out validation sample

Sample Table: Percentage changes in RMSR from bridge equation and bridging with factor models, professional services (PRS)

PCE professional services	ΔRMSR_BE Model	ΔRMSR_BE Outlier Removed	ΔRMSR_BF Model	ΔRMSR_BF Outliers Removed
GAL	**-48.15**	-44.36	-27.64	-27.64
GAH	**-48.04**	-44.10	-29.49	-29.49
AXS	**-43.62**	-10.83	-25.28	-25.28
AXO	**-59.76**	-47.08	-44.97	-38.90
AOO	**-74.15**	-66.03	-62.21	-68.52

Notes: GAL-Legal services; GAH-Private and public legal services; AXS-Nonprofit professional association services; AXO-Private professional association services; AOO-All other organizational services.

Model selection

- Of 85 components, 63 show improvement
- BE outperforms BF in 48/63 (76% of the time)
- RMSR reductions range to 74%

Model averaging

- Of 85 components, 69 show improvement
- RMSR reductions range to 76%

Jeffrey.Chen@bea.gov
Kyle.Hood@bea.gov

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE