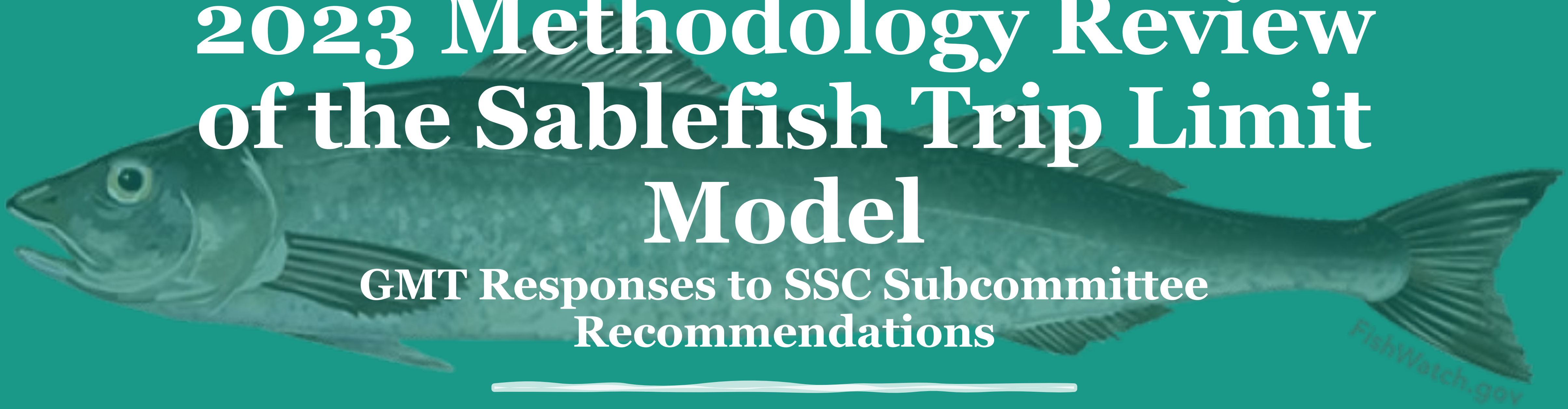
2023 Methodology Review of the Sablefish Trip Limit Model

GMT Responses to SSC Subcommittee Recommendations



October 12, 2023

Groundfish and Economic Subcommittees Workshop #2

Whitney Roberts (WDFW, GMT)

Outline

- Section 1. Introduction
- Section 2. General Updates and Evaluation Methods
- Section 3. Limited Entry North of 36° N. lat.
- Section 4. Open Access North of 36° N. lat.
- Section 5. Cumulative Risk to the Sablefish North ACL

Report Corrections

- Figure 15: models in the figure are predicting number of vessels, not average pounds per vessel

- Figure 19: x-axis is bimonthly period, not year

- Section 4.2.1. Tendency to Under- or Over-predict (OAN) should be listed as Section 4.2.2. instead

Section 1. Introduction

SSC subcommittees' recommendation categories:

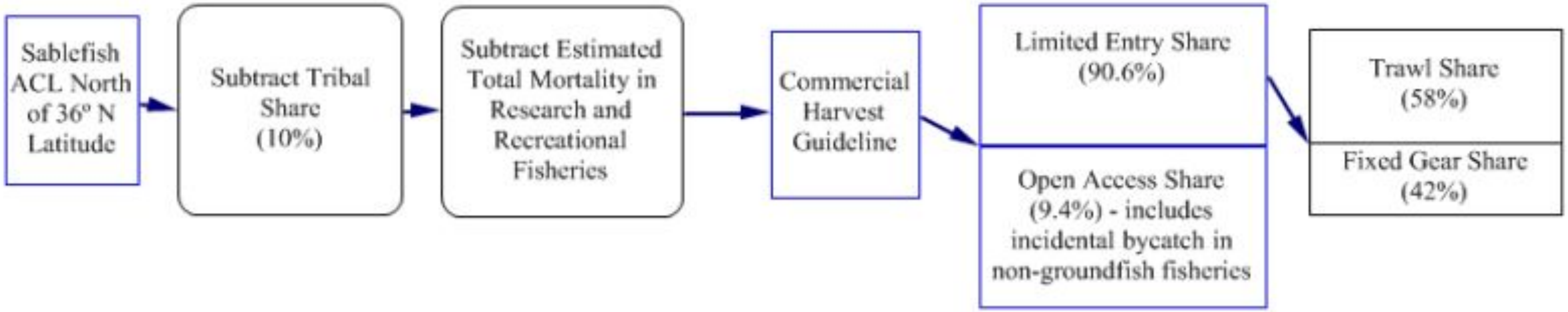
1. **Alternative models should be evaluated using different metrics based on out-of-sample prediction performance.**

2. **Forecast methods should incorporate period-specific outcomes in a transparent and systematic way.**

3. Consider the following when choosing explanatory variables to include in the models.

+ "other recommendations"

Section 5. Cumulative Risk to Sablefish North ACL



- 85% of the sablefish north commercial harvest guideline (HG) is made up of catch share programs
- Remaining 15% is made up of DTL fishery + at-sea whiting set-aside (100 mt in 2023-24)
- 2023 commercial HG was used as proxy for risk to the 2023 sablefish north ACL, because off-the-top deduction mortalities are difficult to predict and full removal of deductions is typically assumed in management
- **Estimated annual risk based on TSCV error: 106-269 mt, which equates to 1.4-3.5% of the total 2023 sablefish north commercial HG**

Section 2. General Updates and Evaluation Methods
Cross-validation Methods

- Time series cross validation with `train()` function in R package "`caret`"

```
myTimeControl <- trainControl(method = "timeslice",
                              initialWindow = 18,
                              horizon = 12,
                              fixedWindow = FALSE)
```

```
lm.mod3.pfe.cov <- train(AVG_LB ~ TL.BIMON + factor(PERIOD) + factor(COVID),
                         data = data,
                         method = "lm",
                         trControl = myTimeControl,
                         metric = "RMSE")
```

- "timeslice" uses function `createTimeSlices()` to partition the data into training and test sets and move those datasets in time for each run
- also performed leave-one-out cross validation using method = "loocv"

Section 2. General Updates and Evaluation Methods
COVID Fixed Effects (2020 and 2021) + Period Fixed Effects

COVID Fixed Effects

$$Average\ lbs.\ per\ vessel \sim bimonthly\ trip\ limit\ +\ factor(COVID)$$

Time data	COVID variable value
2020 periods 1-6 2021 periods 1-6	1
all other data	0

- TSCV generated identical forecasts when training data did not include at least 2020
- Therefore, very little difference in TSCV results with and without COVID variable

Period Fixed Effects

$$Average\ lbs.\ per\ vessel \sim bimonthly\ trip\ limit\ +\ factor(PERIOD)$$

- We did not compare the pooled model with period-specific fixed effects to using separate regressions for each period (status quo)
- However, we did compare the pooled model with period-specific fixed effects to a "baseline" model that is also pooled but does not use period-specific fixed effects

Section 2. General Updates and Evaluation Methods
Time Components (LDV vs. Time Trend)

- TSCV to test performance using lagged dependent variable (LDV) and time trend (Linear Trend Model)
  - R package "`caret`" was not used to perform CV but same parameters were used

all predicting average pounds per vessel

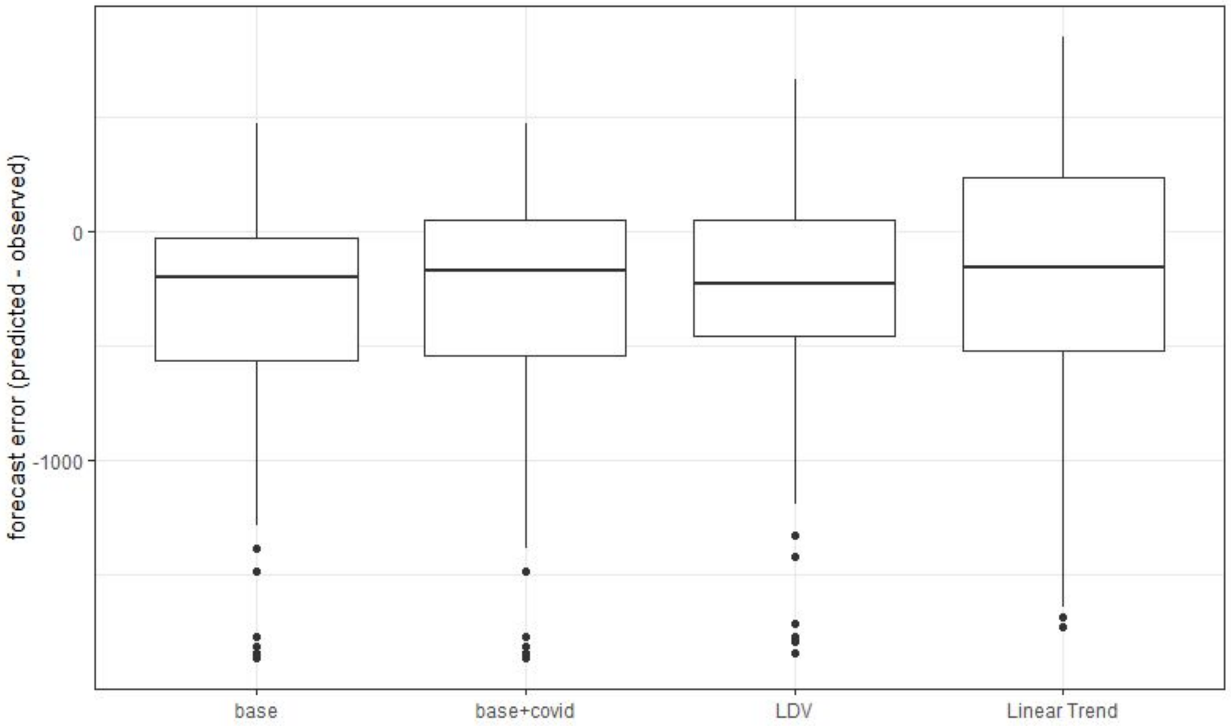
1. The *Base Model*: $AVG\,LB_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 TL.BIMON_t$
2. The Base + Covid Model: $AVG\,LB_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 TL.BIMON_t + \eta D_t^{covid}$
3. The *LDV Model*: $AVG\,LB_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 TL.BIMON_t + \phi AVG\,LB_{(t-1)}$
4. The *Linear Trend Model*: $AVG\,LB_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 TL.BIMON_t + \delta t$

where,
t is a time ordered index of observations, and
$PERIOD_t^j = 1$ if observation t occurs in period j and 0 otherwise.

y-axis limits are not the same

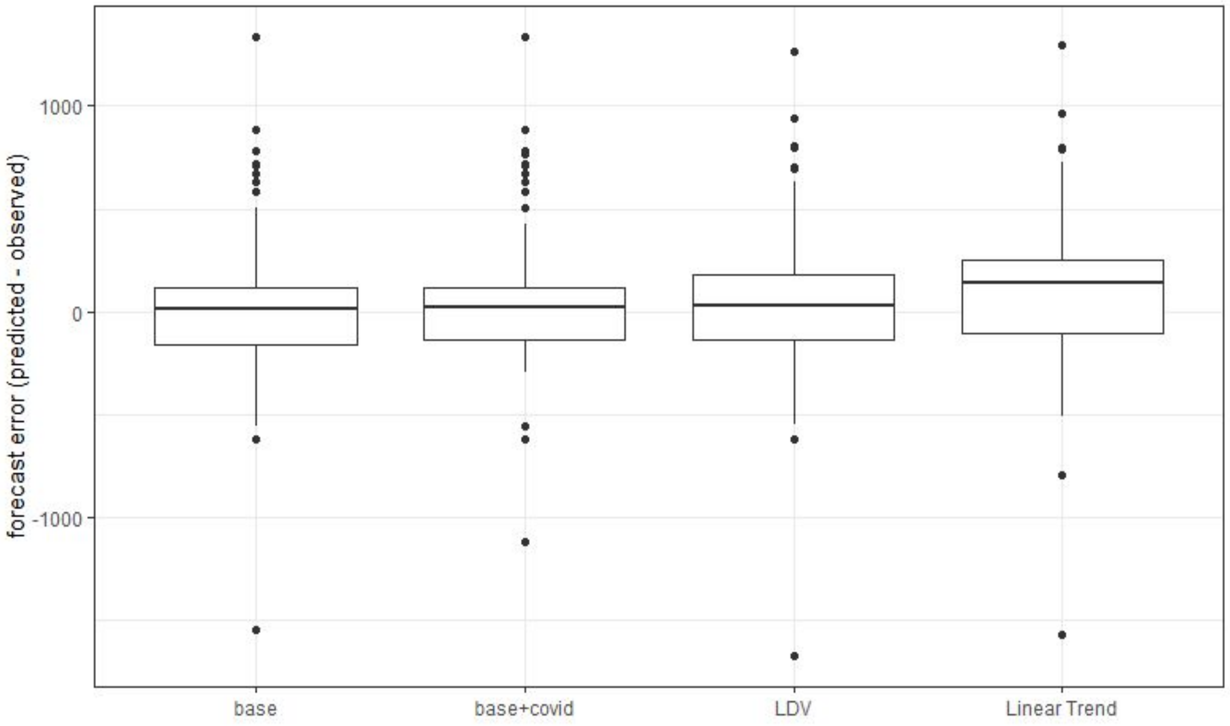
Section 2. General Updates and Evaluation Methods
Time Components (LDV vs. Time Trend)

LEN



RMSE = 639.27 | 638.43 | 599.12 | 598.18 ★

OAN



RMSE = 359.41 ★ | 342.62 ★ | 369.54 | 392.84

Section 2. General Updates and Evaluation Methods
Time Components (LDV vs. Time Trend)

all predicting number of vessels

1. The *Base Model*: $VESSELS_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 Price_t$
2. The Base + Covid Model: $VESSELS_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 Price + \eta D_t^{covid}$
3. The *LDV Model*: $VESSELS_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 Price + \phi VESSELS_{t-1}$
4. The *Linear Trend Model*: $VESSELS_t = \alpha + \sum_{j=2}^{6} \gamma_j PERIOD_t^j + \beta_1 Price + \delta t$

Model	RMSE	
	LEN	OAN
Base Model	6.89	31.07
Base + Covid Model	**6.79**	**30.58**
LDV Model	12.5	65.48
Linear Trend Model	8.46	42.34

Section 2. General Updates and Evaluation Methods
Upweighting Scheme

- Example of current data weighting approach, using dummy variable "WEIGHT":

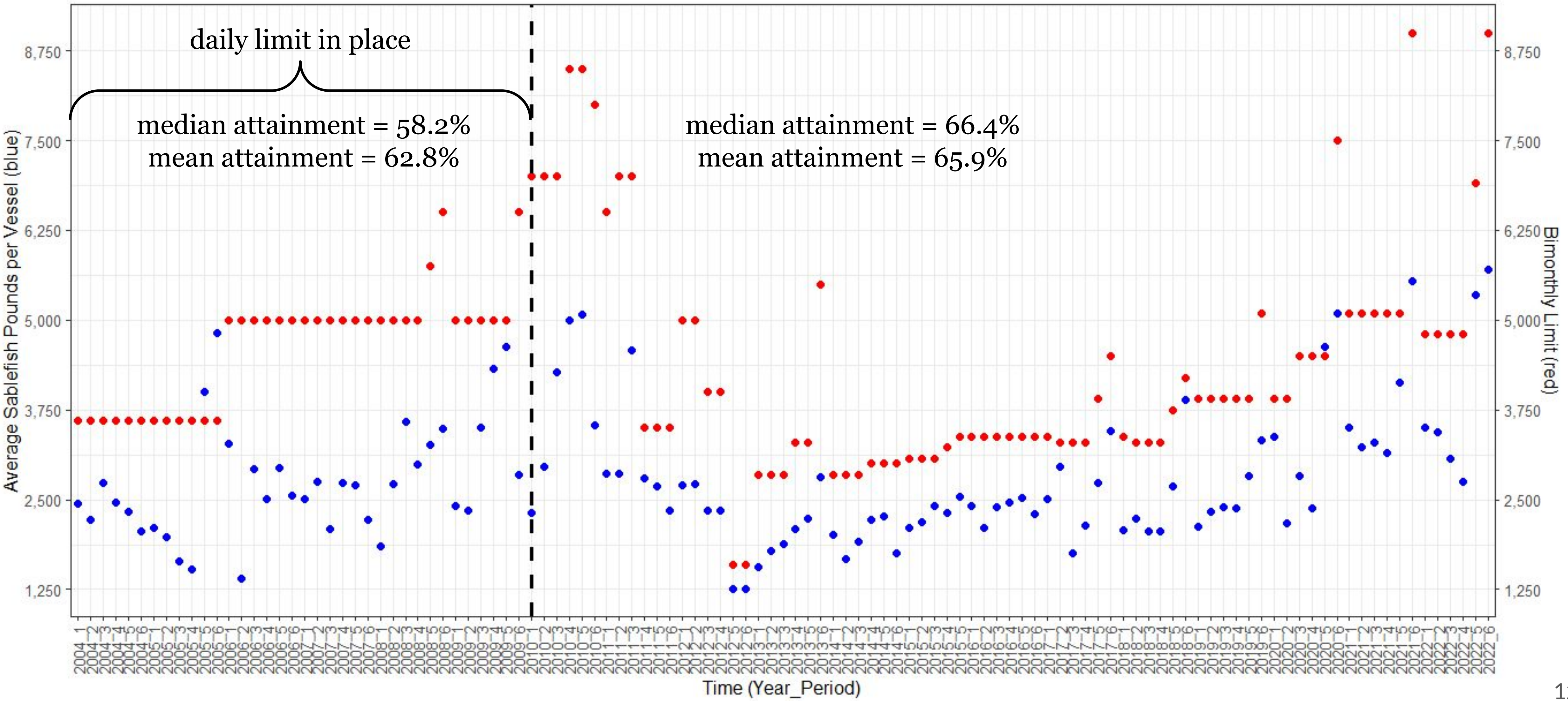
Predicting total 2024 landings

Time data	WEIGHT variable value
2022 periods 1-6 2023 periods 1-6	5
all other data	1

- Very little difference when data weights are included in TSCV and generally arbitrary method
- May not need data weights if time component is included

LEN

Section 3.1. Evaluating Data Prior to 2011

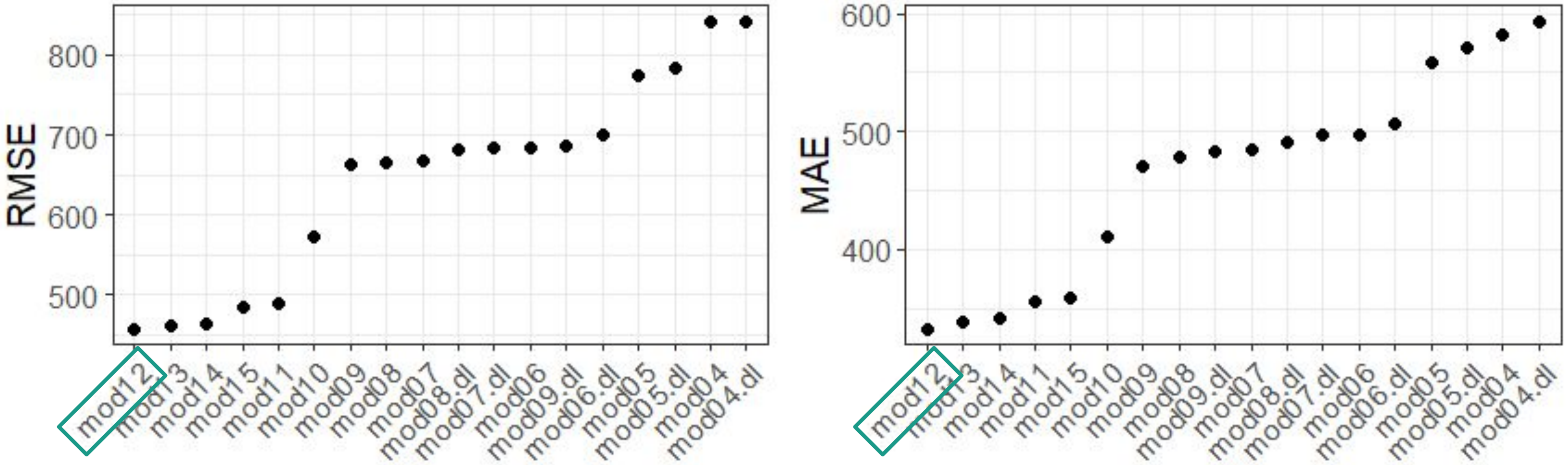


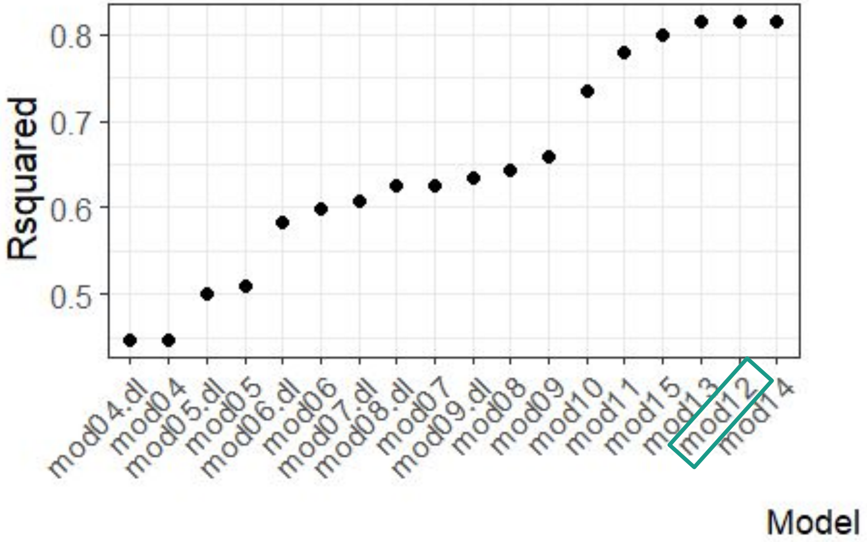
LEN

**y-axis limits are not the same*

Section 3.1. Evaluating Data Prior to 2011

- Leave-one-out cross validation
- All models predict lbs. per vessel using status quo variables
- Order of x-axis varies by plot
- Model used for all further analysis: 2012-2023 (N = 69)

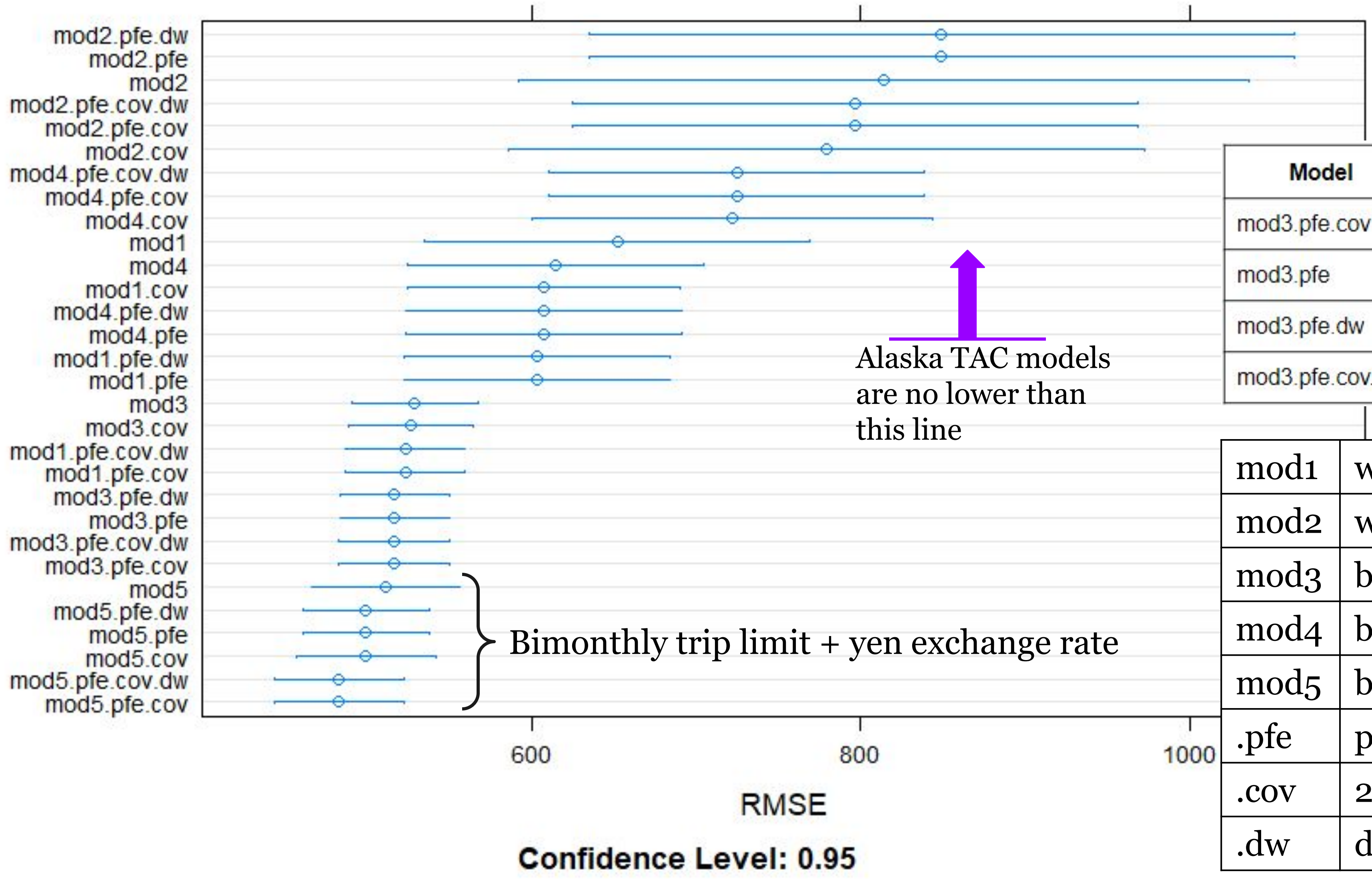




.dl = daily limit factor variable

LEN - pounds per vessel

Section 3.2.1. Time Series Cross Validation Results



Model	MAE	Annual error estimate (mt)	% of LEN Target	% of LEN Share
mod3.pfe.cov	384	58.5	14.0%	2.0%
mod3.pfe	381	58.1	13.9%	2.0%
mod3.pfe.dw	381	58.1	13.9%	2.0%
mod3.pfe.cov.dw	384	58.5	14.0%	2.0%

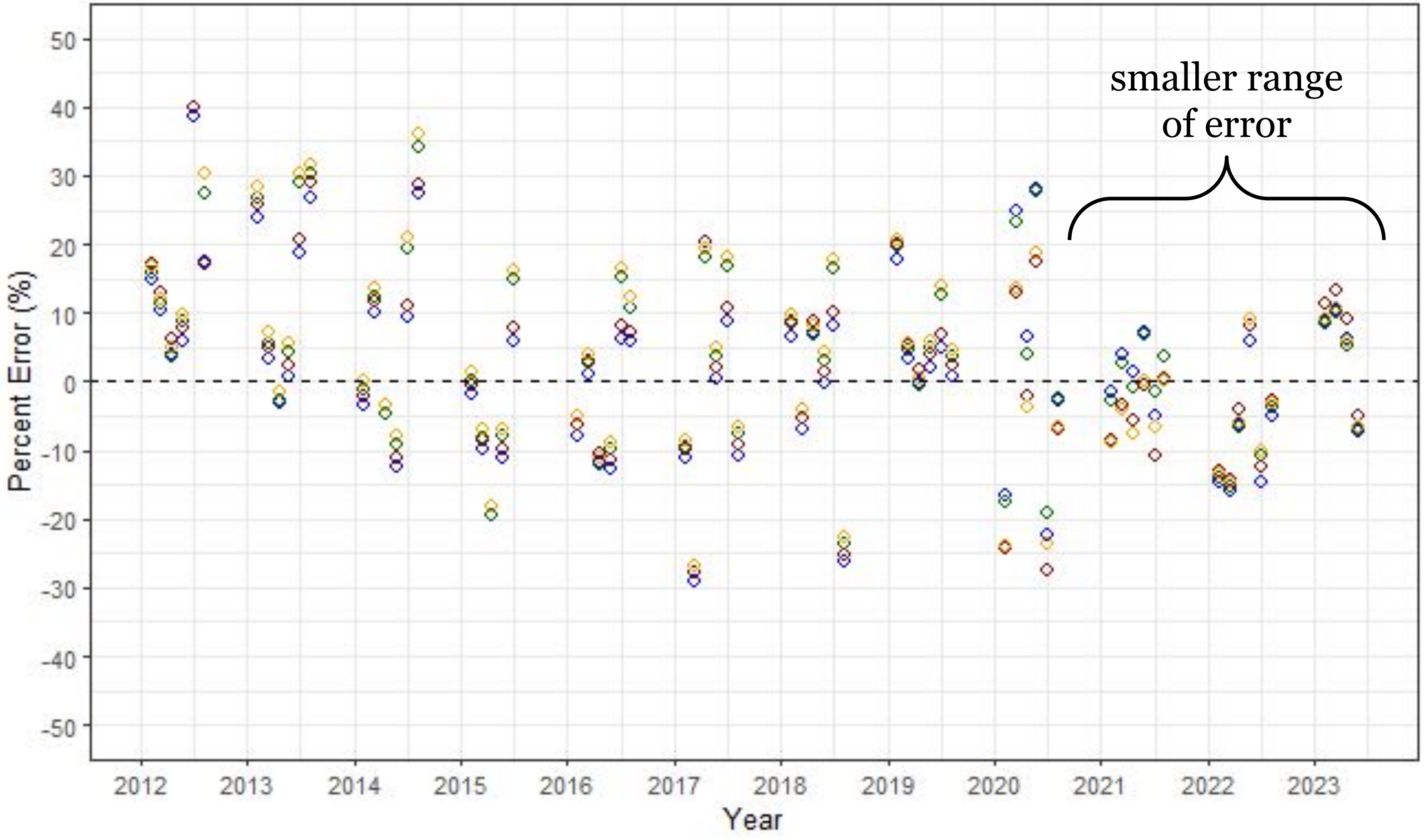
mod1	weekly limit + bimonthly limit (SQ)
mod2	weekly limit
mod3	bimonthly limit
mod4	bimonthly limit + Alaska sablefish TAC
mod5	bimonthly limit + yen exchange rate
.pfe	period fixed effects
.cov	2020 and 2021 as fixed effects (COVID)
.dw	data weights

LEN - pounds per vessel

Section 3.2.2. Tendency to Over- or Under-predict

- top four models from TSCV, excluding models with yen exchange rate (inability to forecast)
- mod3.pfe.cov = lowest **in-sample** MPE
    - underpredicted a maximum of 29% in 2017

Model	Color	MPE
mod3.pfe.cov	blue	1.7568
mod3.pfe	red	1.9660
mod3.pfe.dw	orange	4.0876
mod3.pfe.cov.dw	green	4.2509

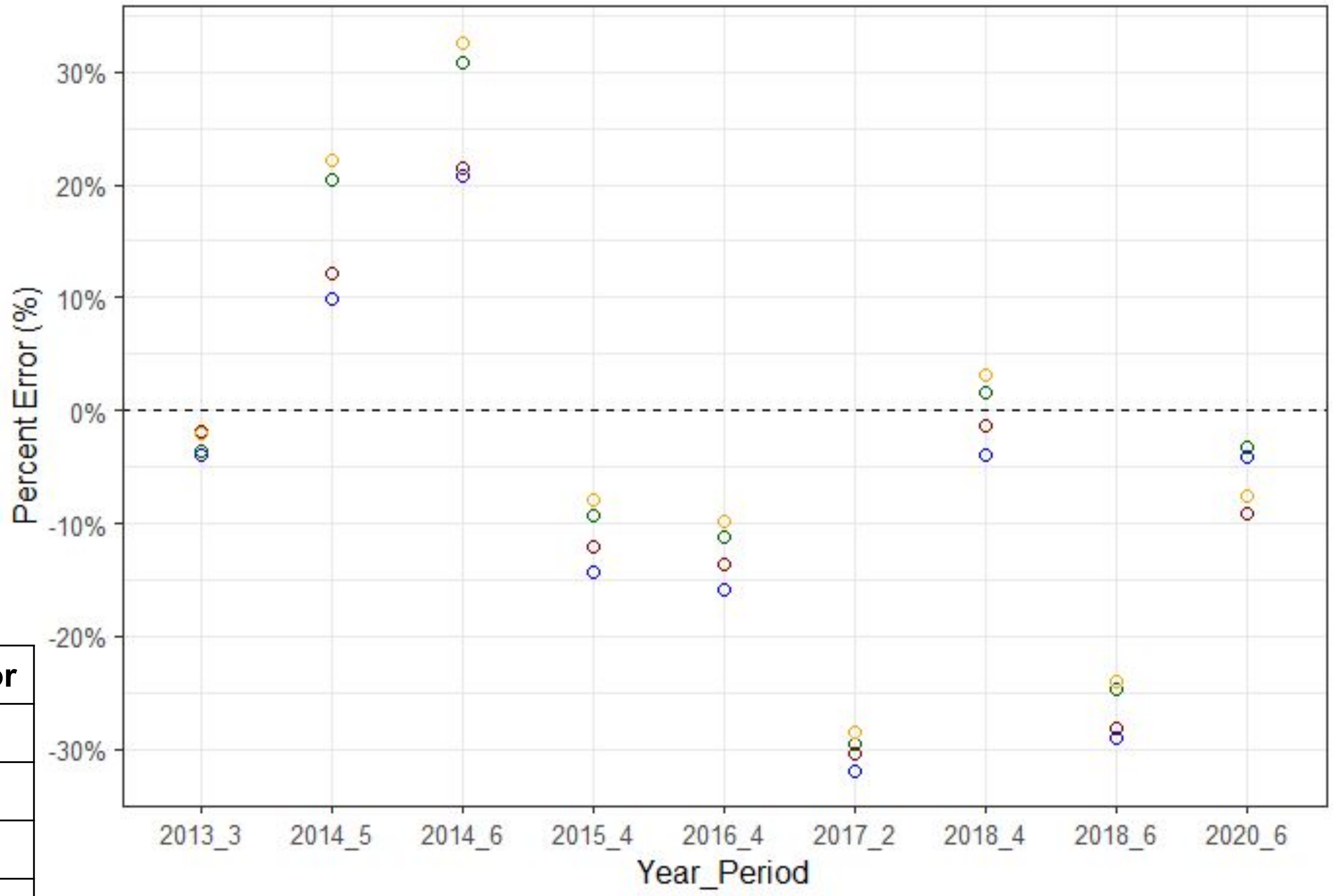


LEN - pounds per vessel

Section 3.2.2. Tendency to Over- or Under-predict

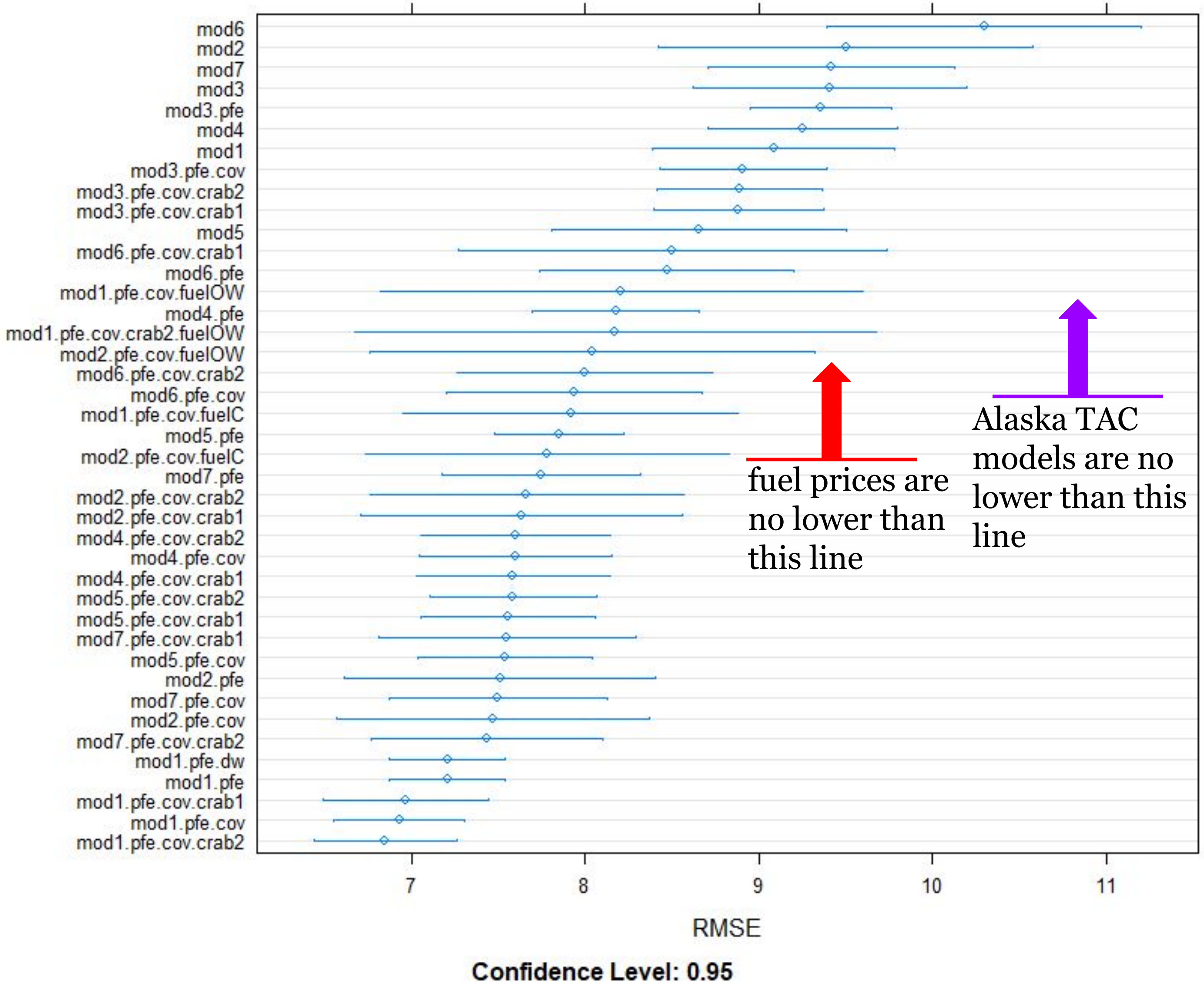
- **Out-of-sample** tendency to under- or -overpredict
- Held out 17% of the dataset (2 years' worth) using `CreateDataPartition()`
- Made predictions of the held-out data based on the training set (excludes the held-out data)
- Underpredicted a maximum of ~30% in period 2 of 2017

Model	Color	MPE	Mean Error
mod3.pfe.cov	**blue**	**-8.1%**	**-286**
mod3.pfe	red	-7.0%	-275
mod3.pfe.dw	orange	-2.4%	-164
mod3.pfe.cov.dw	green	-3.2%	-168



LEN - number of vessels

Section 3.3.1. Time Series Cross Validation Results

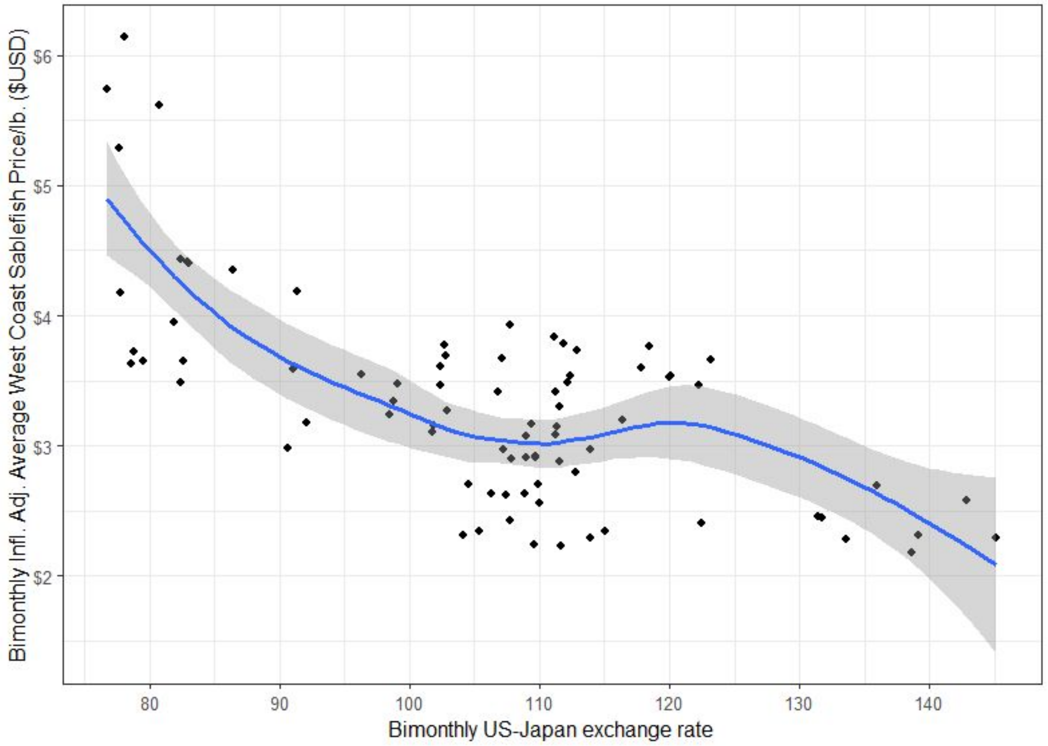


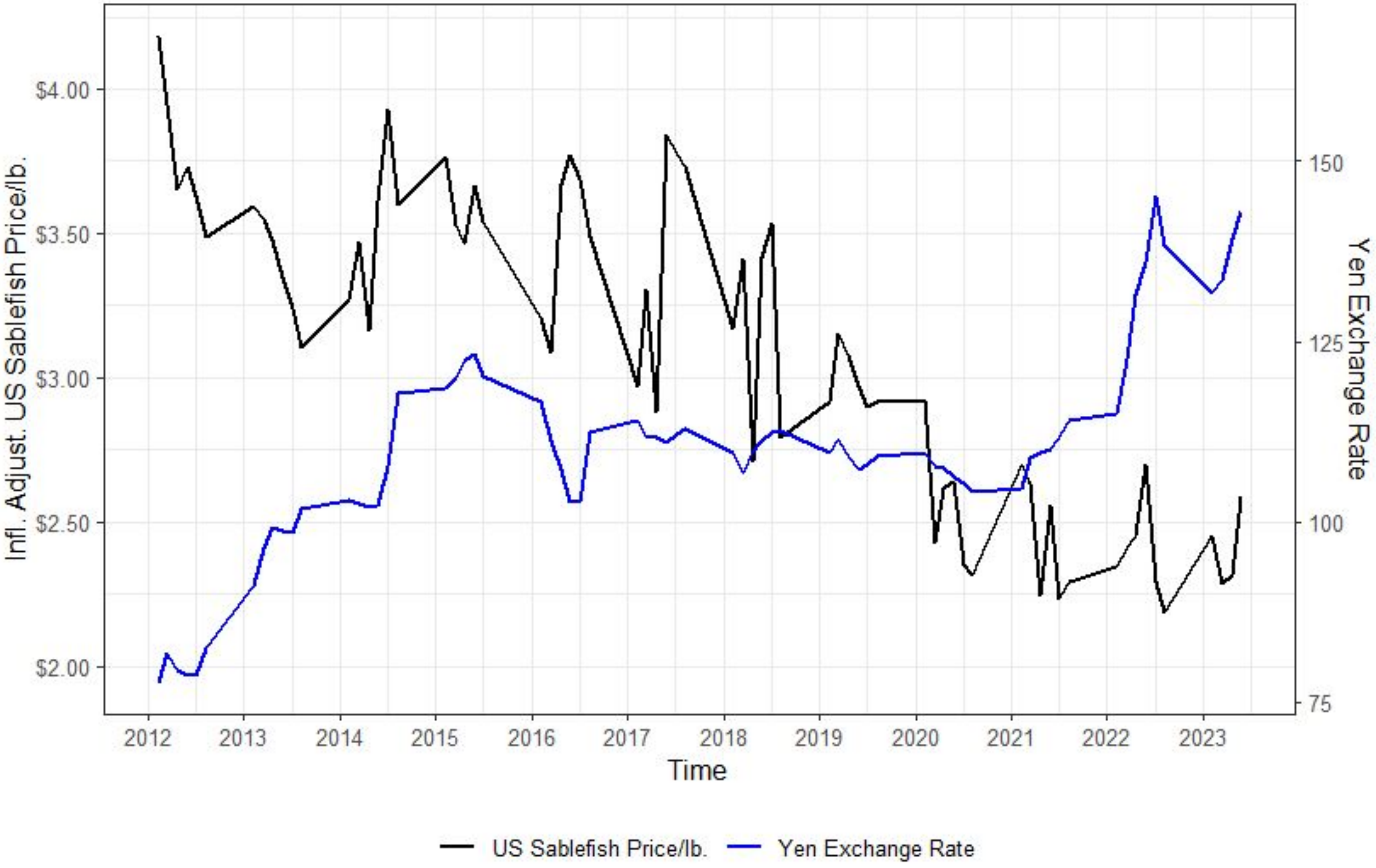
Model	MAE	Annual error estimate (mt)	% of LEN Target	% of LEN Share
mod1.pfe.cov.crab2	5.69	76.9	18.4%	2.8%
mod1.pfe.cov	5.74	77.6	18.6%	2.8%
mod1.pfe.cov.crab1	5.76	77.9	18.7%	2.8%

mod1	avg. sablefish price
mod2	avg. sablefish price + max. sablefish price
mod3	max. sablefish price
mod4	median sablefish price
mod5	median sablefish price + max. sablefish price
mod6	avg. sablefish price + Alaska sablefish TAC
mod7	avg. sablefish price + yen exchange rate
.pfe	period fixed effects
.cov	2020 and 2021 as fixed effects (COVID)
.crab1	+ avg. Dungeness crab price
.crab2	+ max. Dungeness crab price
.fuelOW	OR and WA fuel prices (combined)
.fuelC	CA fuel prices north of 36° N. lat.
.dw	data weights

Section 3.3.1. Time Series Cross Validation Results

West Coast LEN sablefish prices and the yen exchange rate are likely correlated (inverse relationship)



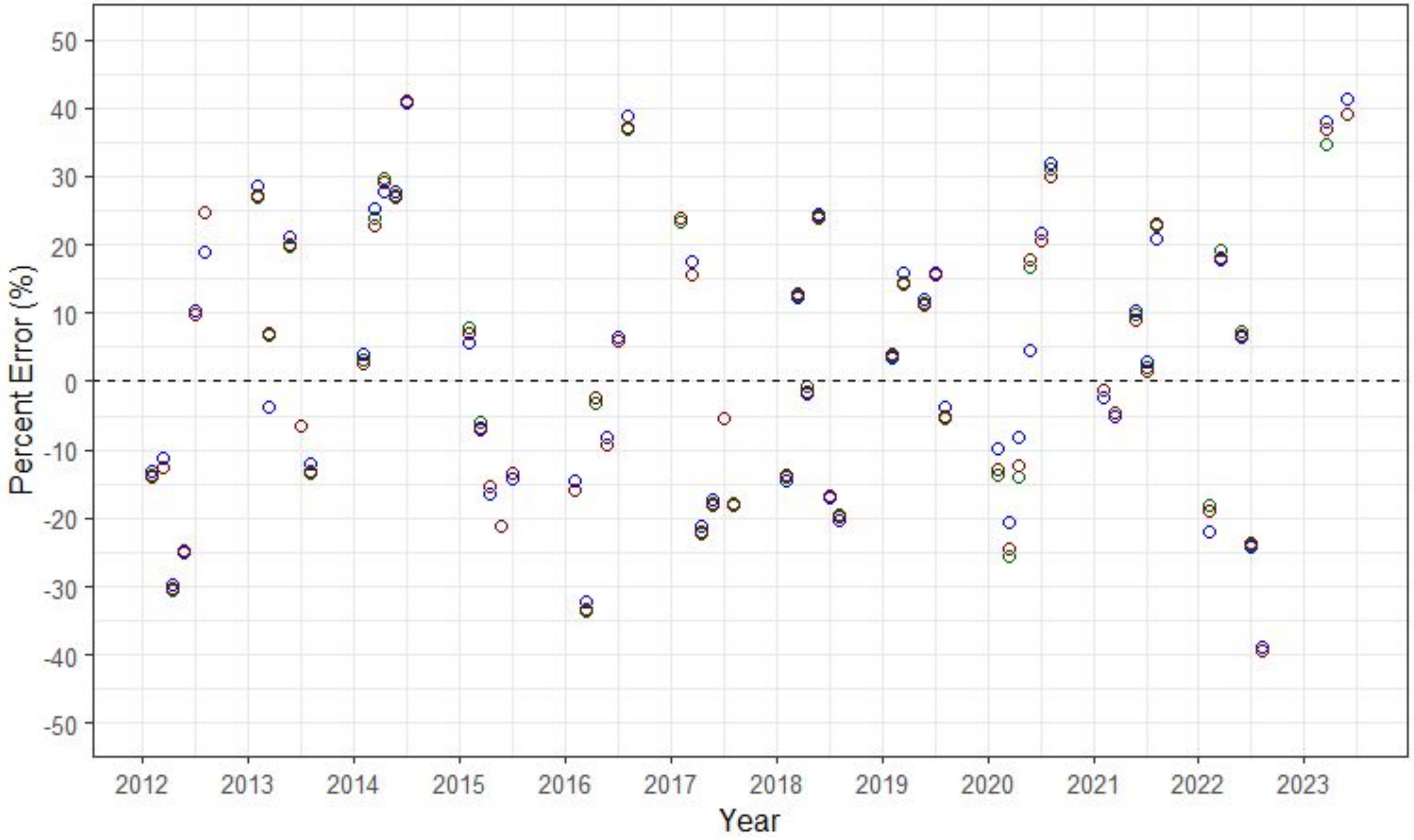


LEN - number of vessels

Section 3.3.2. Tendency to Over- or Under-predict

- top three models from TSCV
- strong underprediction in late 2022 and strong overprediction in early 2023
- mod1.pfe.cov.crab2 = lowest **in-sample** MPE
    - underpredicted a maximum of 39% in 2022

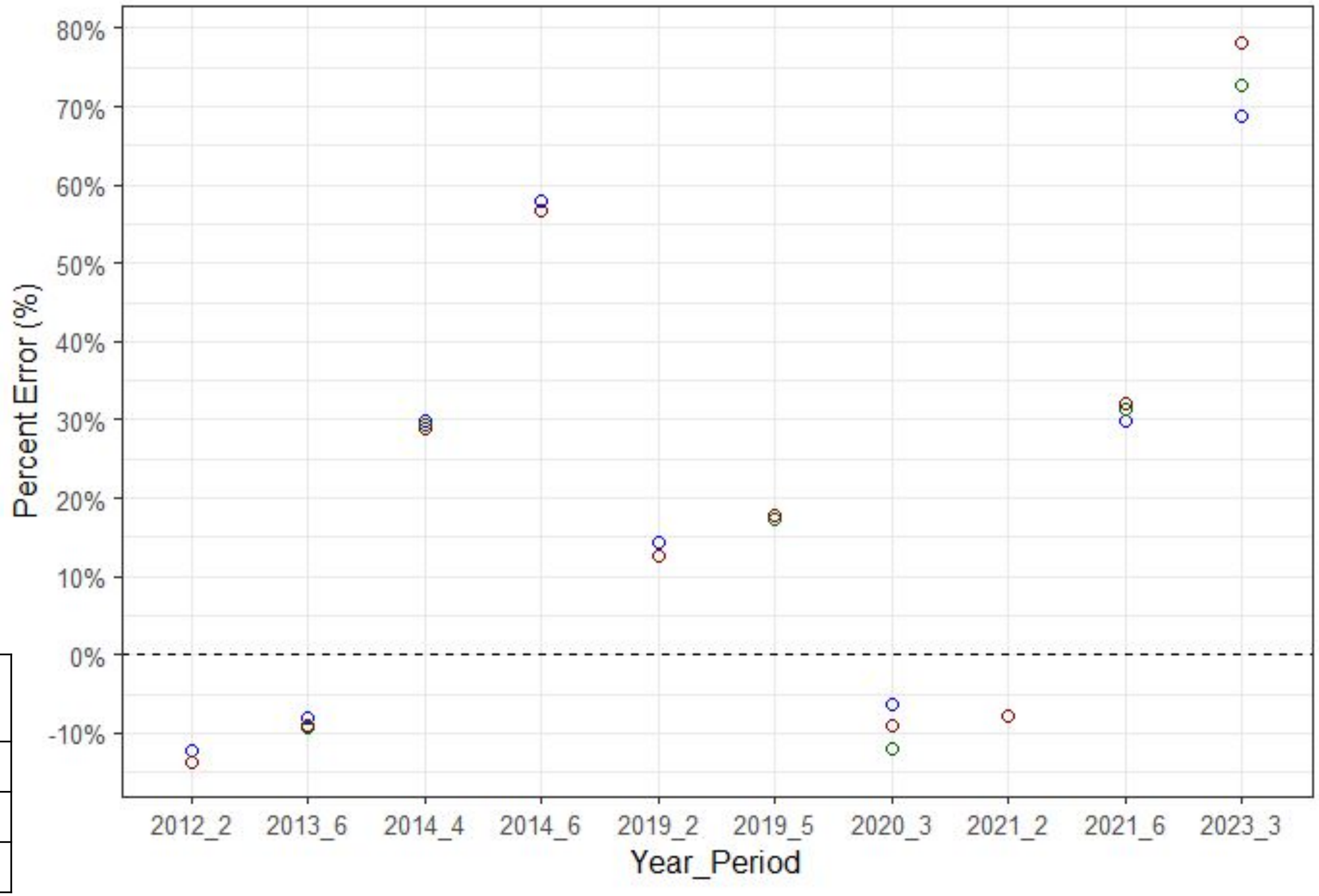
Model	Color	MPE
mod1.pfe.cov.crab2	blue	6.1493
mod1.pfe.cov	green	6.2366
mod1.pfe.cov.crab1	red	6.2431



Section 3.3.2. Tendency to Over- or Under-predict

- **Out-of-sample** tendency to under- or -overpredict
- Held out 17% of the dataset (2 years' worth) using `CreateDataPartition()`
- Made predictions of the held-out data based on the training set (excludes the held-out data)
- Underpredicted a maximum of ~12% in period 2 of 2012 and 2020

Model	Color	MPE	Mean Error
mod1.pfe.cov.crab2	blue	18.4%	3.2
mod1.pfe.cov	**green**	**17.7%**	**3.0**
mod1.pfe.cov.crab1	red	18.7%	3.1

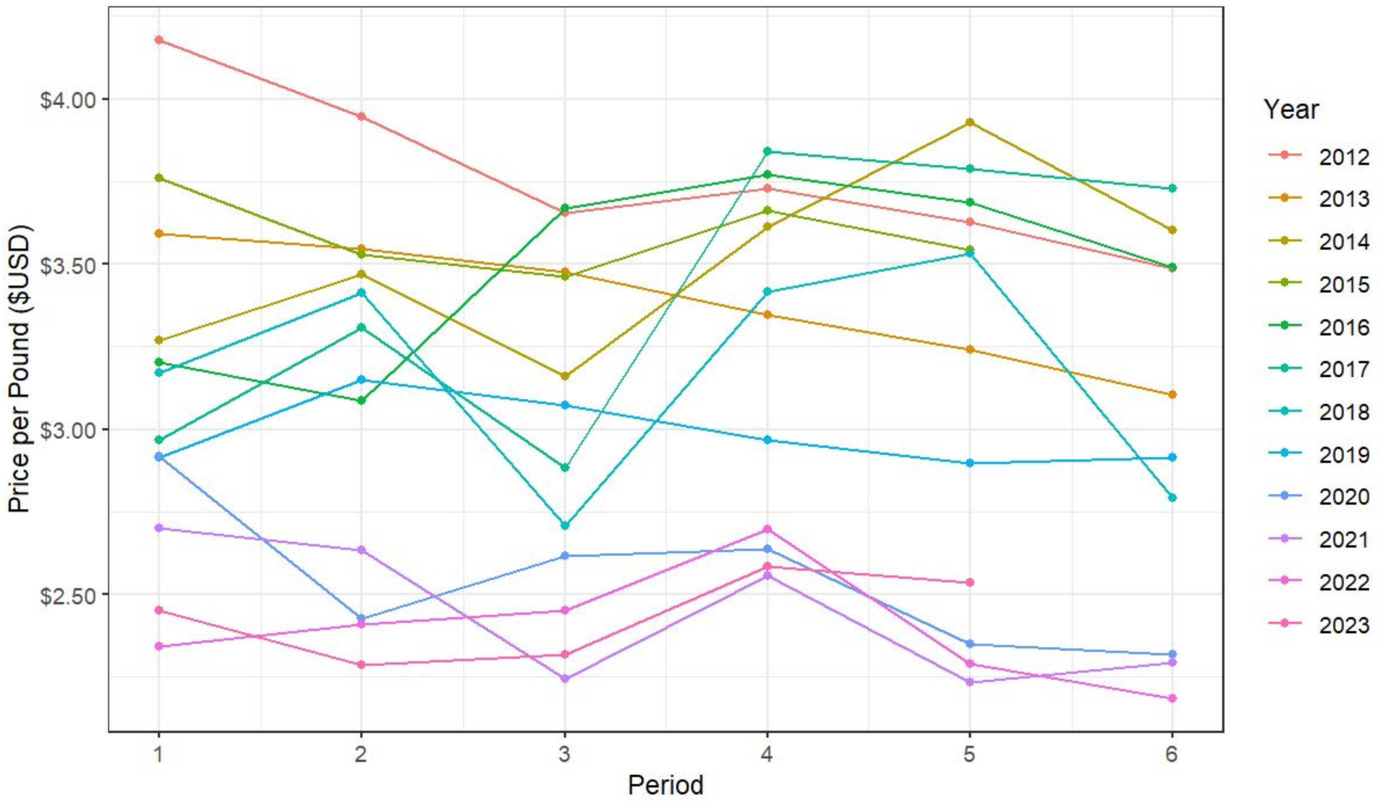


LEN

Section 3.3.3. Price Forecasting Evaluation

- Current price forecasting method: use average price by period from most similar recent years
- Ex: if predicting 2023 periods 5-6, would use period 5 and 6 average prices from years 2020-2022 (Figure)
- Greater uncertainty in price forecasting during SPEX process, but Council can adjust trip limits inseason as sablefish prices change

Average LEN sablefish price per pound ($USD) by period and year, 2012-2023



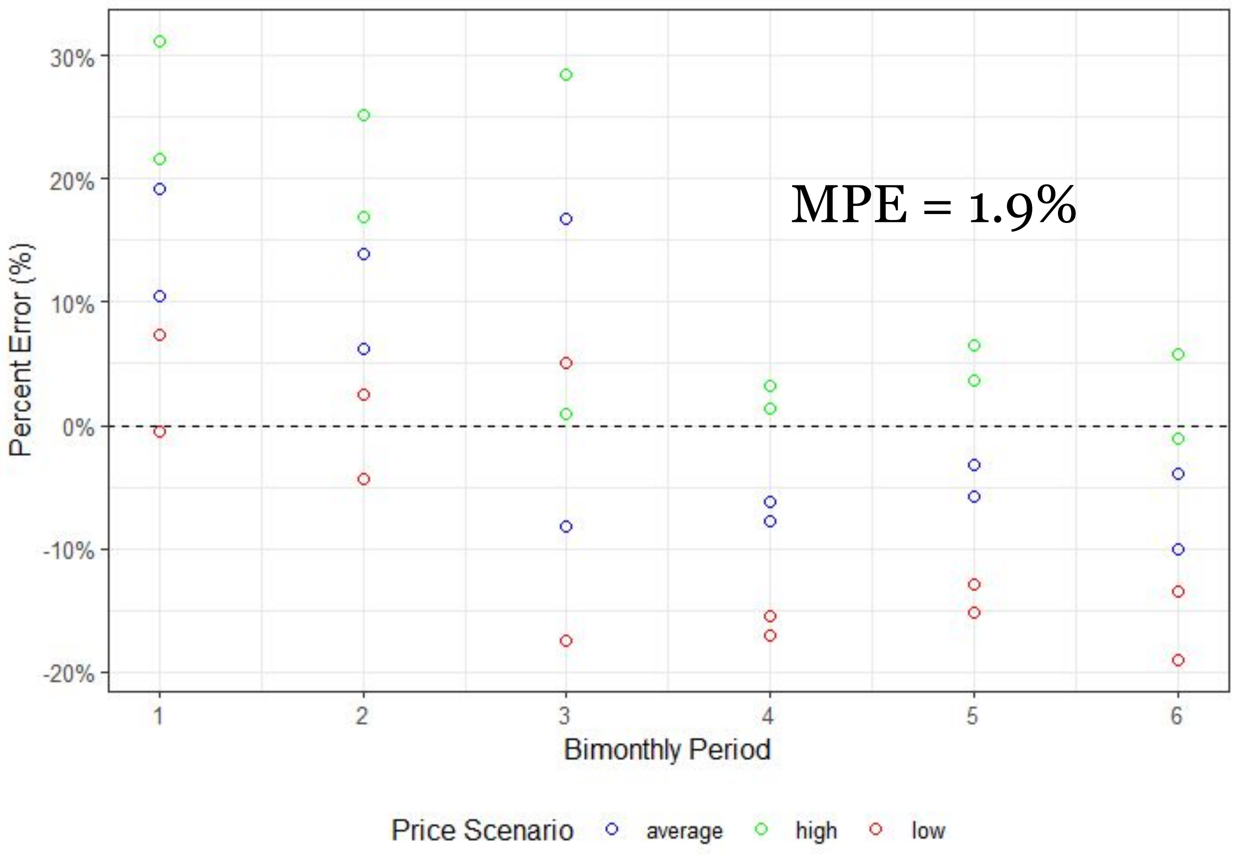
LEN

y-axis limits are not the same

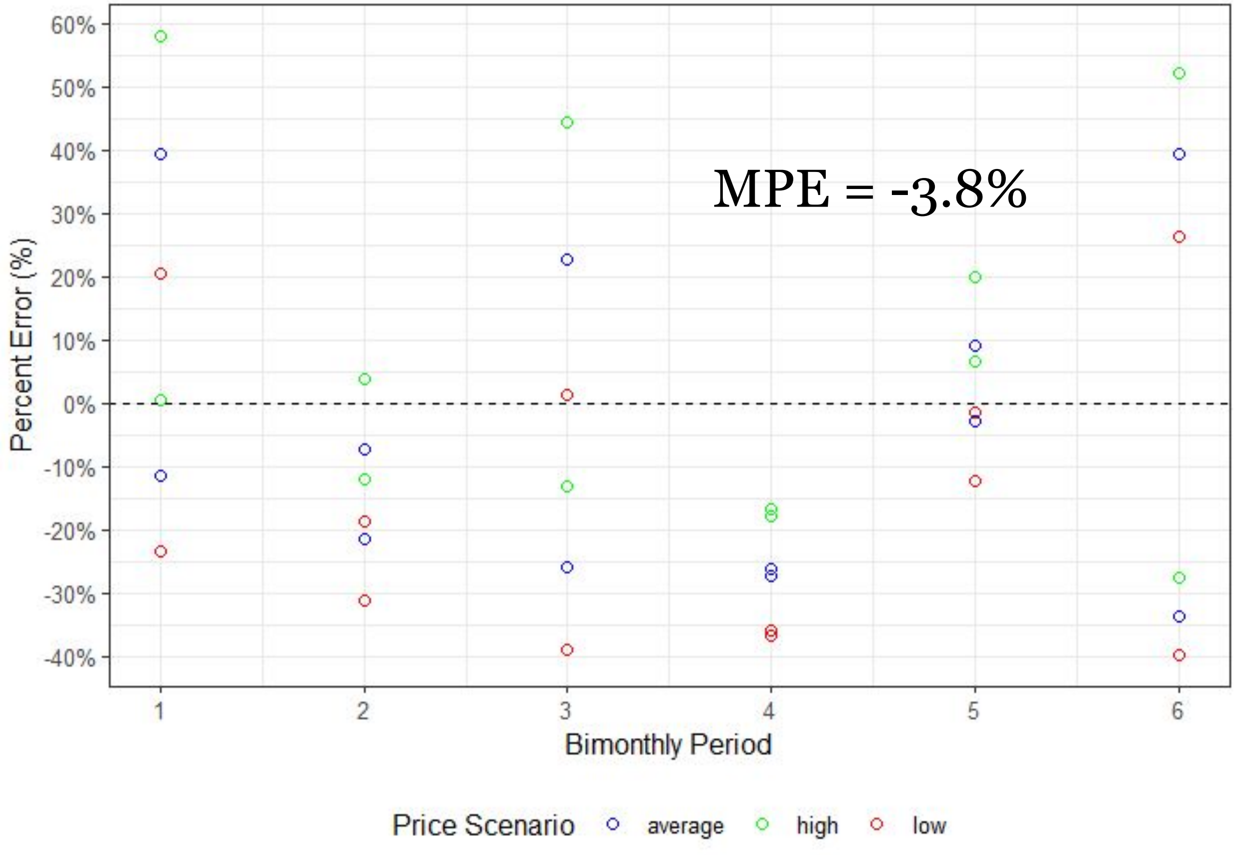
Section 3.3.3. Price Forecasting Evaluation

Using 2013-2015 prices to predict 2016 and 2017 prices as an example to test out-of-sample prediction error

Price prediction error



Total LEN landings prediction error



Section 3.4. Best Performing Models for LEN

both linear regressions:

$$\text{Average lbs. per vessel} \sim \text{bimonthly trip limit} + factor(PERIOD) + factor(COVID)$$

$$\text{Number of vessels} \sim \text{avg. sablefish price per lb.} + factor(PERIOD) + factor(COVID)$$

- Despite the models with lowest RMSE from TSCV including these, we:
  - excluded yen exchange rate from landings model due to inability to forecast
  - excluded maximum Dungeness crab prices from participation model due to wide variability in maximum prices ($4.81 - $29.38 per pound in dataset)
  - excluded average Dungeness crab prices from participation model because very little difference with vs. without and would need to be forecasted

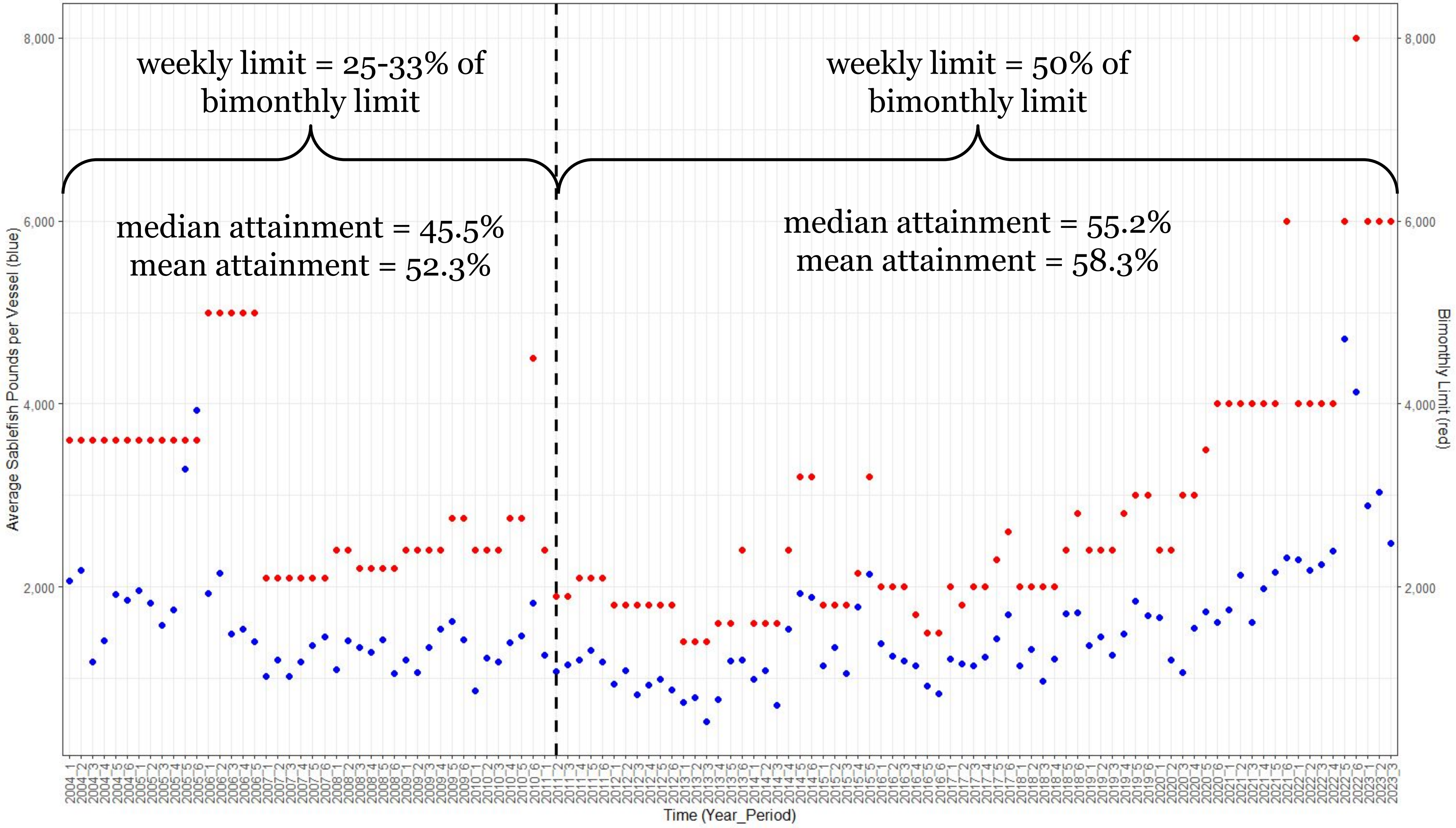
Section 3.5. Risk to the Sablefish North ACL

Year	Sum annual prediction error (mt)	Under- or over-prediction	Actual LEN Target attainment	Hypothetical LEN Target attainment	LE Share (mt)	Total prediction error as % of LE Share
2017	-32.3	under	98%	111%	4,252	0.76%
2018	-30.9	under	84%	96%	4,434	0.70%
2019	25.6	over	65%	56%	4,537	0.56%
2020	5.1	over	56%	54%	4,636	0.11%
2021	3.3	over	51%	50%	5,586	0.06%
2022	-78.7	under	92%	116%	5,320	1.48%

- These are based on MAE results of the TSCV using the best performing LEN models (previous slide)
- Shaded rows are where the model under-predicted for that year
- Council may want to consider precaution when the model predicts attainment >90%
- Total prediction error as a percent of the LE share is less than 2%

OAN

Section 4.1. Evaluating Data Prior to 2011

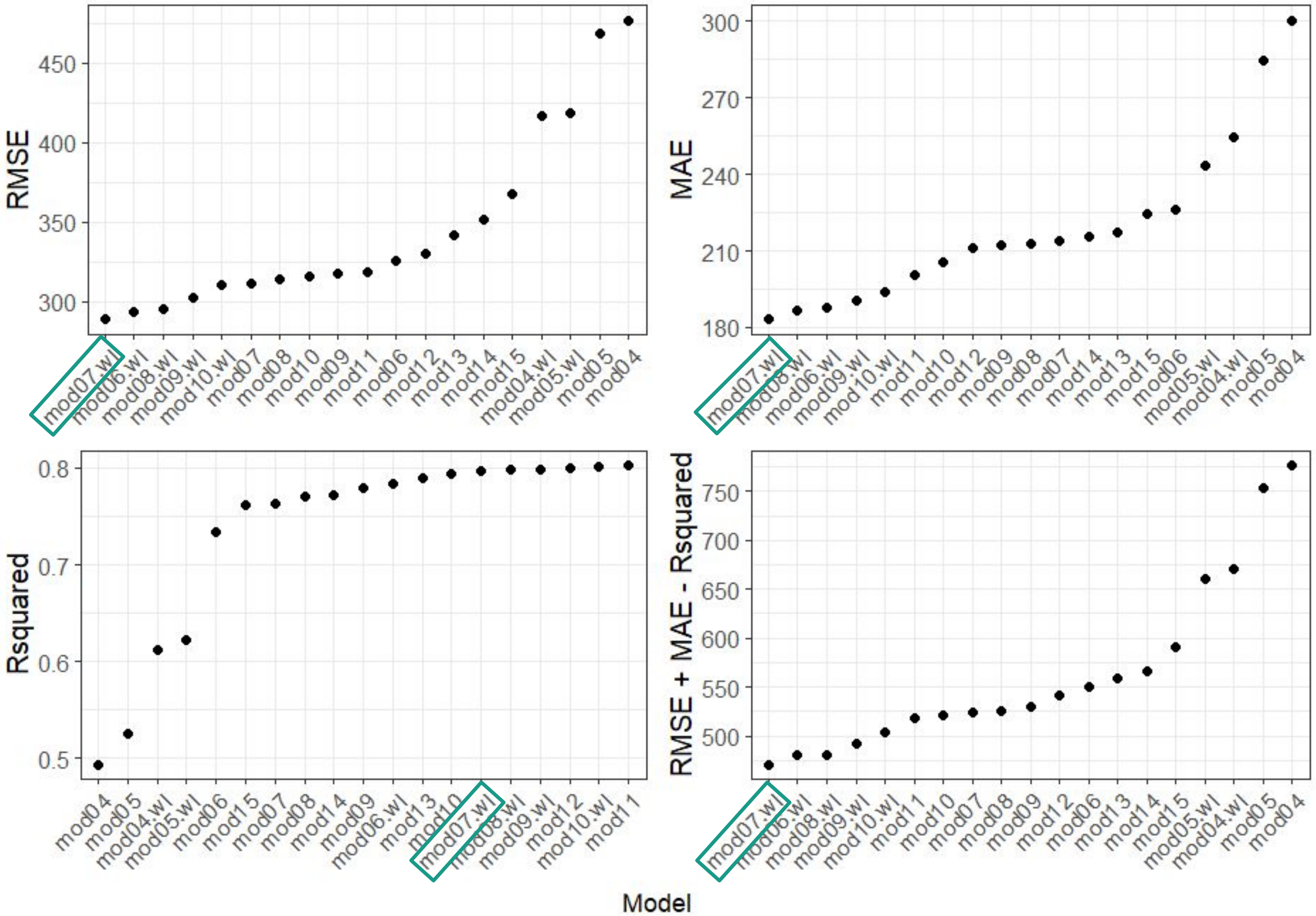


OAN

**y-axis limits are not the same*

Section 4.1. Evaluating Data Prior to 2011

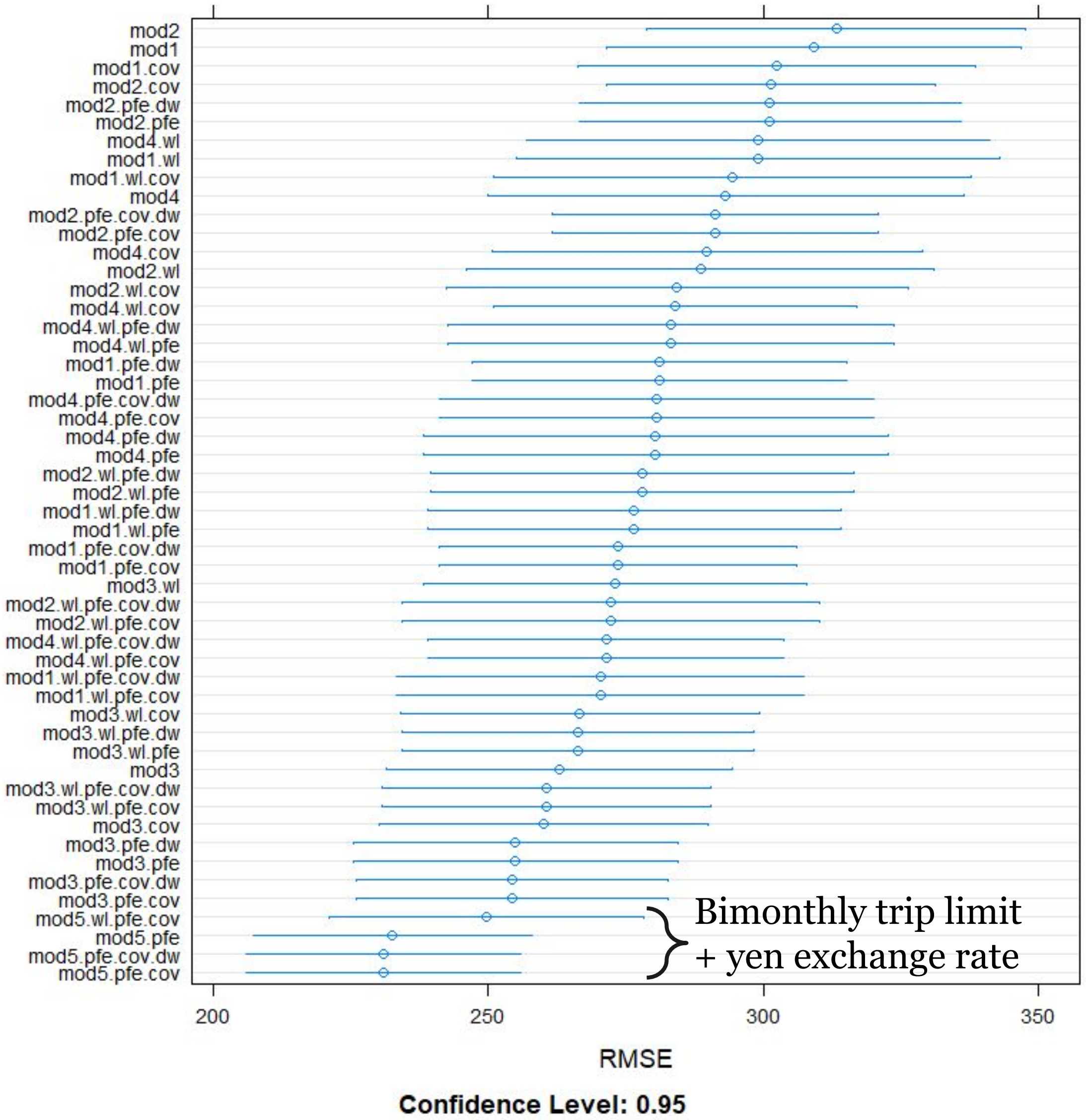
- Leave-one-out cross validation
- All models predict number of vessels using status quo variable (avg. sablefish price)
- Order of x-axis varies by plot
- Model used for all further analysis: 2007-2023 (N = 95)
- Included models with and without weekly limit factor variable for comparison



.wl = weekly limit factor variable

OAN - pounds per vessel

Section 4.2.1. Time Series Cross Validation Results



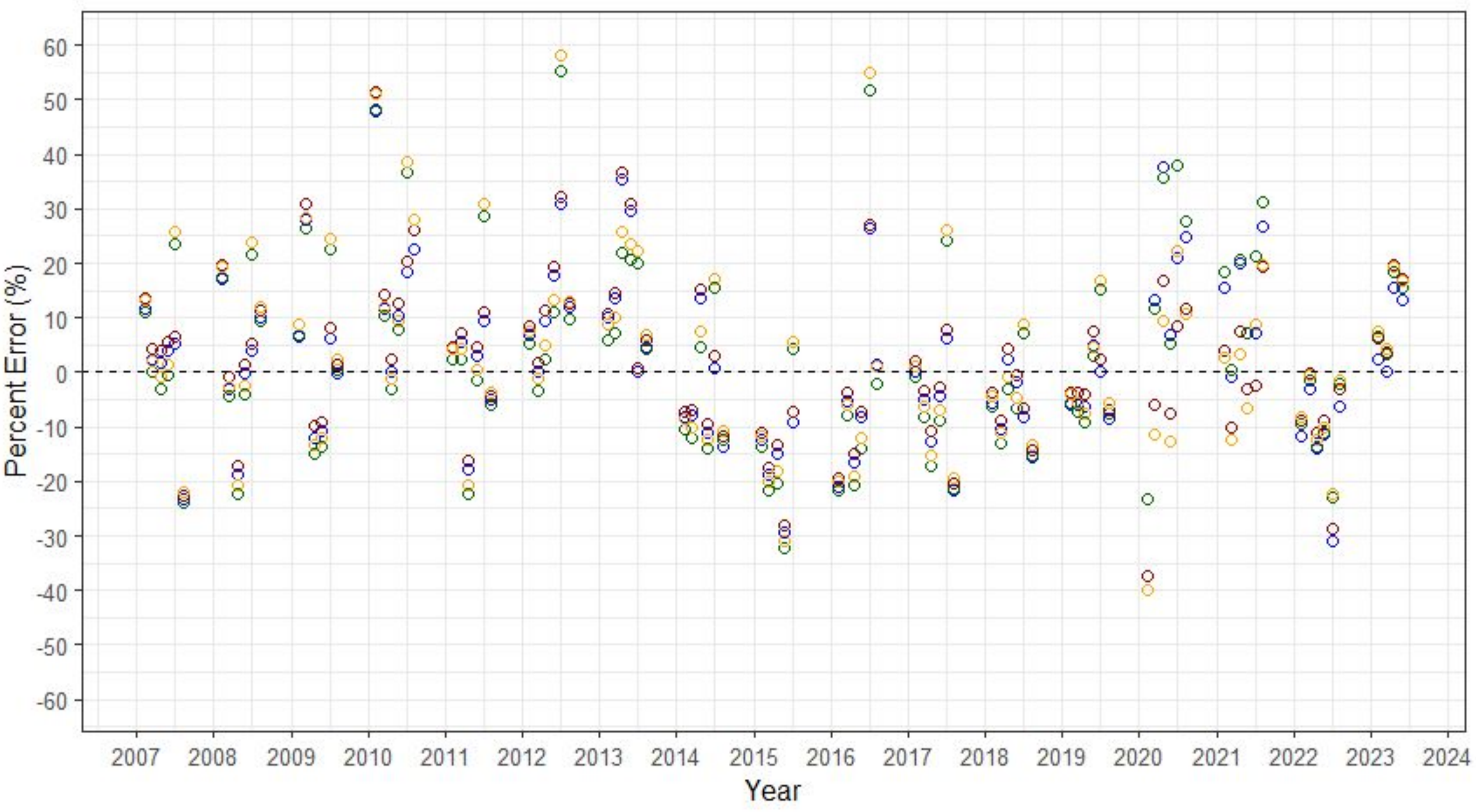
Model	MAE	Annual error estimate (mt)	% of OAN Target	% of OA Share
mod3.pfe	202.5	47.5	6.9%	6.6%
mod3.pfe.cov	202.8	47.6	6.9%	6.7%
mod3.pfe.cov.dw	202.8	47.6	6.9%	6.7%
mod3.pfe.dw	202.5	47.5	6.9%	6.6%

mod1	weekly limit + bimonthly limit
mod2	weekly limit
mod3	bimonthly limit
mod4	bimonthly limit + Alaska sablefish TAC
mod5	bimonthly limit + yen exchange rate
.wl	pre-2011 fixed effects
.pfe	period fixed effects
.cov	2020 and 2021 as fixed effects (COVID)
.dw	data weights

Section 4.2.2. Tendency to Under- or Over-predict

- top four models from TSCV, excluding models with yen exchange rate (inability to forecast)
- mod3.pfe = lowest **in-sample** MPE
  - underpredicted a maximum of ~30% in 2022

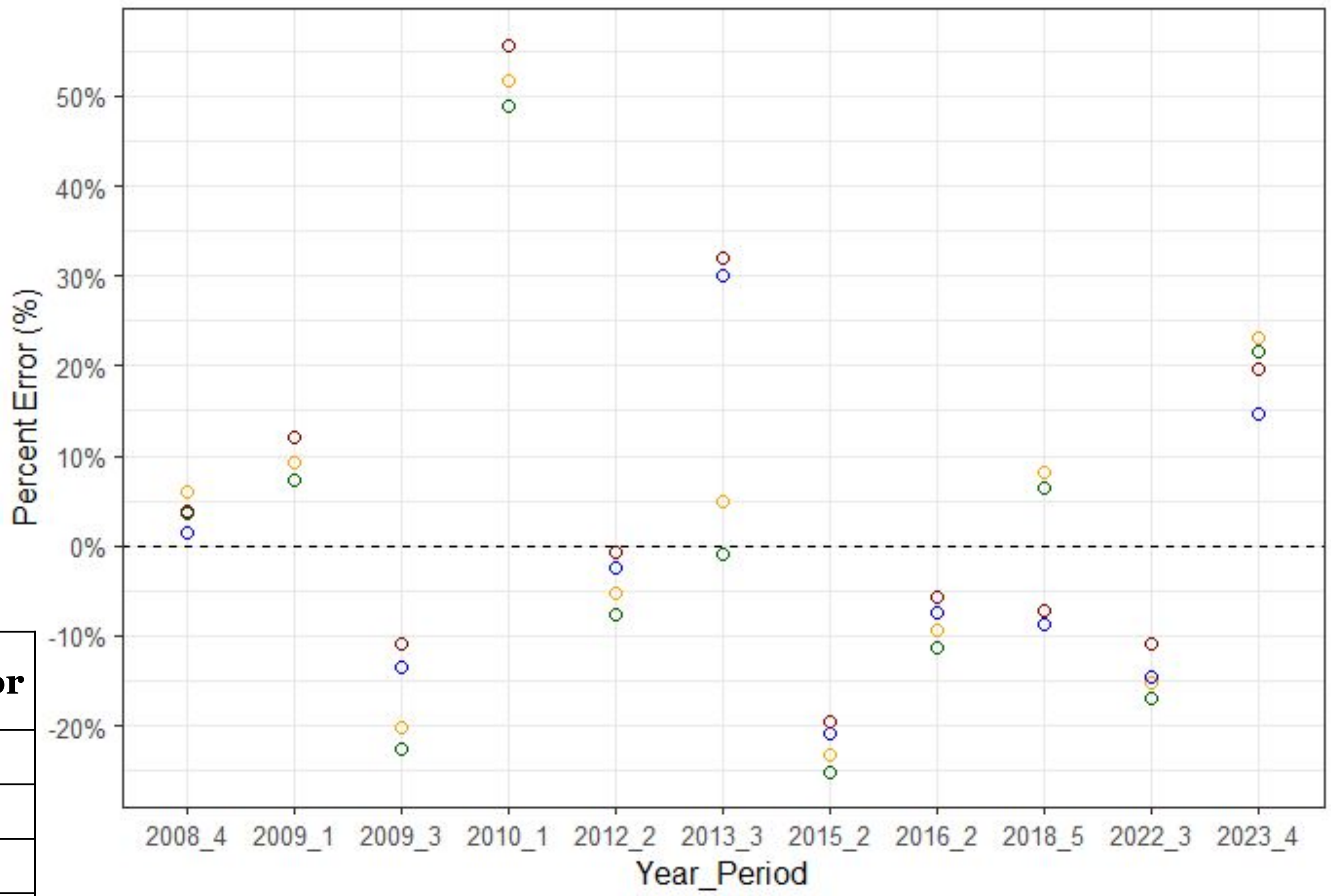
Model	Color	MPE
mod3.pfe	blue	1.8821
mod3.pfe.cov	red	1.9065
mod3.pfe.cov.dw	orange	2.8138
mod3.pfe.dw	green	3.1615



Section 4.2.2. Tendency to Under- or Over-predict

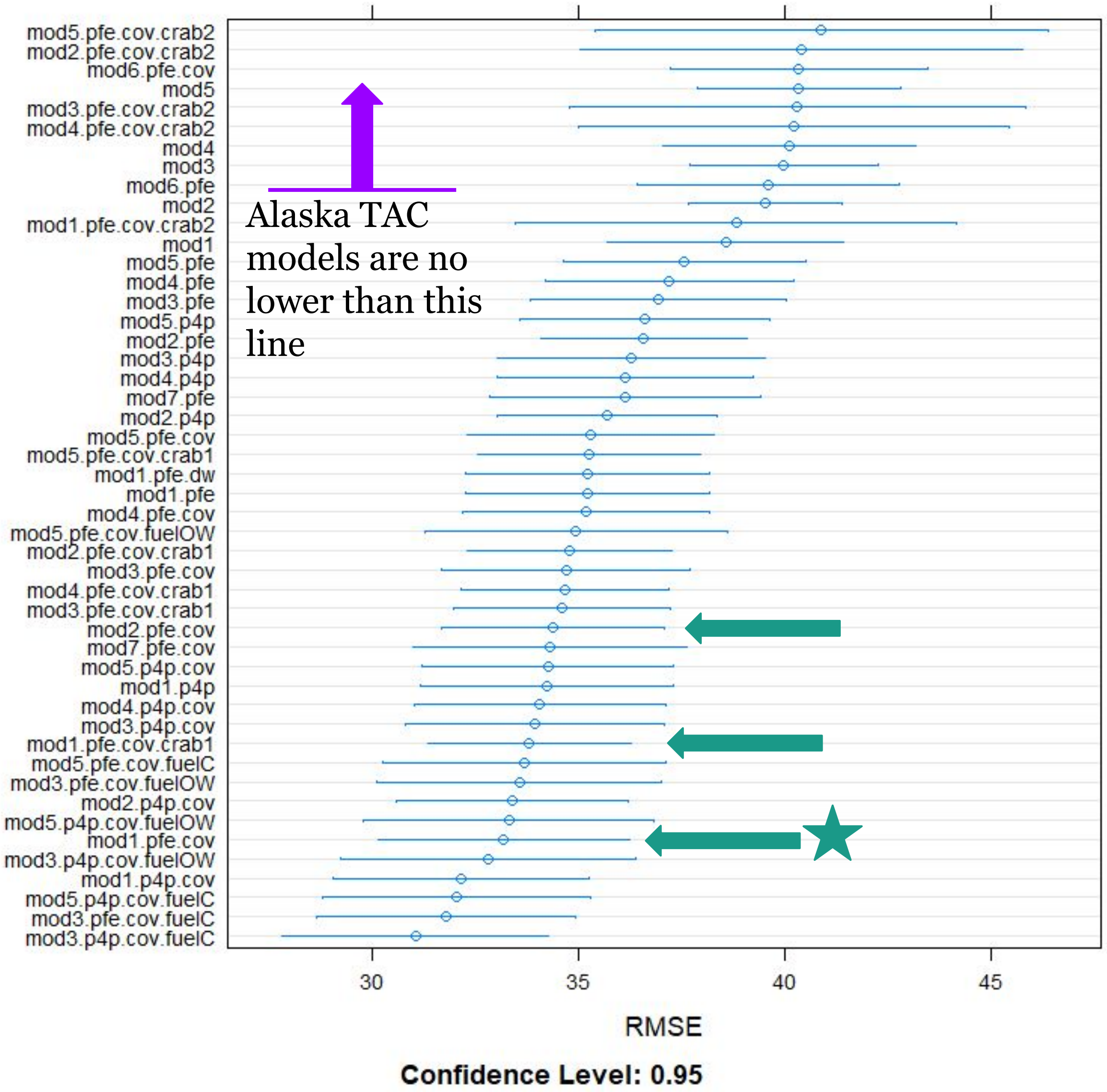
- **Out-of-sample** tendency to under- or -overpredict
- Held out 12% of the dataset (2 years' worth) using `CreateDataPartition()`
- Made predictions of the held-out data based on the training set (excludes the held-out data)
- Underpredicted a maximum of ~30% in period 2 of 2017

Model	Color	MPE	Mean Error
mod3.pfe	**blue**	**3.6%**	**6.9**
mod3.pfe.cov	red	6.2%	47.3
mod3.pfe.cov.dw	orange	2.7%	29.5
mod3.pfe.dw	green	0.3%	-0.4



OAN - number of vessels

Section 4.3.1. Time Series Cross Validation Results



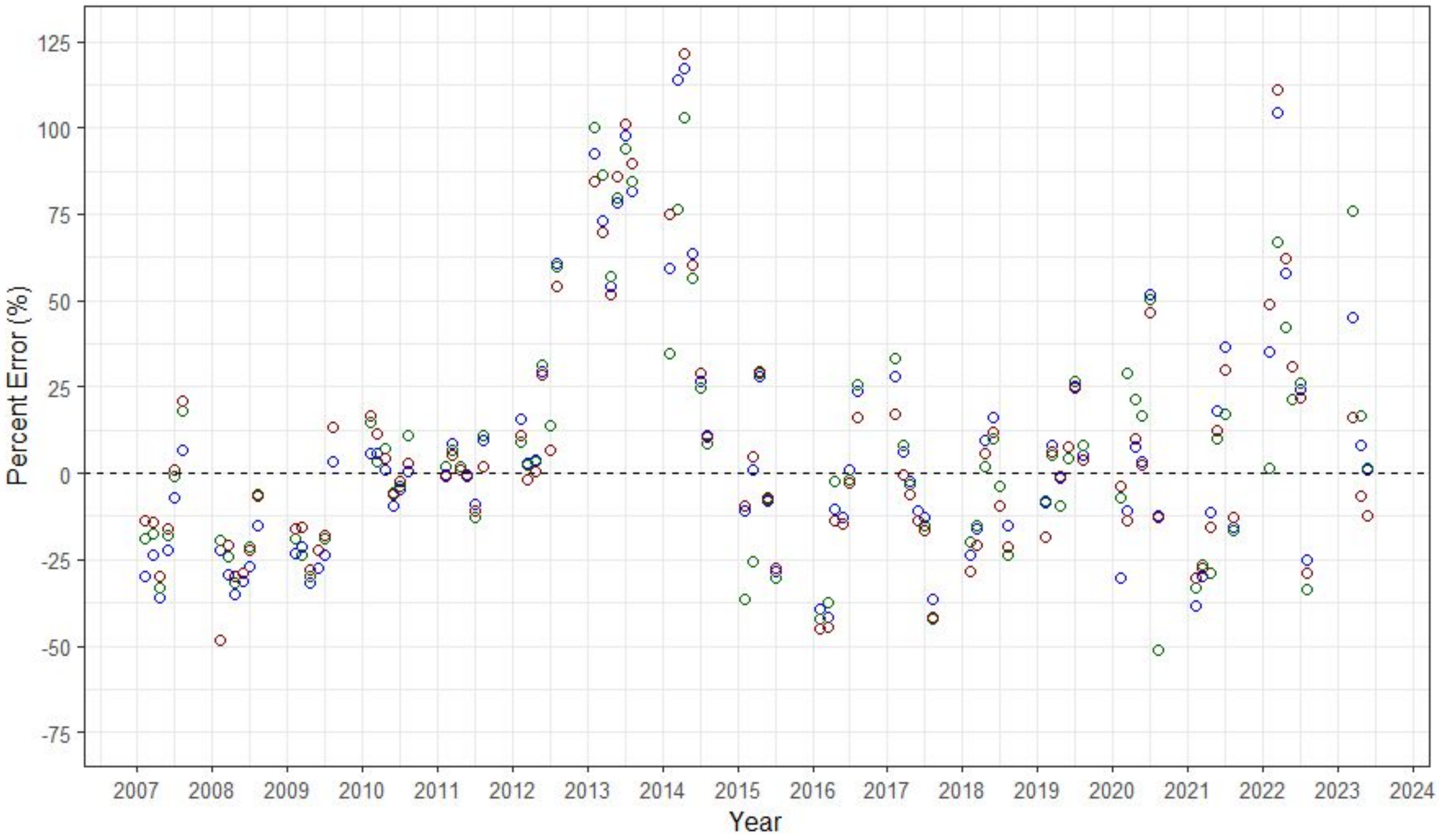
Model	MAE	Annual error estimate (mt)	% of OAN Target	% of OA Share
mod1.pfe.cov.crab1	25.7	194.1	28.2%	27.2%
mod2.pfe.cov	21.6	163.1	23.7%	22.8%
mod1.pfe.cov ★	**25.3**	**191.1**	**27.8%**	**26.7%**

mod1	avg. sablefish price
mod2	avg. sablefish price + max. sablefish price
mod3	max. sablefish price
mod4	median sablefish price
mod5	median sablefish price + max. sablefish price
mod6	avg. sablefish price + Alaska sablefish TAC
mod7	avg. sablefish price + yen exchange rate
.p4p	Period 4 peak adjuster fixed effect
.pfe	period fixed effects
.cov	2020 and 2021 as fixed effects (COVID)
.crab1	+ avg. crab price
.crab2	+ max. crab price
.dw	data weights

Section 4.3.2. Tendency to Over- or Under-predict

- top three models from previous slide
- strong overprediction in 2013-2014 as well as 2022
- mod1.pfe.cov.crab1 = lowest **in-sample** MPE
    - underpredicted a maximum of ~50% in 2022
- mod1.pfe.cov underpredicted a maximum of ~40% in 2016

Model	Color	MPE
mod1.pfe.cov.crab1	green	10.2239
mod2.pfe.cov	red	10.3481
mod1.pfe.cov	blue	10.4308

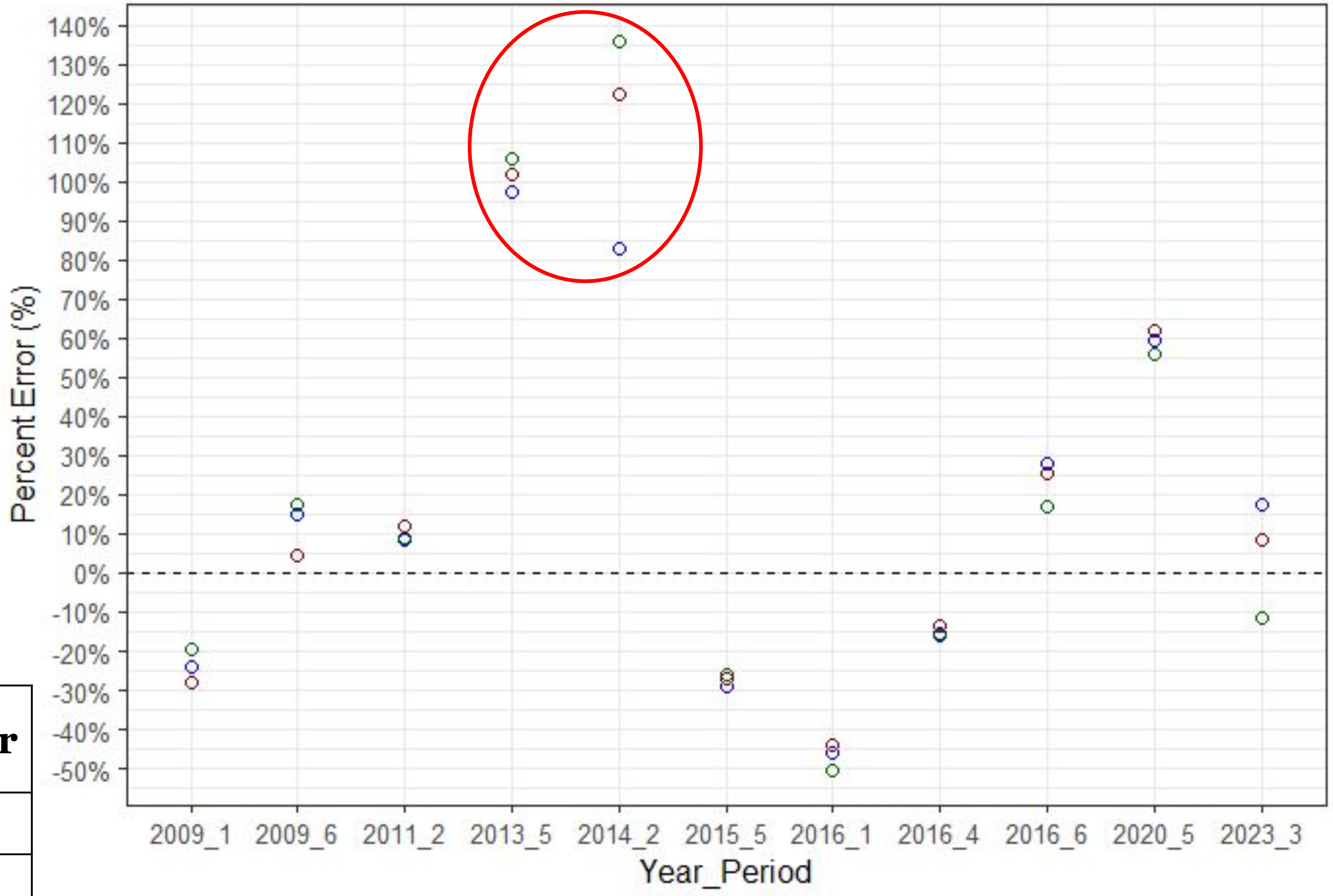


OAN - number of vessels

Section 4.3.2. Tendency to Over- or Under-predict

- **Out-of-sample** tendency to under- or -overpredict
- Held out 12% of the dataset (2 years' worth) using `CreateDataPartition()`
- Made predictions of the held-out data based on the training set (excludes the held-out data)
- mod1.pfe.cov underpredicted a maximum of ~40% in period 1 of 2016

Model	Color	MPE	Mean Error
mod1.pfe.cov.crab1	blue	17.7%	3.0
mod2.pfe.cov	green	19.8%	2.5
mod1.pfe.cov	**red**	**20.4%**	**3.5**



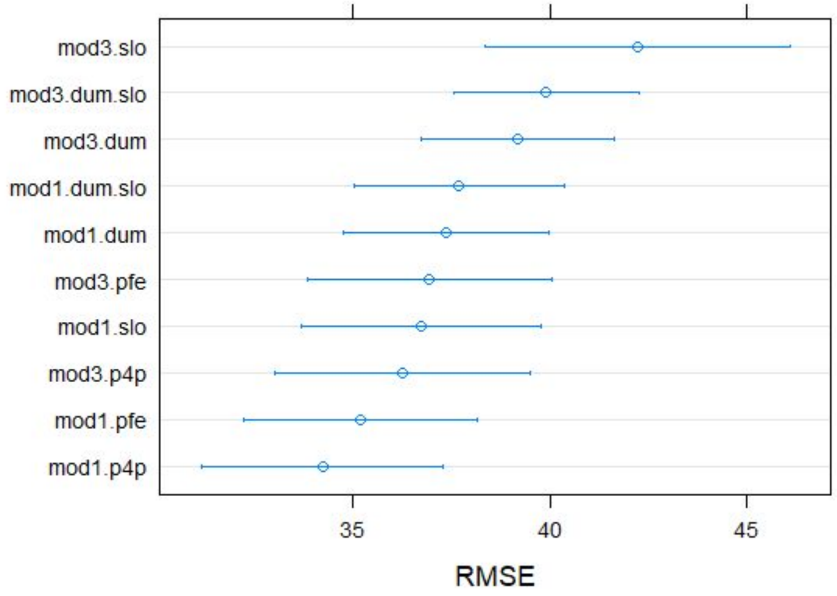
Section 4.3.2. Tendency to Over- or Under-predict
Crab Season Start Date as Predictor

- Dungeness crab season start dates vary by state and year-to-year
  - Typically sometime late in the year (~December) to early Spring of the following year (~February), with the season ending around September
  - Therefore, not feasible to include in biennial management projections
- Inseason: March is the earliest the Council could consider inseason action based on a model that uses Dungeness crab season start dates as a predictor variable
  - However, likely greatest influence would be on period 1 around the time season dates are actively being set
  - Dungeness crab markets/prices likely have a stronger influence on OAN participation in periods 2-6

Section 4.3.3. Comparison of Period-specific Effects

```
> summary(oan_mod_slopes1)

Call:
lm(formula = VES_NUM ~ ADJ_PRICE * factor(PERIOD), data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-60.057 -24.067   0.912  17.323  64.799 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 44.910     45.603   0.985    0.328
ADJ_PRICE                    3.293     15.048   0.219    0.827
factor(PERIOD)2            -10.588     70.290  -0.151    0.881
factor(PERIOD)3              6.740     62.665   0.108    0.915
factor(PERIOD)4            -38.175     59.825  -0.638    0.525
factor(PERIOD)5            -58.563     64.262  -0.911    0.365
factor(PERIOD)6            -20.477     63.495  -0.323    0.748
ADJ_PRICE:factor(PERIOD)2    9.490     22.911   0.414    0.680
ADJ_PRICE:factor(PERIOD)3   14.588     19.880   0.734    0.465
ADJ_PRICE:factor(PERIOD)4   28.894     18.871   1.531    0.130
ADJ_PRICE:factor(PERIOD)5   32.166     20.306   1.584    0.117
ADJ_PRICE:factor(PERIOD)6   11.003     20.561   0.535    0.594

Residual standard error: 30.31 on 83 degrees of freedom
Multiple R-squared:  0.4609,    Adjusted R-squared:  0.3895 
F-statistic: 6.452 on 11 and 83 DF,  p-value: 0.0000001235
```

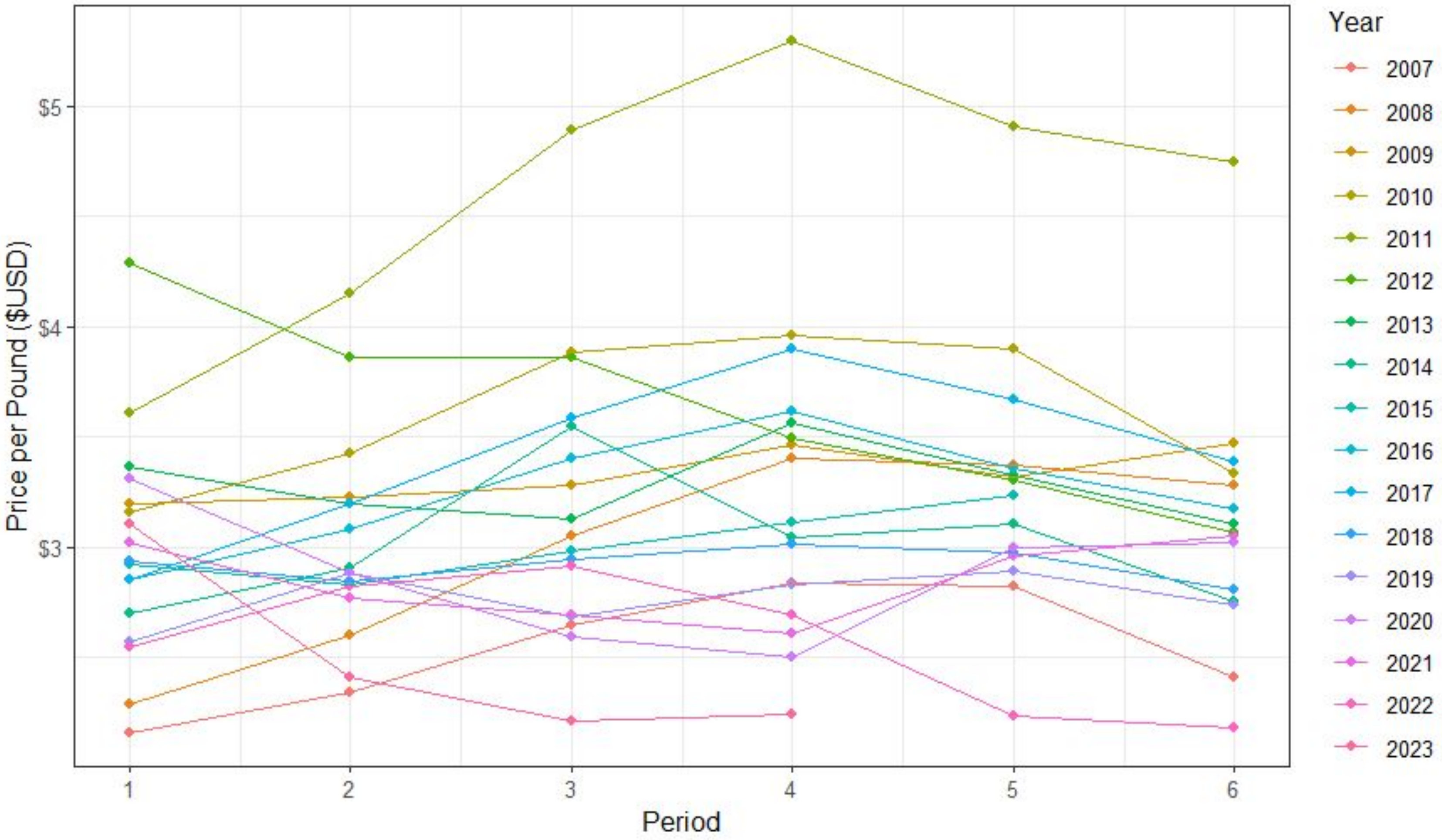
mod1	avg. sablefish price
mod3	max. sablefish price
.pfe	period fixed effects
.p4p	period 4 peak adjuster fixed effects
.dum	Dummy variable for period 4 Period 4 = 1 Periods 1-3 and 5-6 = 0
.slo	interaction between period and price variables



Confidence Level: 0.95

Section 4.3.4. Price Forecasting Evaluation

Average OAN sablefish price per pound ($USD) by period and year, 2007-2023



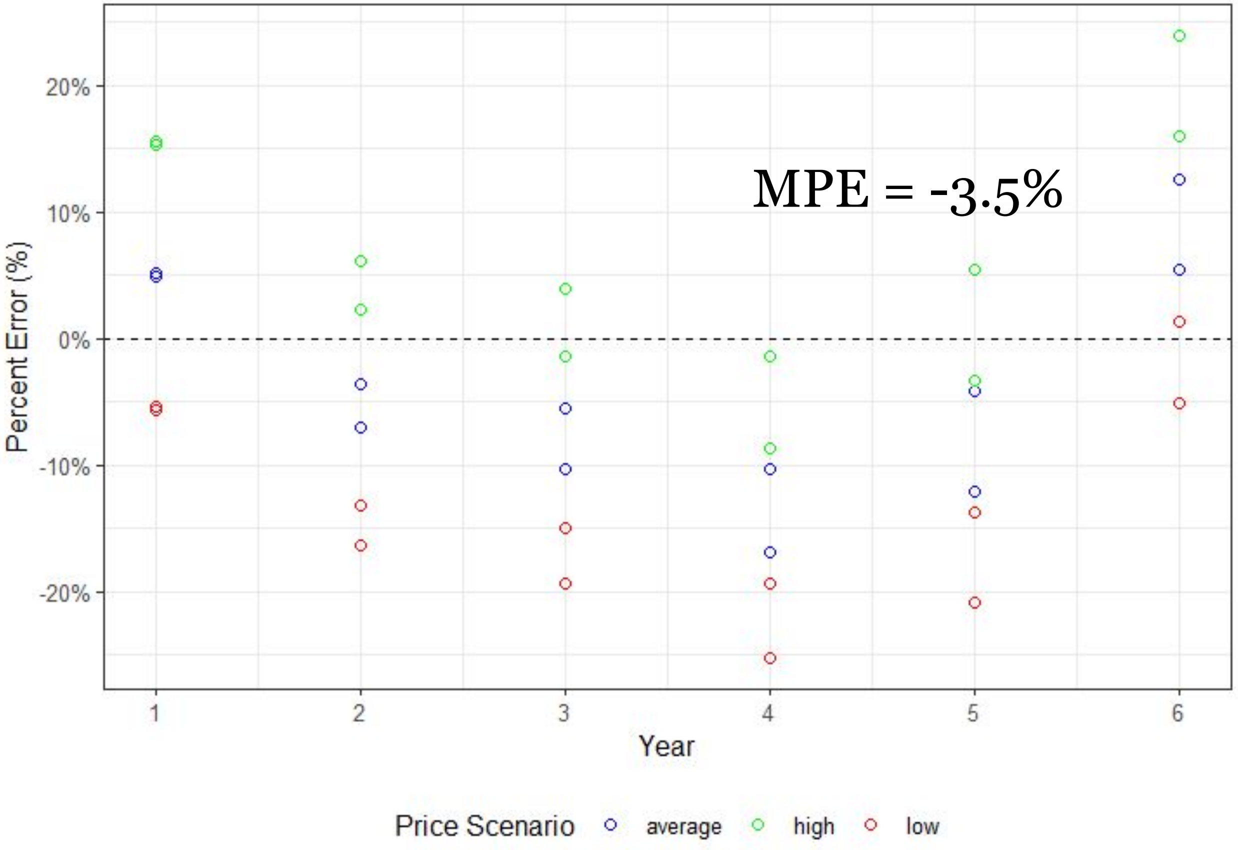
OAN

y-axis limits are not the same

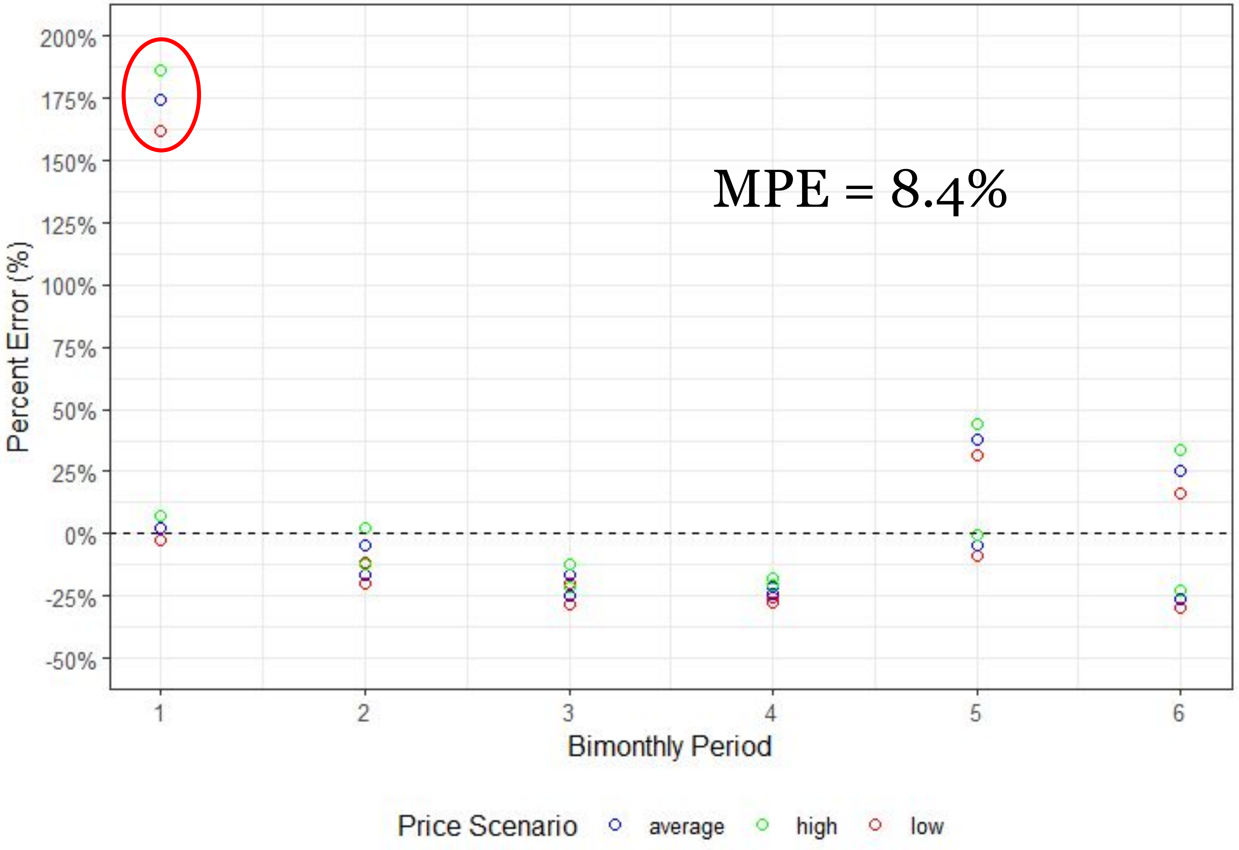
Section 4.3.4. Price Forecasting Evaluation

Using 2013-2015 prices to predict 2016 and 2017 prices as an example to test out-of-sample prediction error

Price prediction error



Total OAN landings prediction error



Section 4.4. Best Performing Models for OAN

both linear regressions:

$$Average\ lbs.\ per\ vessel \sim bimonthly\ trip\ limit\ +\ factor(PERIOD)$$

$$Number\ of\ vessels \sim avg.\ sablefish\ price\ per\ lb.\ +\ factor(PERIOD)\ +\ factor(COVID)$$

- Despite the models with lowest RMSE from TSCV including these, we:
  - excluded yen exchange rate from landings model due to inability to forecast
  - excluded fuel prices from participation model due to inability to forecast
  - excluded average Dungeness crab prices from participation model because small difference with vs. without and would need to be forecasted
  - excluded period 4 peak adjuster from the participation model due to subcommittee recommendation and arbitrariness compared to period fixed effects which are also effective

OAN

Section 4.5. Risk to the Sablefish North ACL

Year	Sum annual prediction error (mt)	Under- or over-prediction	Actual OAN Target attainment	Hypothetical[a/] OAN Target attainment	OA Share (mt)	Total prediction error as % of OA Share
2017	-69.9	under	91%	107%	441	15.8%
2018	-38.7	under	76%	84%	460	8.4%
2019	16.2	over	73%	69%	471	3.4%
2020	36.3	over	37%	29%	481	7.6%
2021	20.6	over	44%	41%	580	3.6%
2022	15.4	over	102%	99%	552	2.8%

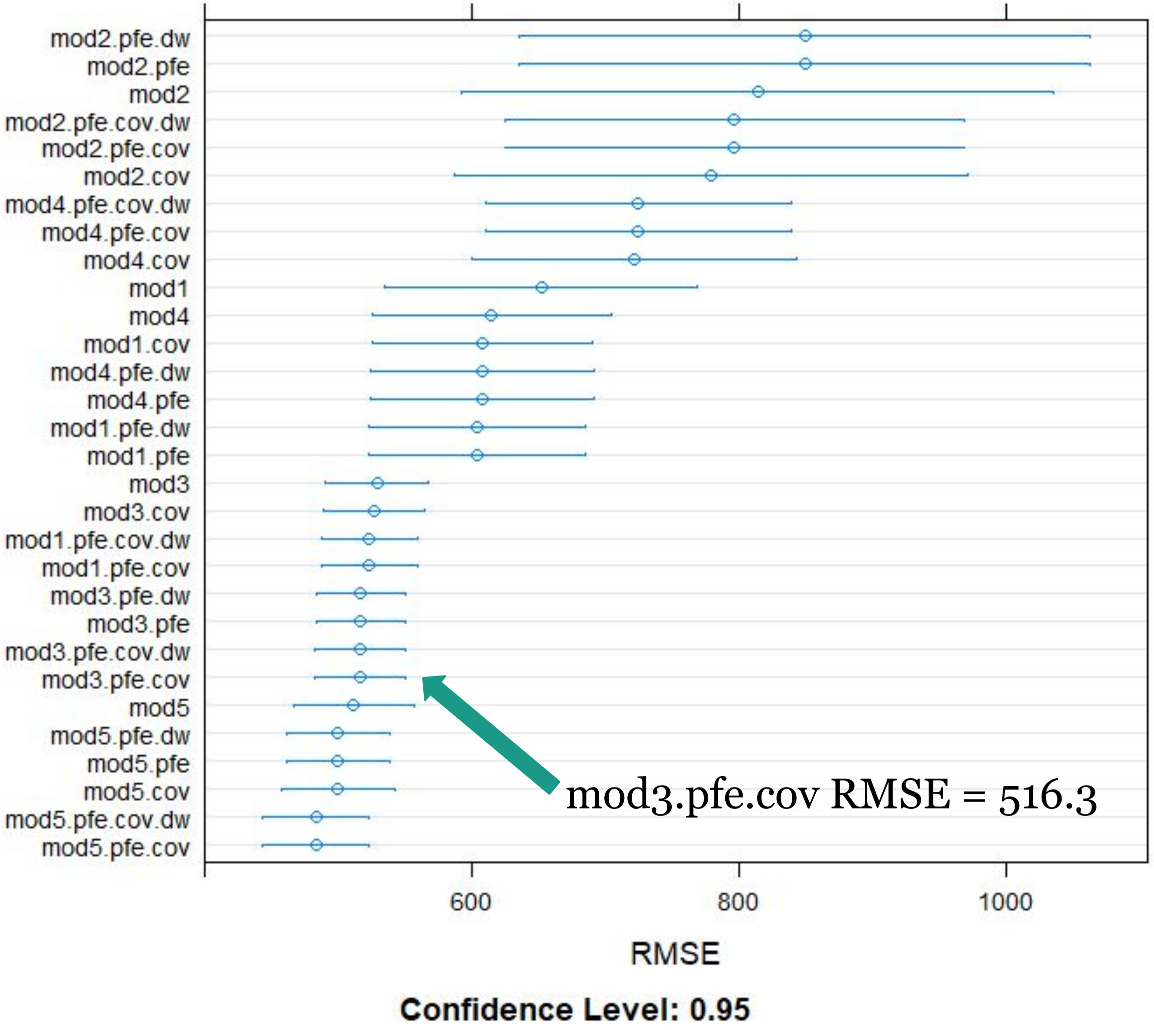
- Shaded rows are where the model under-predicted for that year
- Total prediction error as a percent of the OA share is approximately 3-16% but the OA share makes up 9.4% of the sablefish north commercial HG

Questions?

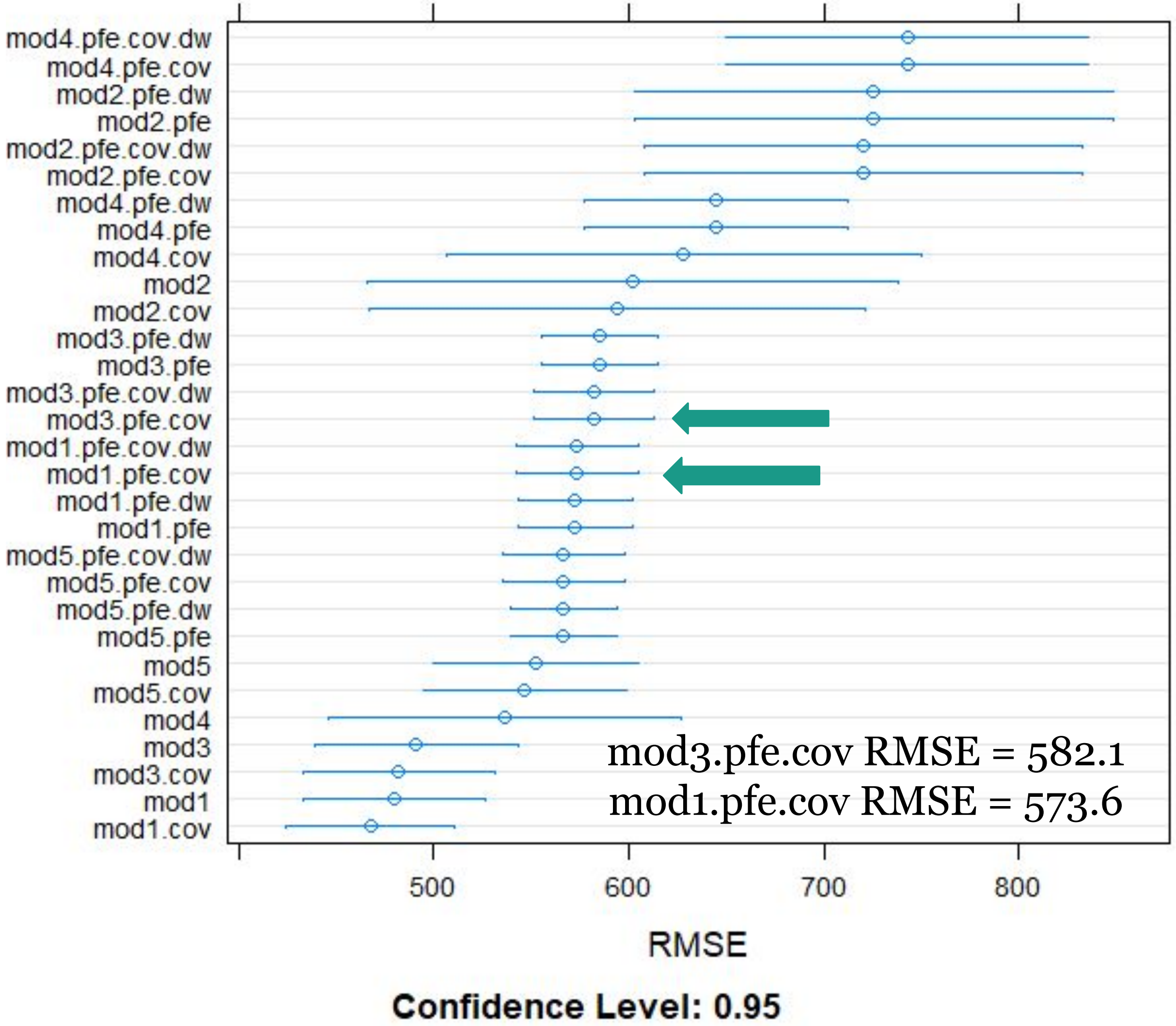
Extra Slides

LEN - pounds per vessel

2012-2023

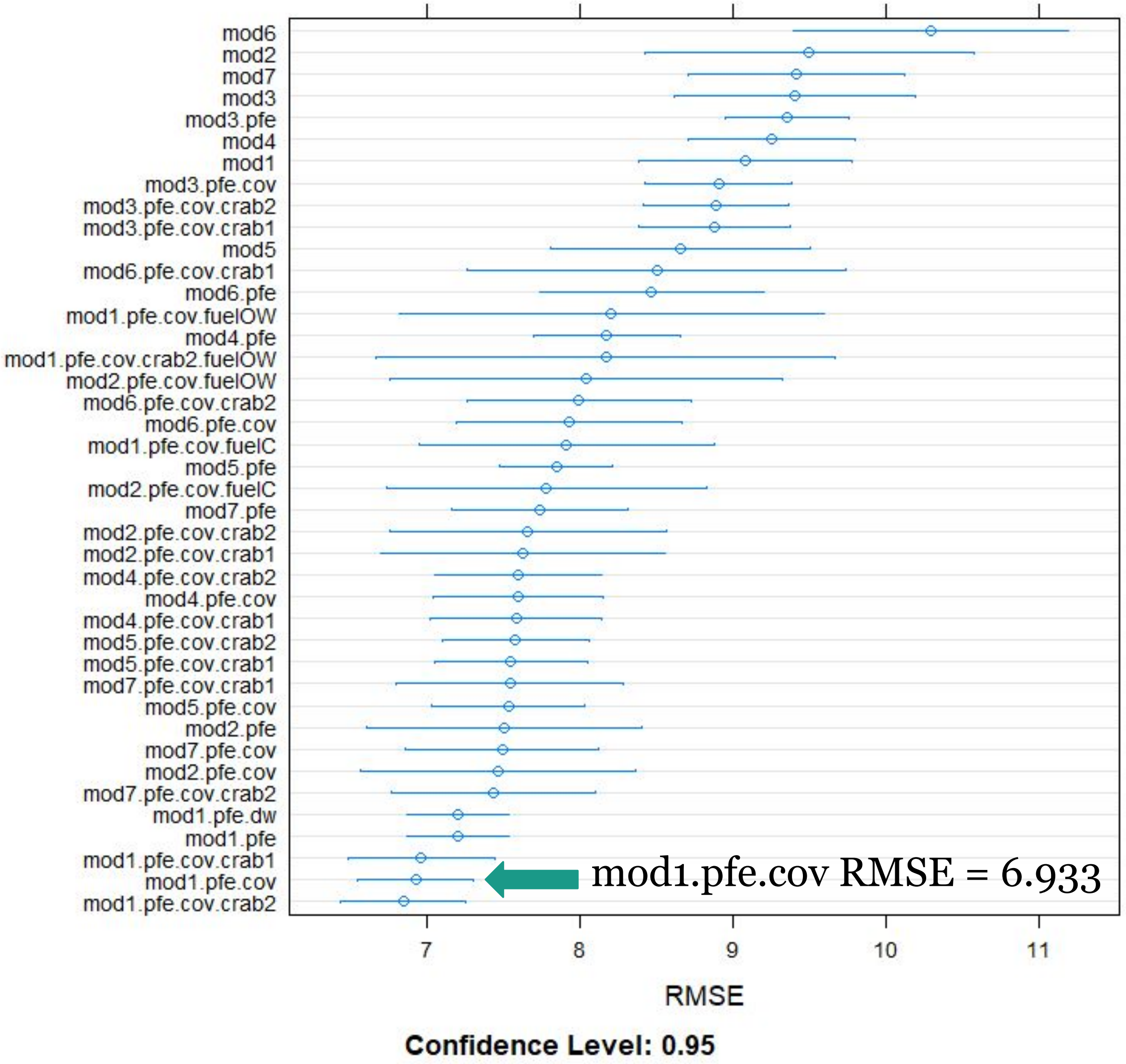


2010-2023

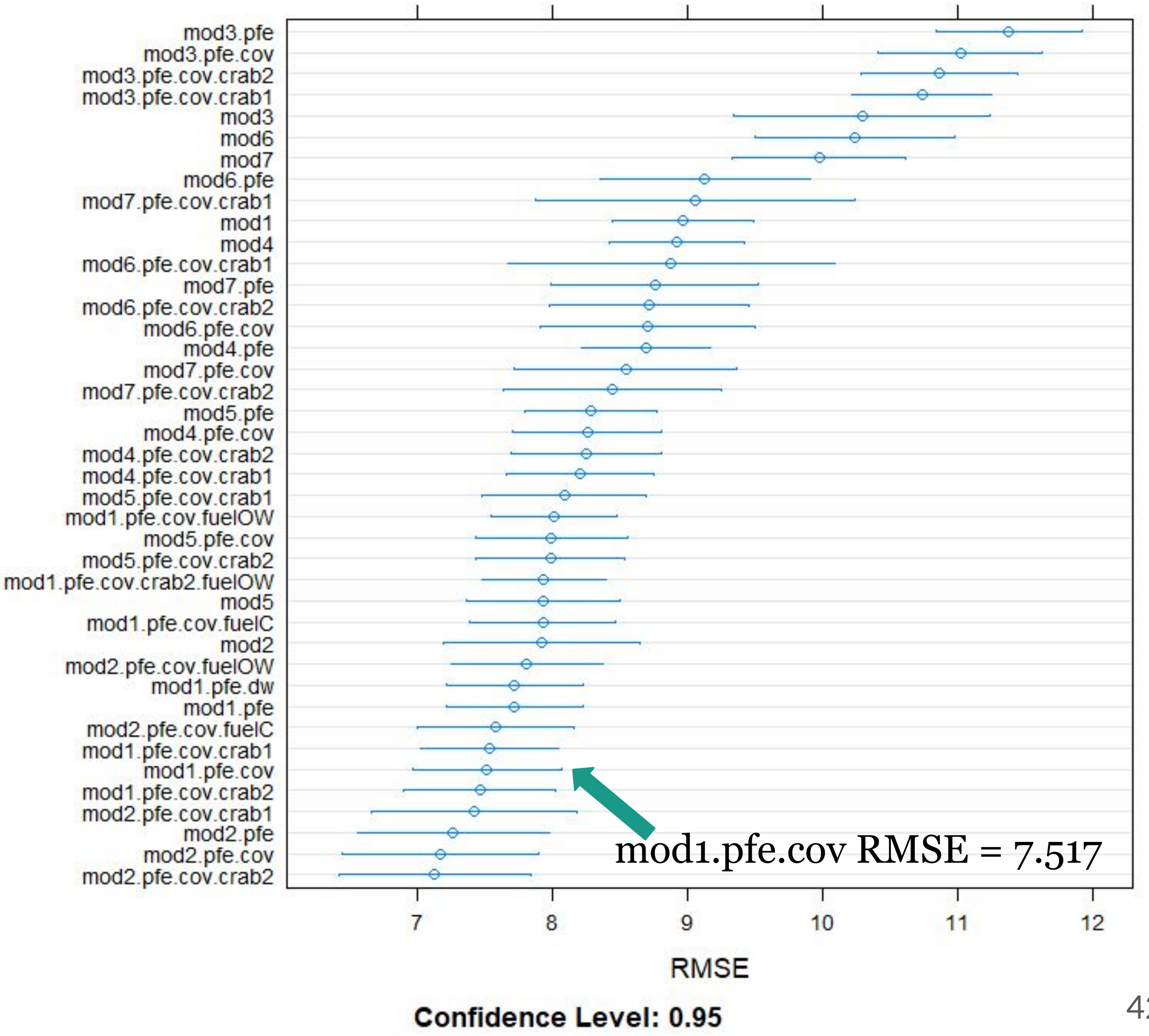


LEN - number of vessels

2012-2023

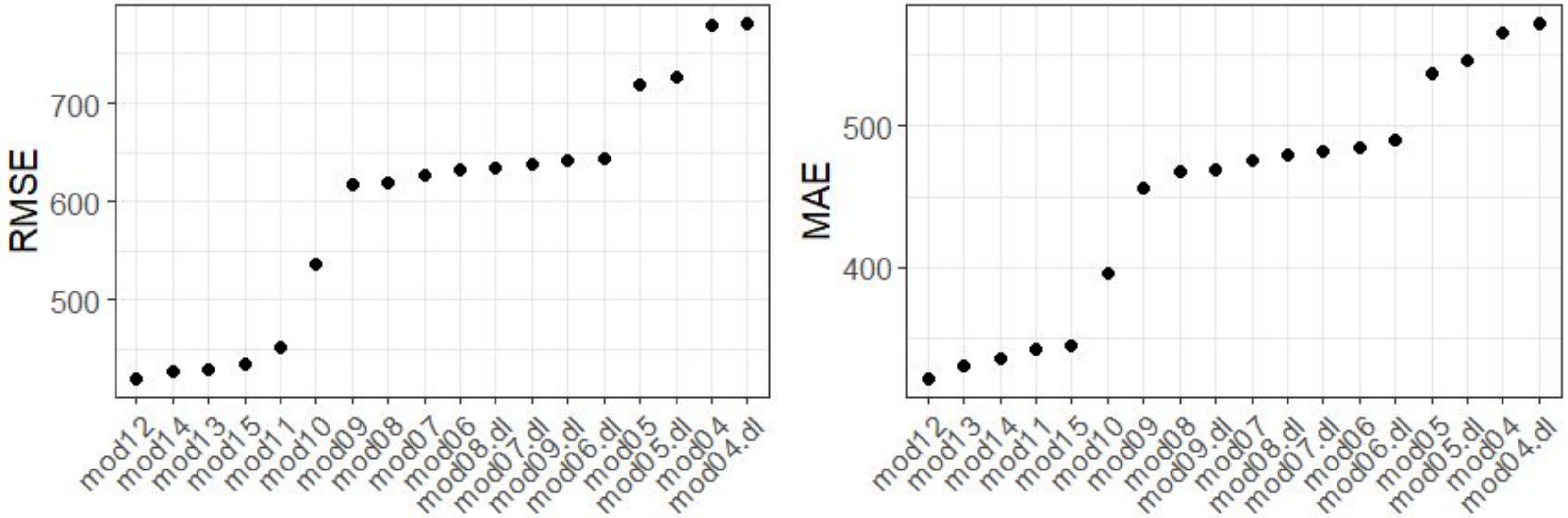


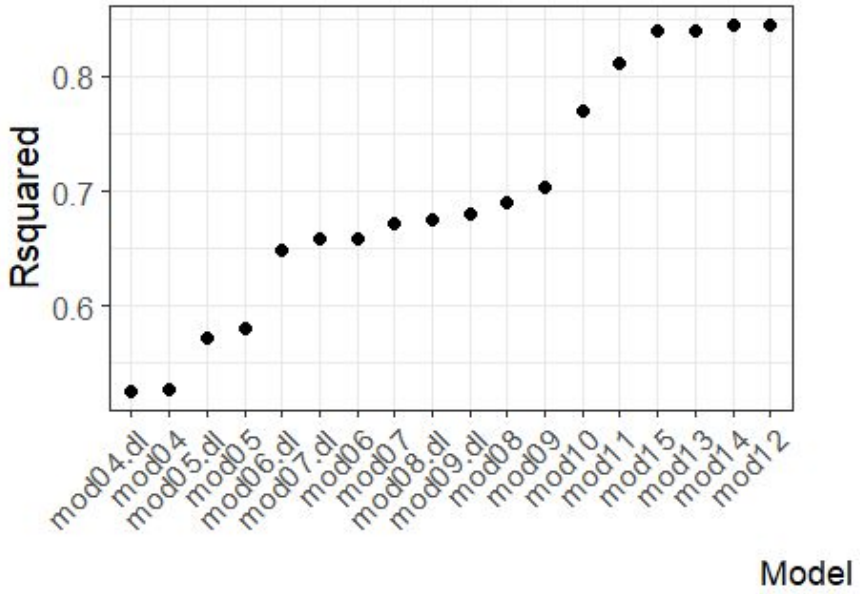
2010-2023



LEN

- LOOCV using the “best performing model” for LEN instead of the status quo model
- The same set of years perform the best (2012-2023), and the RMSE is lower for the “best” model

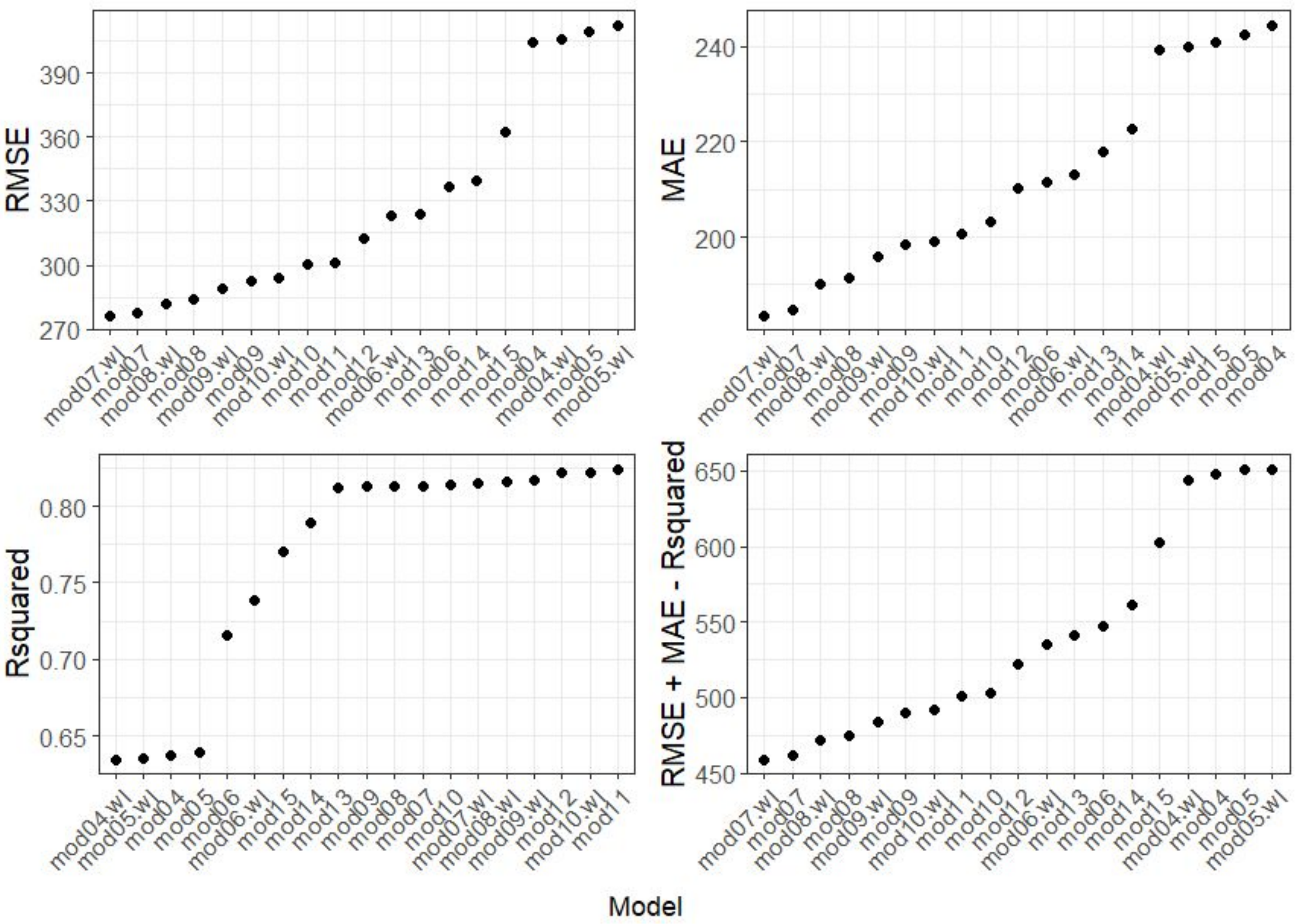




Model	RMSE (2012-2023)
Status Quo	458
“Best” Model	419

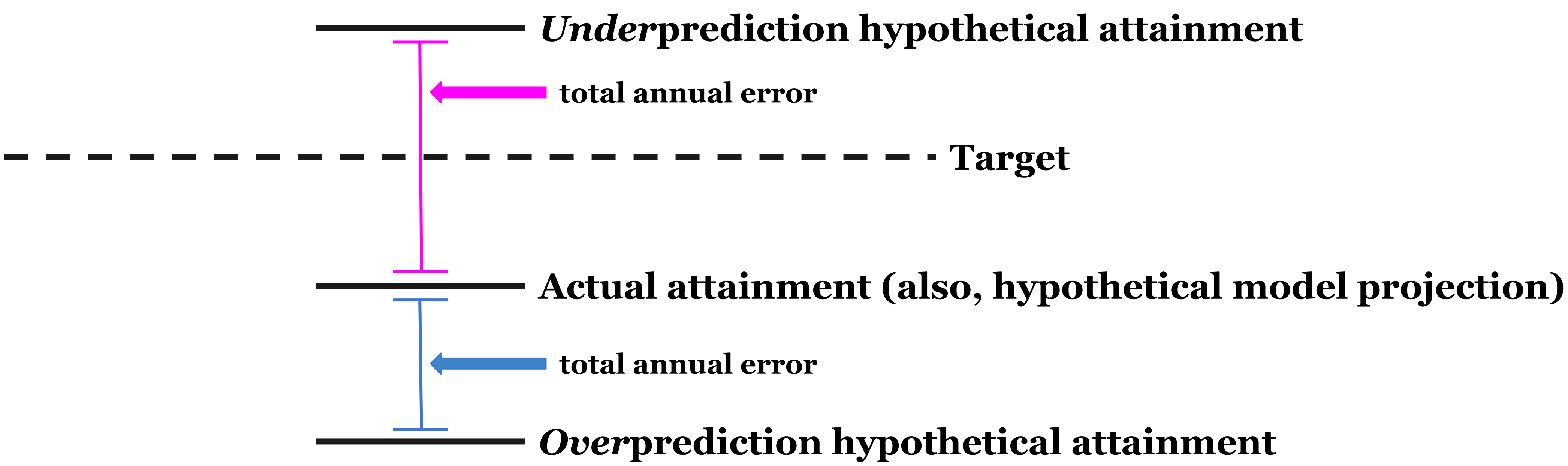
OAN

- LOOCV using the “best performing model” for OAN instead of the status quo model
- The same set of years perform the best (2012-2023), and the RMSE is lower for the “best” model
- The models *without* pre-2011 fixed effects (“.dl”) perform better using the “best” model



Model	RMSE (2007-2023)
Status Quo	311
“Best” Model	278

Illustration of the Hypothetical Attainment Calculation



total annual error:

- underprediction = added
- overprediction = subtracted