IMF and Benchmark Forecasts

1 Extracting error quantiles

Consider a forecast that stems from a source s for a specific target k in a country j, for target year t and with forecast horizon h:

$$\hat{y}_{s,k,j,t,h}$$

For example, this could be a forecast stemming from the International Monetary Fund World Economic Outlook ($s = IMF$) for real GDP growth ($k = gdpg$) in Canada ($j = Canada$) for the year 2022 ($t = 2022$). h then indexes the forecast horizon, where we code:

$$h = \begin{cases} 0, & \text{for forecasts made in October of the same year} \\ 0.5, & \text{for forecasts made in April of the same year} \\ 1, & \text{for forecasts made in October of the previous year} \\ 1.5, & \text{for forecasts made in April of the previous year} \end{cases}$$

After the target year has completed, we obtain the realized value for the quantity of interest. For these, the WEO updates publishes biannual updates for two years, yielding 4 versions of the realized value. In accordance with previous literature (*cite Timmermann 2008*), we use the version that is published in October of the following year and thereby don't index the true value by its publishing date (*rephrase*). We thus write the true value as

$$\hat{y}_{k,j,t}$$

Given the forecast and the realized value for the quantity of interest, we can calculate the respective forecast error as

$$e^{d}_{s,k,j,t,h} = y_{k,j,t} - \hat{y}_{s,k,j,t,h}$$

for the "directional" error method and as

$$e^{a}_{s,k,j,t,h} = |y_{k,j,t} - \hat{y}_{s,k,j,t,h}|$$

for the "absolute" error method.

The objective is to extract quantiles from sets of errors $\mathcal{E}_{s,k,j,t,h}$ constructed of certain years, depending on the estimation method m, to be able to quantify the uncertainty inherent in the forecasts via central prediction intervals of level $\alpha = \{0.5, 0.8\}$. For the estimation method, we consider a "rolling window" method, an "expanding window" method, and a "leave-one-out" method. For the rolling window method ($m = rw$), the errors of the last nine years enter into the estimation. For the expanding window method ($m = ew$), all previous years are considered, leaving a nine year window up front for the first estimation. For the leave-one-out method, all years except the current target year enter the estimation set. The latter is of course equivalent to the expanding window method in a real time setting and is considered in the scope of this analysis as a mere check *rephrase*. As an example, the error set for the "directional" error method and the rolling window approach is

$$\mathcal{E}^{d,rw}_{s,k,j,t,h} = \left\{ e^{d}_{s,k,j,t^*,h} | t - 9 \leq t^* < t \right\}$$

Insert reasoning to use the past 9 errors.

To now obtain the lower l and upper u values for a central prediction interval of level α, we take quantiles of these sets and add them to the current prediction:

For the directional method:

$$l^{\alpha,d}_{t,h,v,l,j} = \hat{y}_{t,h,l,j} + q^{0.5-\alpha/2}\left(\mathcal{E}^{d,m}_{t,h,v,l,j}\right)$$

$$u^{\alpha,d}_{t,h,v,l,j} = \hat{y}_{t,h,l,j} + q^{0.5+\alpha/2}\left(\mathcal{E}^{d,m}_{t,h,v,l,j}\right)$$

And for the absolute method:

$$l^{\alpha,a}_{t,h,v,l,j} = \hat{y}_{t,h,l,j} - q^{\alpha}\left(\mathcal{E}^{m,a}_{t,h,v,l,j}\right)$$

$$u^{\alpha,a}_{t,h,v,l,j} = \hat{y}_{t,h,l,j} + q^{\alpha}\left(\mathcal{E}^{m,a}_{t,h,v,l,j}\right)$$

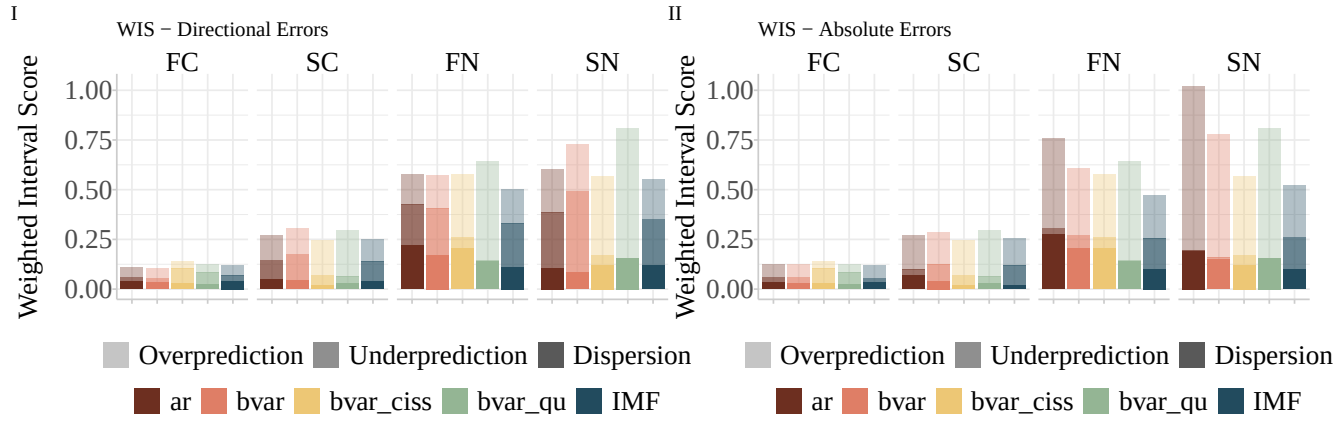
Two different philosophies.

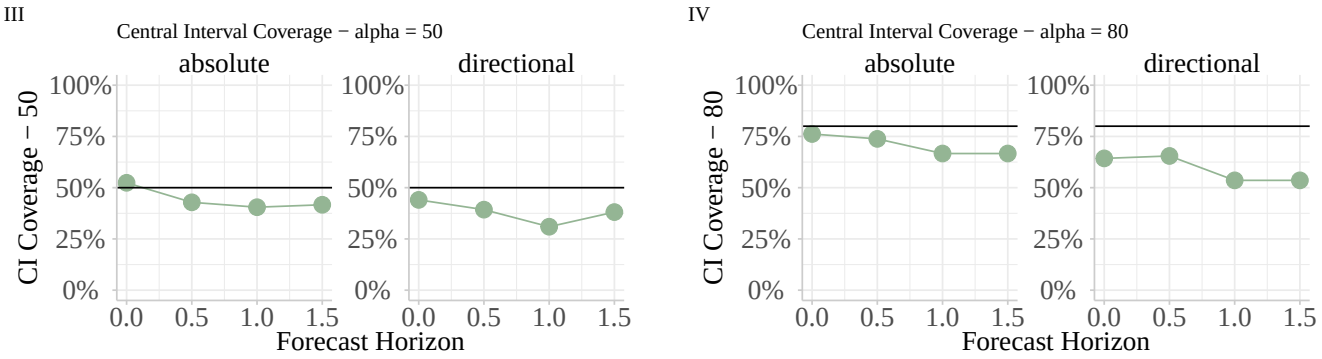
The absolute method will always yield symmetric central prediction intervals around the forecast value, while the directional method will in general yield asymmetric intervals. They thus result in different central intervals, unless the errors in $\mathcal{E}$ are perfectly symmetric around zero[1]. In fact, the directional method can yield central prediction intervals that do not even contain the forecast value, in cases where the $(0.5 - \alpha/2)$-quantile is positive or the $(0.5 + \alpha/2)$-quantile is negative.

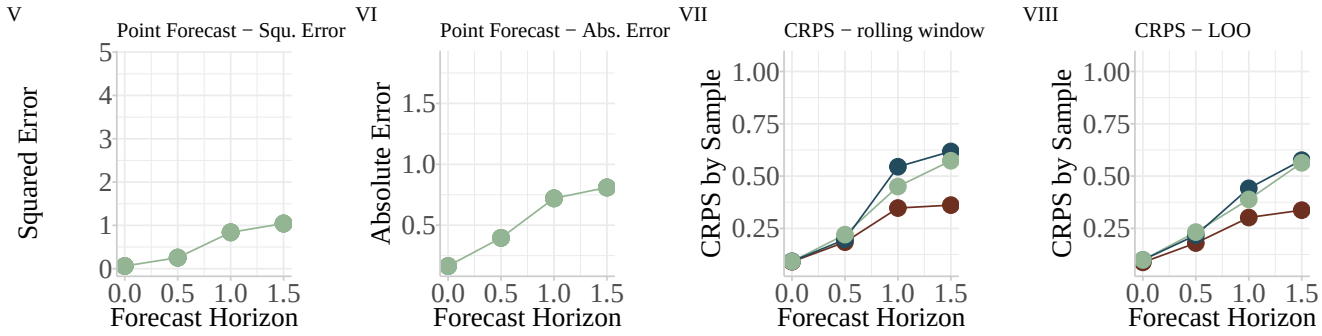
[1]Not totally correct, actually. For this to hold exactly, the error set would need to be augmented with one zero value.

2 Scores, by error method, Horizon and forecast source

2.1 Inflation



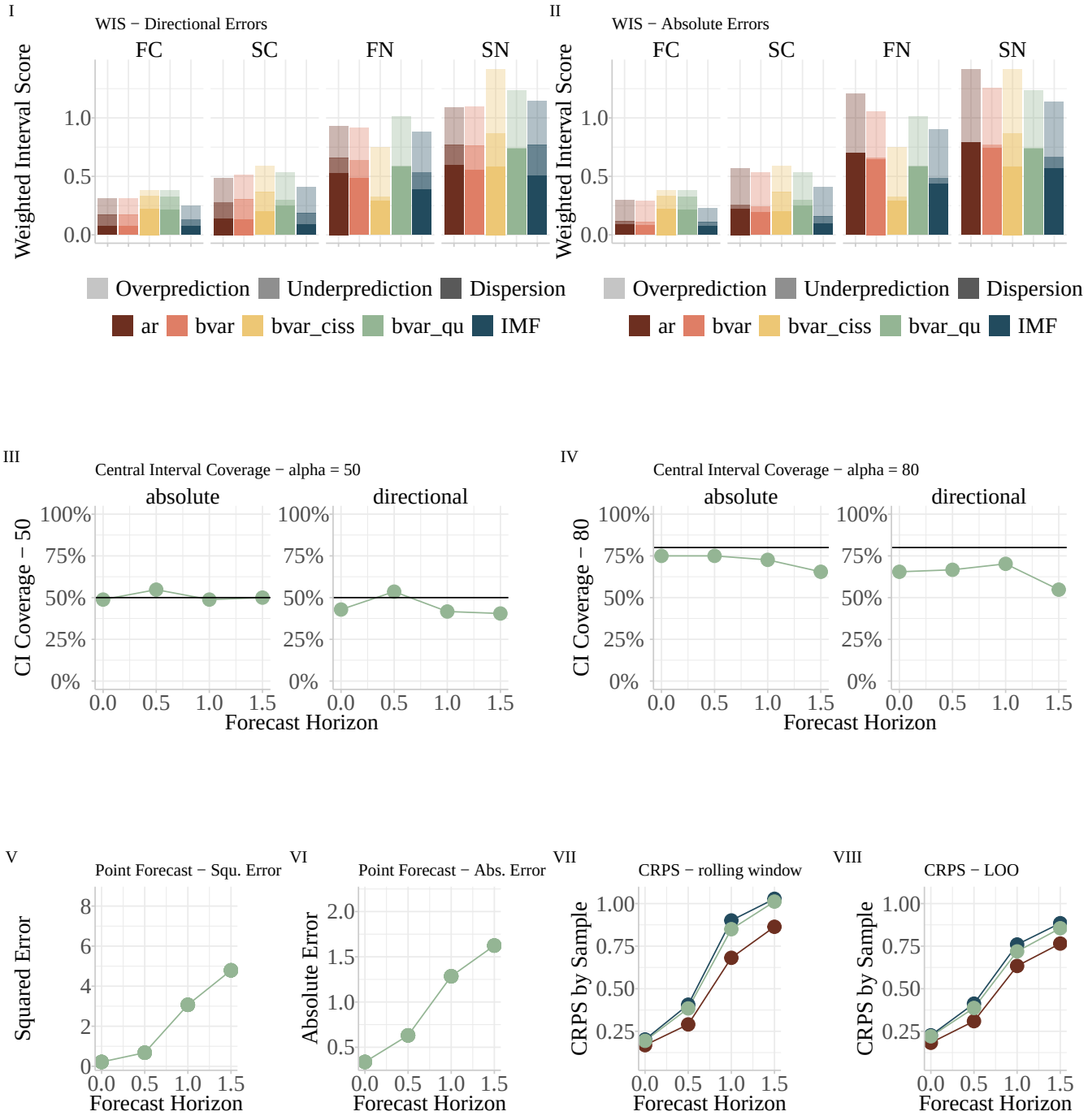




Some notes:

- Inflation: directional vs. absolute errors:
  - difference small for IMF method, absolute slightly better, likely due to longer central intervals
  - AR and BVAR profit more from directional correction (upward bias)
  - for expanding window method, difference in coverage is smaller (-> structural breaks)
- Inflation overall scores: IMF forecasts outperform others
  - lower scores for point forecasts
  - lower WIS
  - lower bias (compute directly?)
- GDP Growth: more similar results for different sources
  - lower scores at shorter horizons, more similar at larger horizons
  - IMF forecasts better only for absolute error method

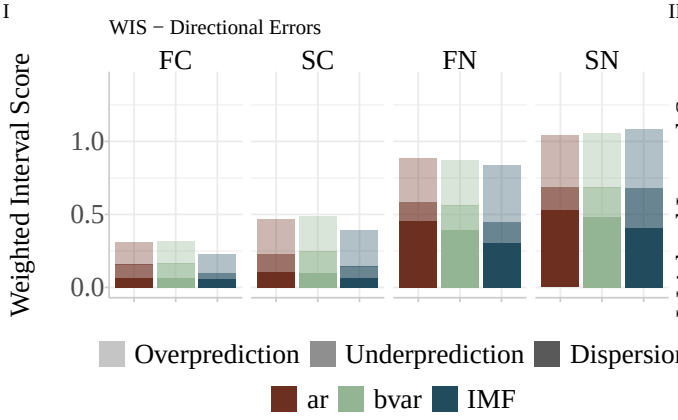
2.2 GDP


I
WIS − Directional Errors
II
WIS − Absolute Errors
FC
SC
FN
SN
Weighted Interval Score
1.0
0.5
0.0
Overprediction
Underprediction
Dispersion
ar
bvar
bvar_ciss
bvar_qu
IMF
III
Central Interval Coverage − alpha = 50
IV
Central Interval Coverage − alpha = 80
absolute
directional
CI Coverage − 50
CI Coverage − 80
100%
75%
50%
25%
0%
0.0
0.5
1.0
1.5
Forecast Horizon
V
Point Forecast − Squ. Error
VI
Point Forecast − Abs. Error
VII
CRPS − rolling window
VIII
CRPS − LOO
Squared Error
Absolute Error
CRPS by Sample
2.0
1.5
1.00
0.75
0.50
0.25

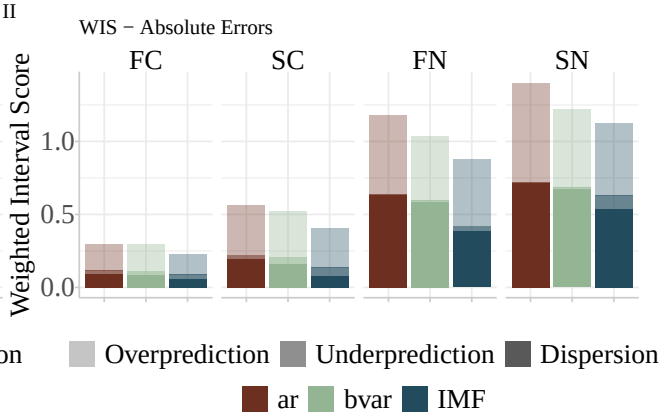
3 Expanding Window - Scores, by error method, Horizon and forecast source

3.1 Inflation

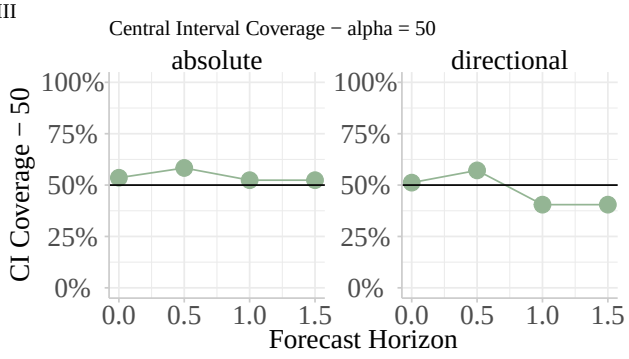
I



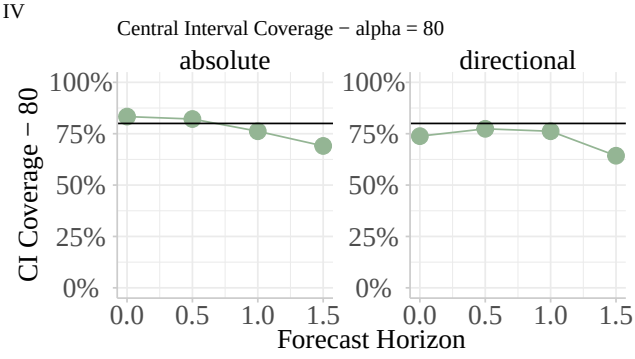
II



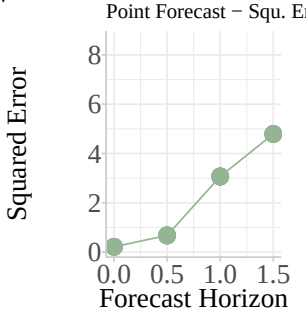
III



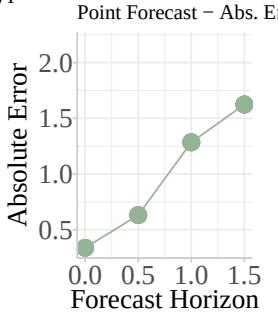
IV



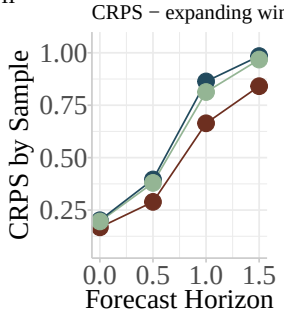
V



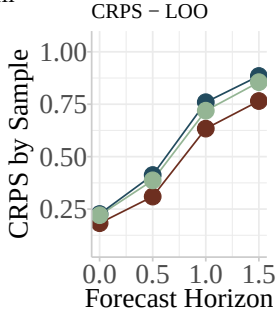
VI



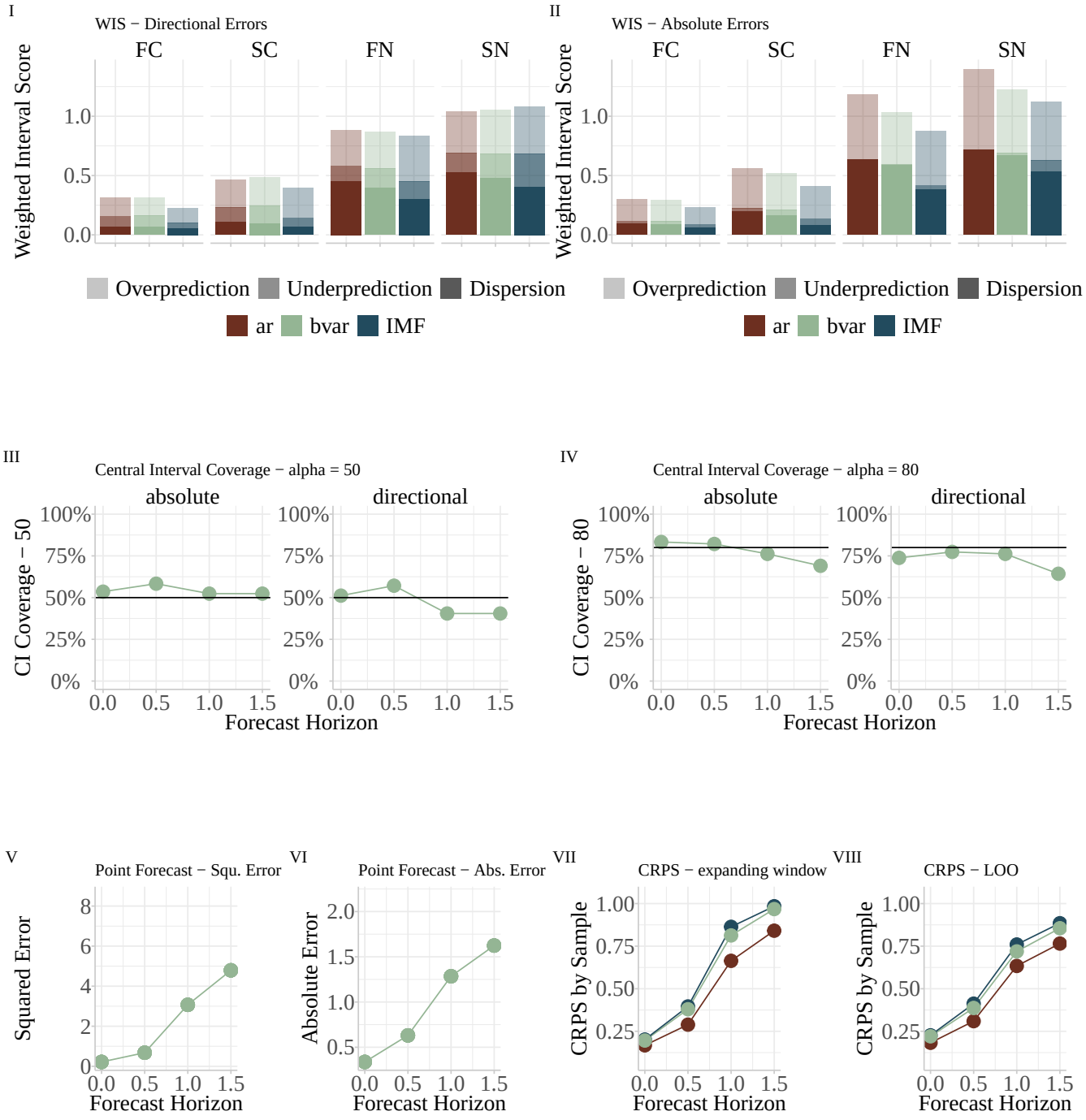
VII



VIII

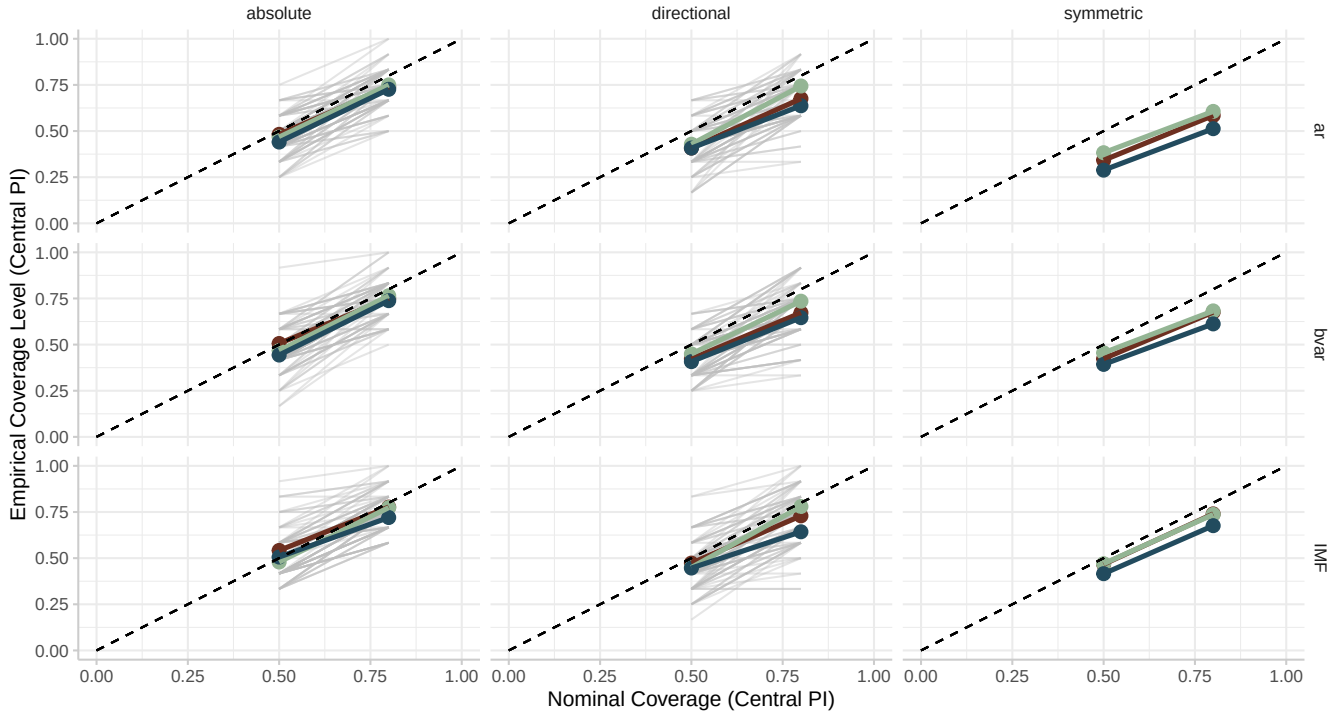


3.2 GDP

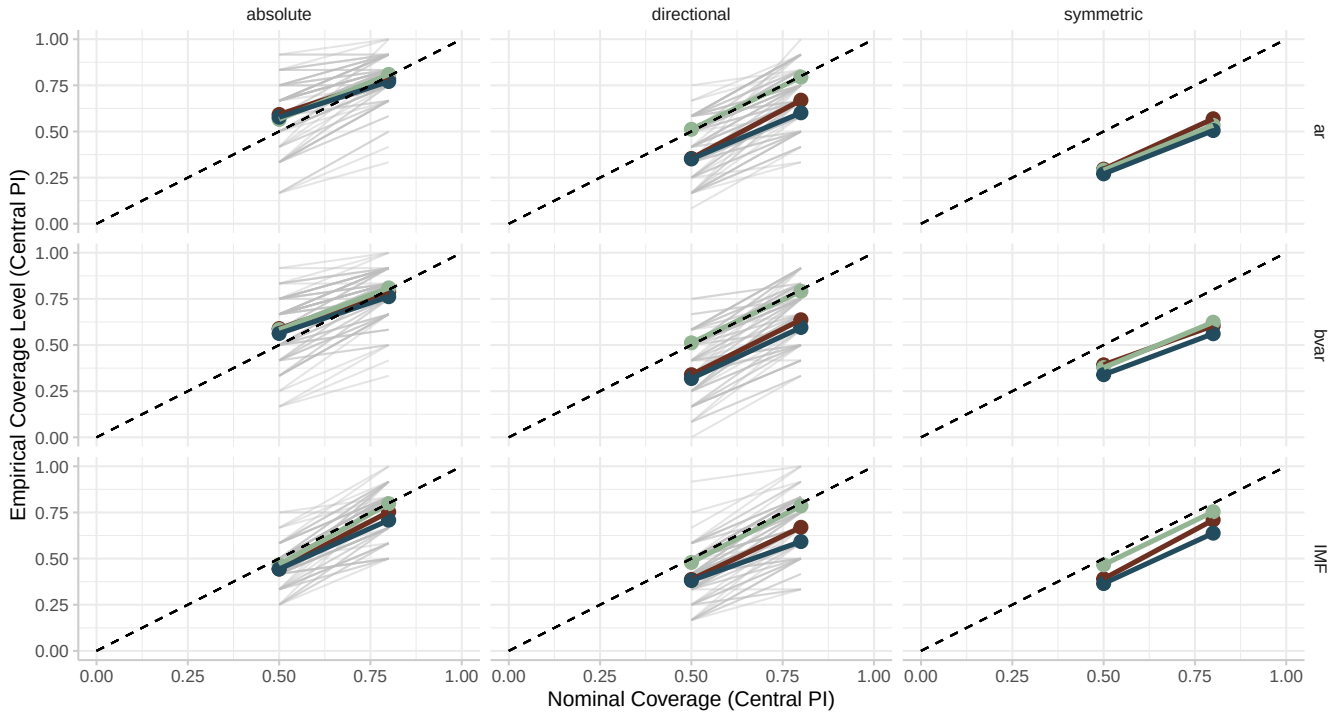

I
WIS − Directional Errors
II
WIS − Absolute Errors
FC
SC
FN
SN
Weighted Interval Score
1.0
0.5
0.0
Overprediction
Underprediction
Dispersion
ar
bvar
IMF
III
Central Interval Coverage − alpha = 50
IV
Central Interval Coverage − alpha = 80
absolute
directional
CI Coverage − 50
CI Coverage − 80
100%
75%
50%
25%
0%
0.0
0.5
1.0
1.5
Forecast Horizon
V
Point Forecast − Squ. Error
VI
Point Forecast − Abs. Error
VII
CRPS − expanding window
VIII
CRPS − LOO
Squared Error
Absolute Error
CRPS by Sample
8
6
4
2
0
2.0
1.5
1.0
0.5
1.00
0.75
0.50
0.25

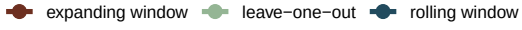
4 Coverage, by target, methods and source

I



II

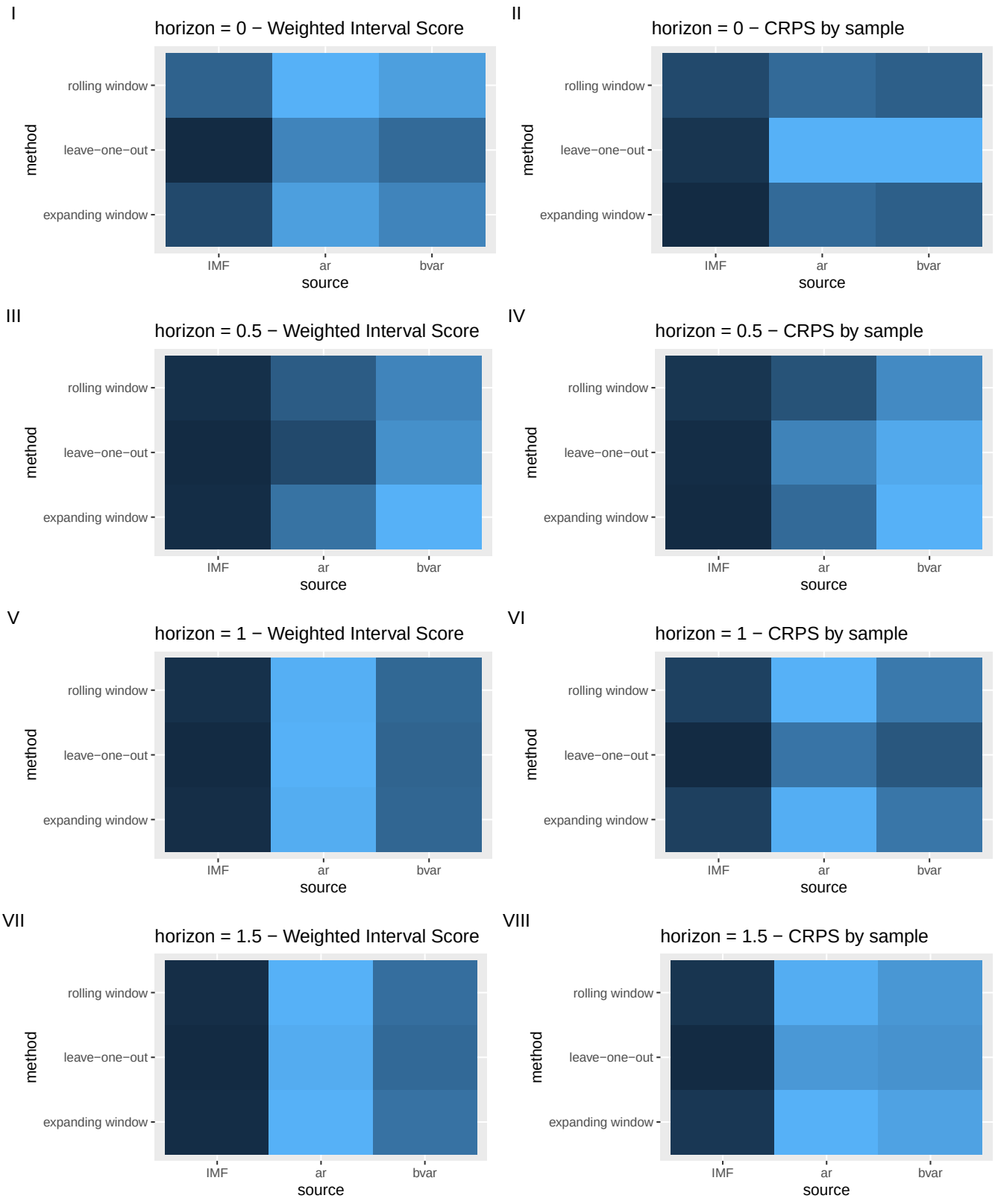




	IMF	ar	bvar
horizon = 0			
expanding window_interval_score	0.114	0.124	0.121
expanding window_sample_crps	0.087	0.093	0.092
leave-one-out_interval_score	0.110	0.121	0.118
leave-one-out_sample_crps	0.088	0.099	0.099
rolling window_interval_score	0.117	0.126	0.124
rolling window_sample_crps	0.090	0.093	0.092
horizon = 0.5			
expanding window_interval_score	0.255	0.281	0.302
expanding window_sample_crps	0.179	0.207	0.235
leave-one-out_interval_score	0.254	0.266	0.291
leave-one-out_sample_crps	0.180	0.217	0.232
rolling window_interval_score	0.256	0.273	0.287
rolling window_sample_crps	0.184	0.197	0.220
horizon = 1			
expanding window_interval_score	0.464	0.757	0.601
expanding window_sample_crps	0.344	0.540	0.445
leave-one-out_interval_score	0.456	0.766	0.596
leave-one-out_sample_crps	0.302	0.442	0.388
rolling window_interval_score	0.471	0.761	0.605
rolling window_sample_crps	0.347	0.545	0.450
horizon = 1.5			
expanding window_interval_score	0.517	1.020	0.795
expanding window_sample_crps	0.367	0.625	0.596
leave-one-out_interval_score	0.507	1.004	0.760

leave-one-out_sample_crps	0.337	0.576	0.564
rolling window_interval_score	0.521	1.021	0.780
rolling window_sample_crps	0.361	0.618	0.573

4.1 Inflation

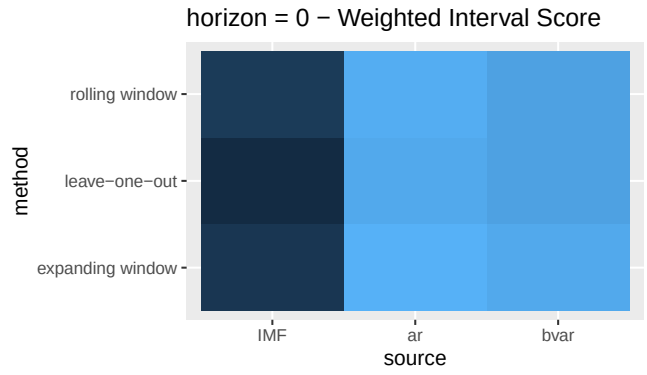


4.2 GDP

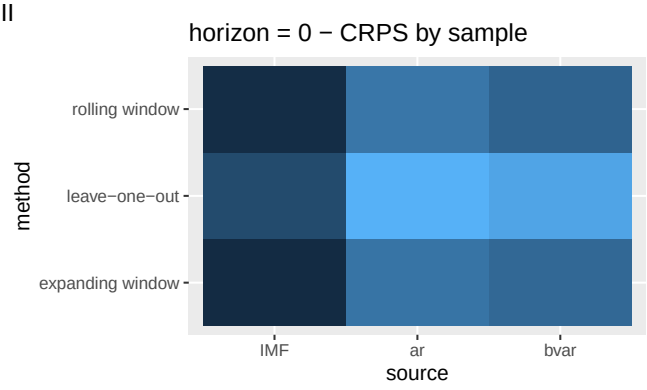
	IMF	ar	bvar
horizon = 0			
expanding window_interval_score	0.229	0.297	0.293
expanding window_sample_crps	0.168	0.201	0.196
leave-one-out_interval_score	0.222	0.293	0.289
leave-one-out_sample_crps	0.183	0.226	0.221
rolling window_interval_score	0.232	0.295	0.289
rolling window_sample_crps	0.169	0.202	0.194
horizon = 0.5			
expanding window_interval_score	0.408	0.562	0.521
expanding window_sample_crps	0.289	0.396	0.380
leave-one-out_interval_score	0.396	0.558	0.515
leave-one-out_sample_crps	0.310	0.412	0.387
rolling window_interval_score	0.411	0.570	0.530
rolling window_sample_crps	0.290	0.406	0.384
horizon = 1			
expanding window_interval_score	0.878	1.182	1.033
expanding window_sample_crps	0.664	0.863	0.813
leave-one-out_interval_score	0.866	1.167	1.011
leave-one-out_sample_crps	0.634	0.760	0.719
rolling window_interval_score	0.901	1.204	1.056
rolling window_sample_crps	0.682	0.901	0.850
horizon = 1.5			
expanding window_interval_score	1.126	1.397	1.222
expanding window_sample_crps	0.841	0.984	0.968
leave-one-out_interval_score	1.097	1.388	1.215

leave-one-out_sample_crps	0.765	0.885	0.855
rolling window_interval_score	1.139	1.417	1.253
rolling window_sample_crps	0.864	1.028	1.012

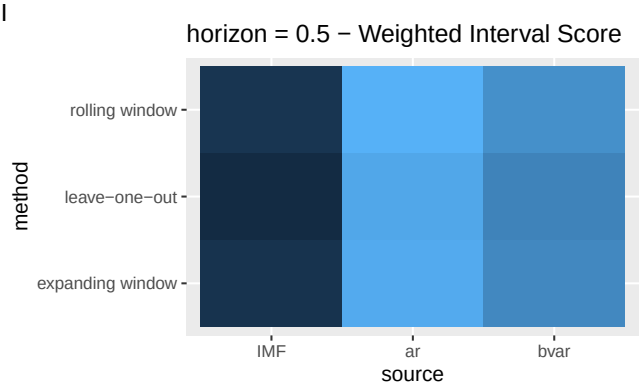
I



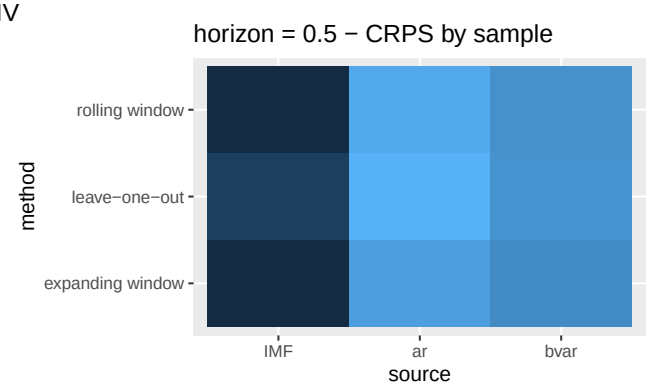
II



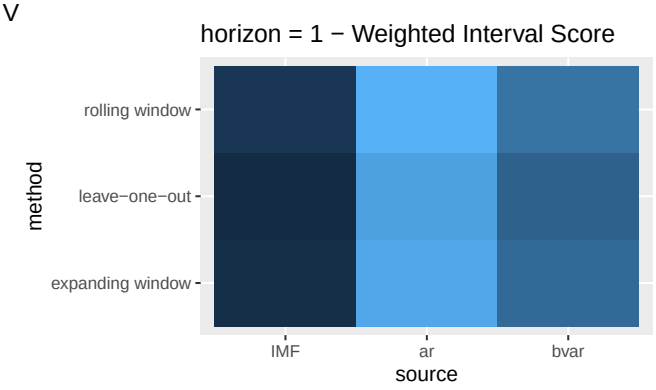
III



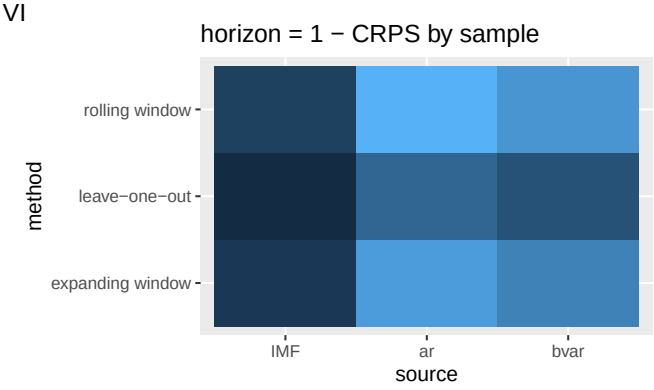
IV



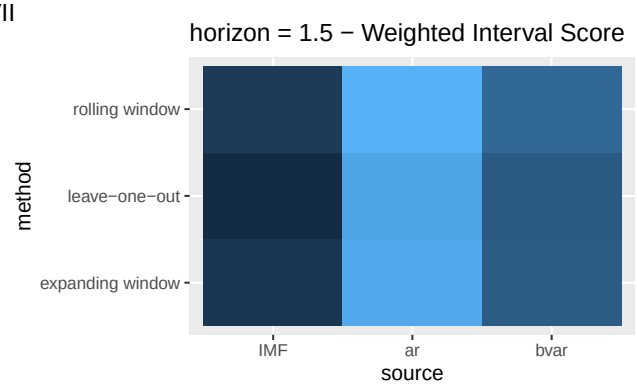
V



VI



VII



VIII