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.

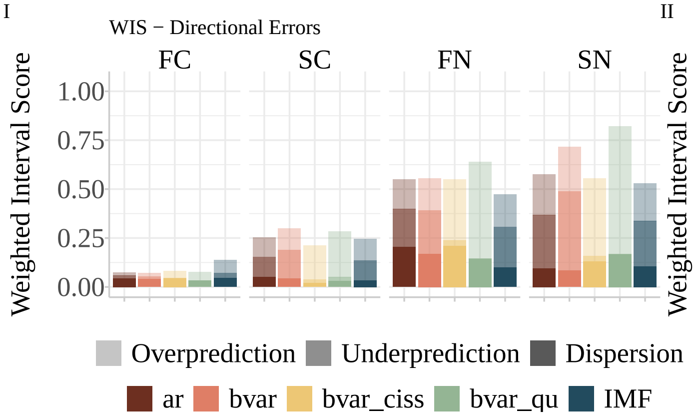
The absolute method will always yield symmetric central prediction intervals around the forecast value, while the directional method will in general yield asymmetric intervals. They thus result in different central intervals, unless the errors in $\mathcal{E}$ are perfectly symmetric around zero[1]. 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.

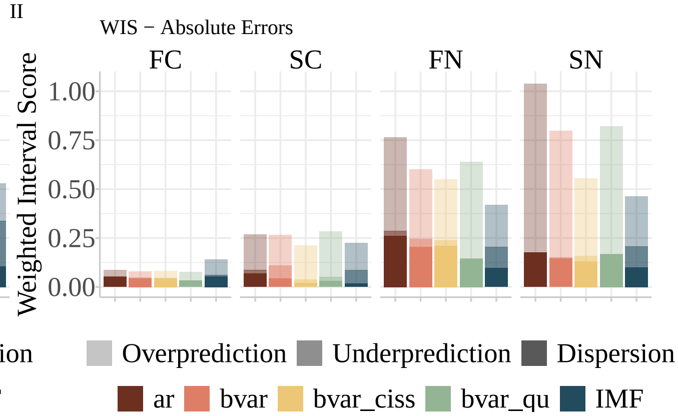
2 Scores, by error method, Horizon and forecast source

2.1 Inflation

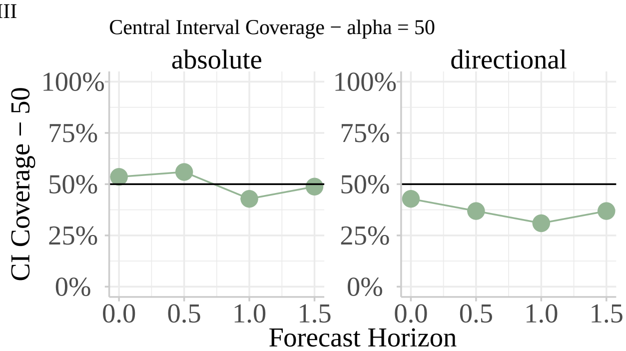
I



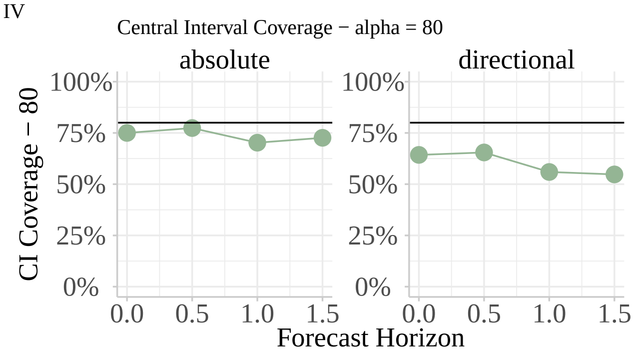
II



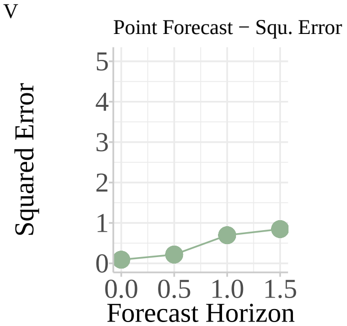
III



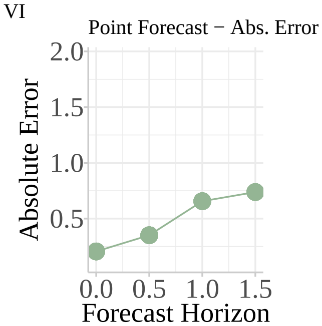
IV



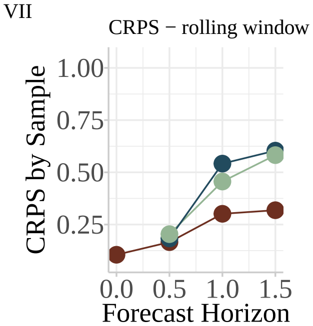
V



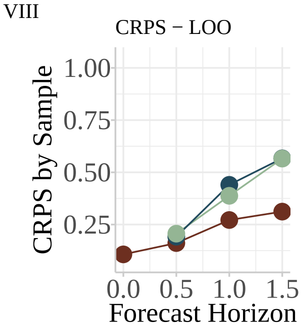
VI



VII



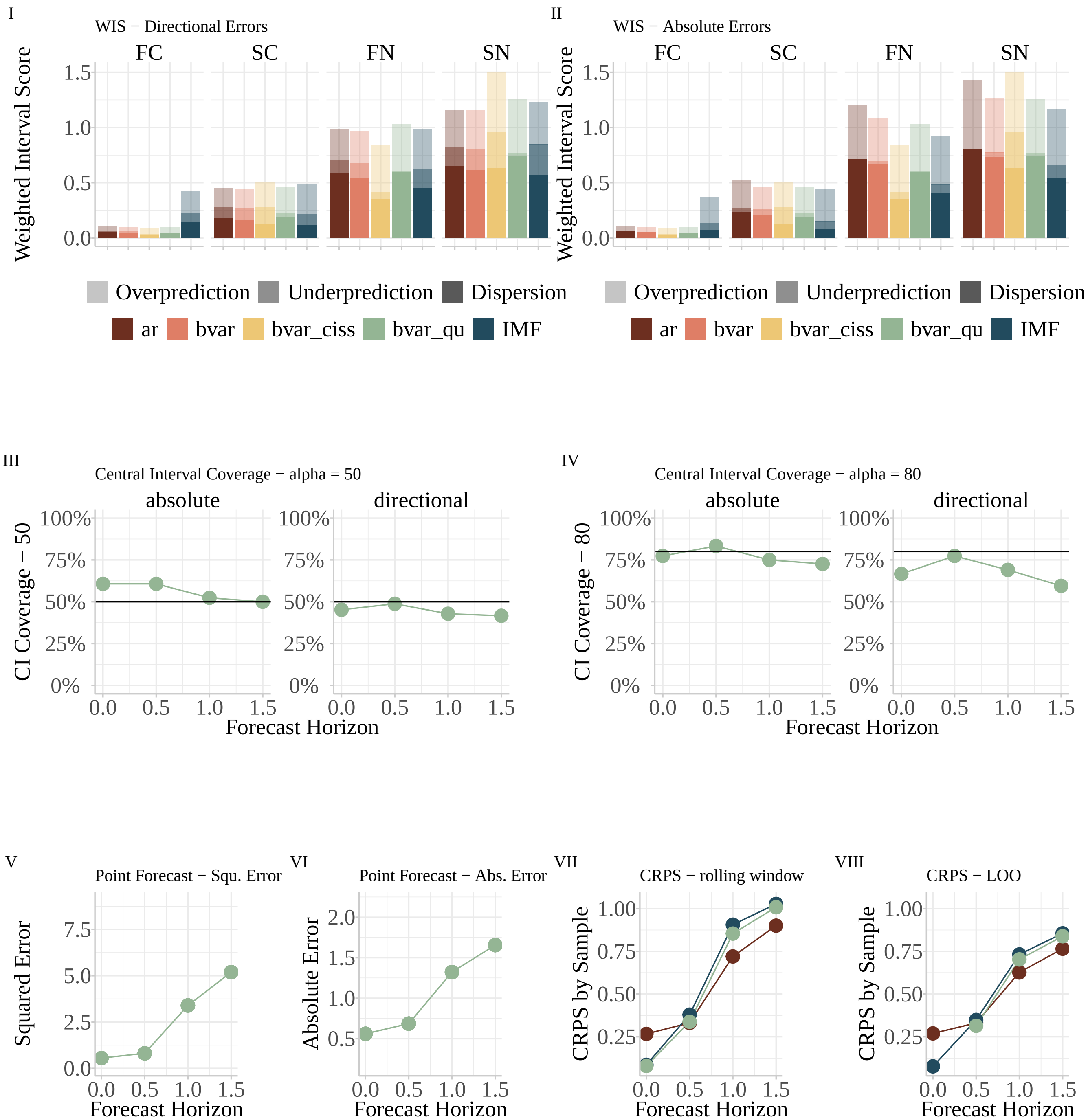
VIII



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

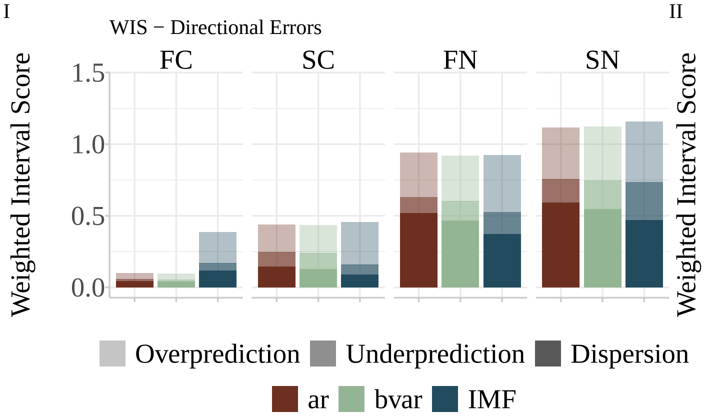
2.2 GDP


I
WIS − Directional Errors
II
WIS − Absolute Errors
FC
SC
FN
SN
Weighted Interval Score
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
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

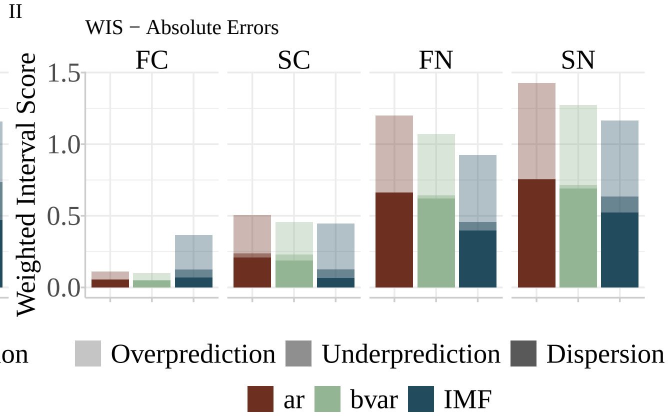
3 Expanding Window - Scores, by error method, Horizon and forecast source

3.1 Inflation

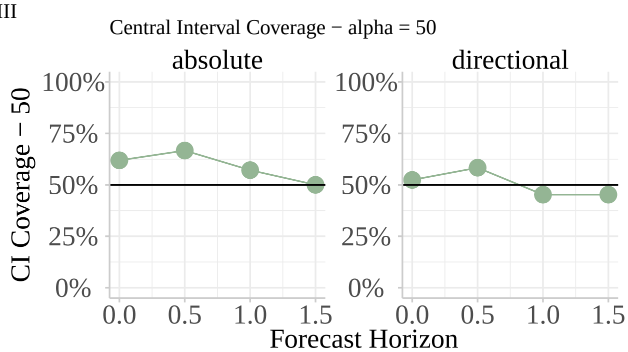
I



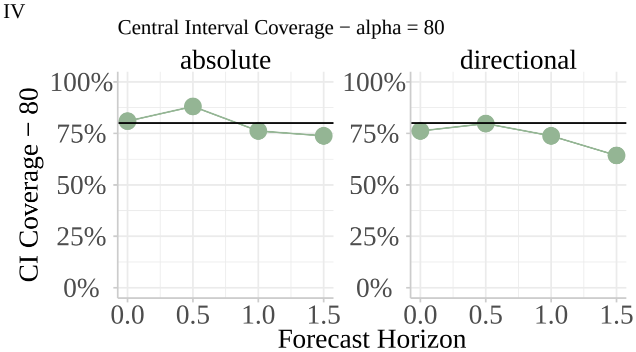
II



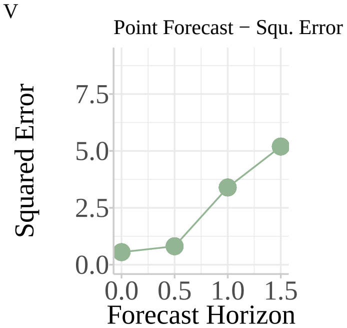
III



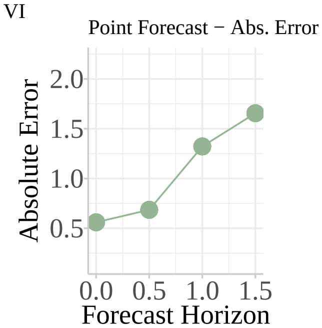
IV



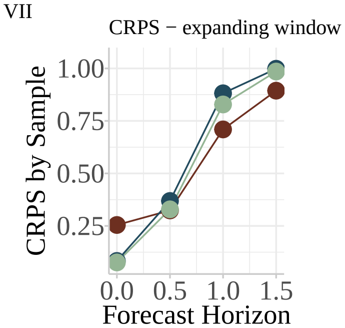
V



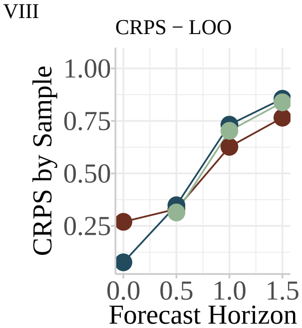
VI



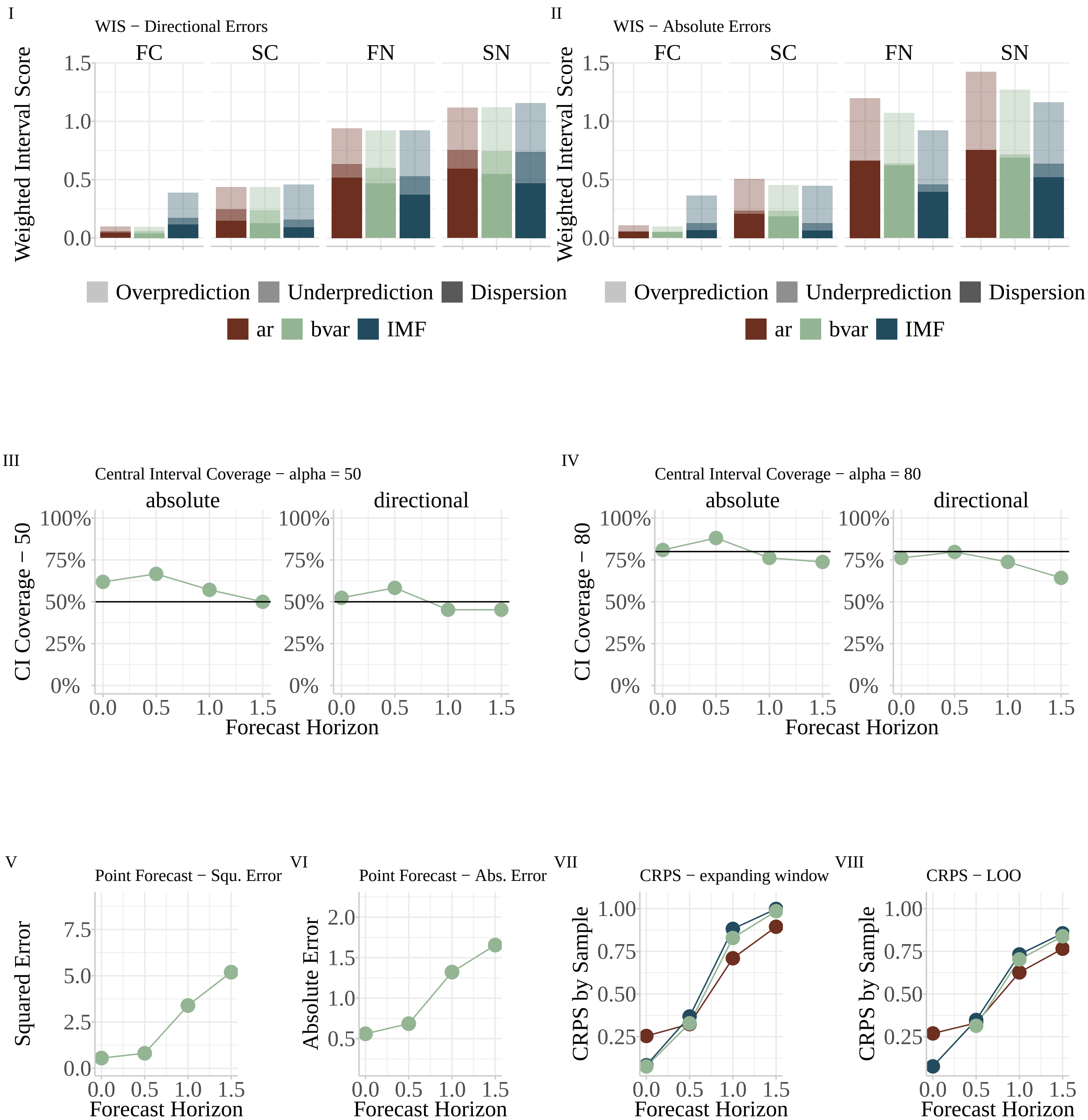
VII



VIII



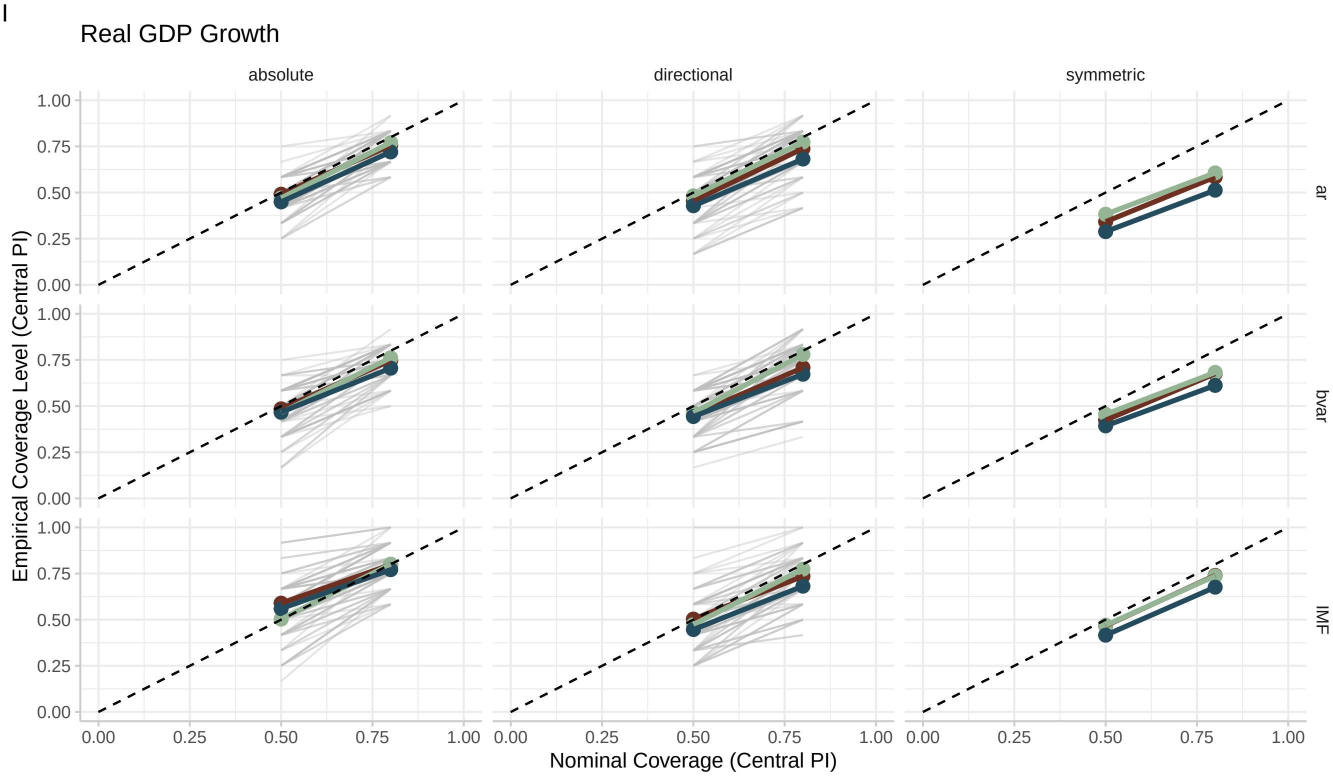
3.2 GDP



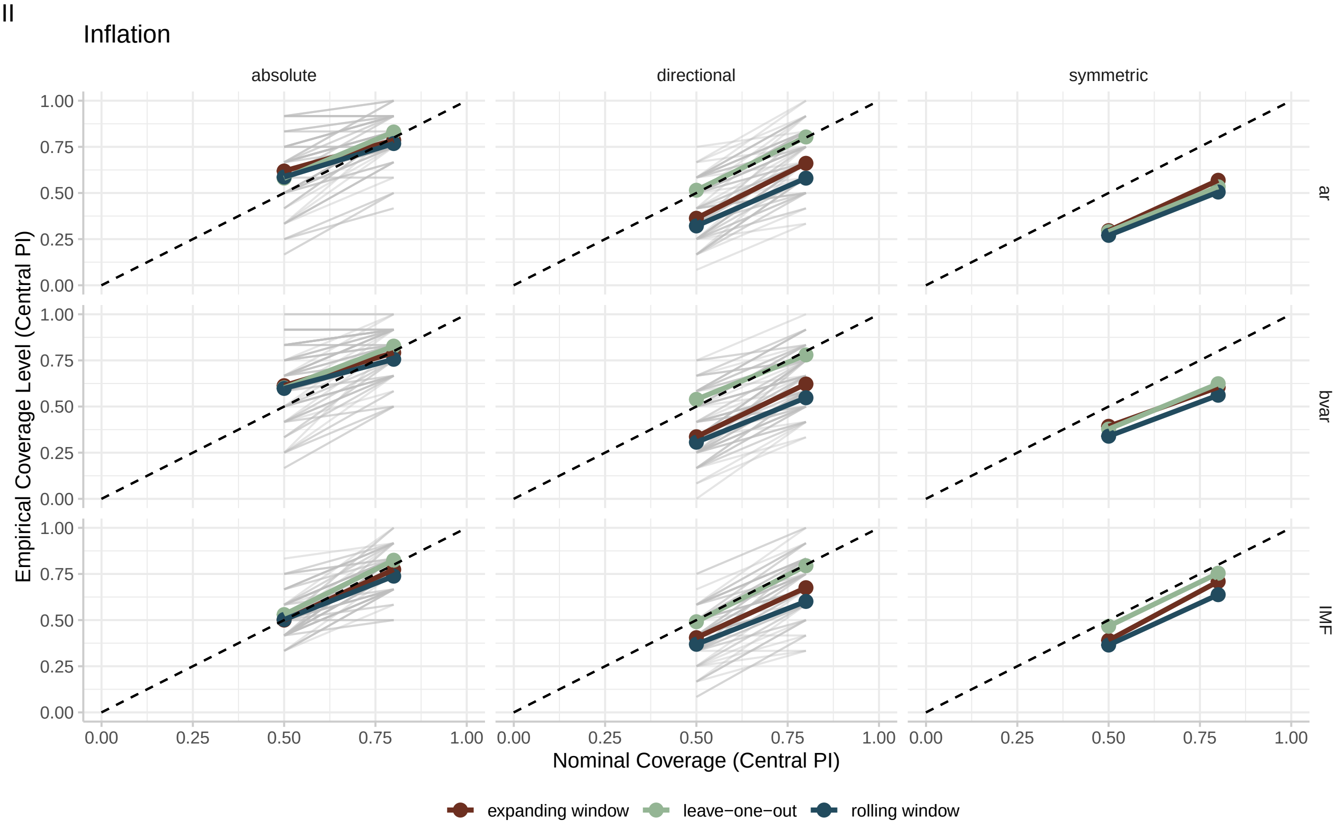
I
WIS − Directional Errors
II
WIS − Absolute Errors
FC
SC
FN
SN
Weighted Interval Score
Overprediction
Underprediction
Dispersion
ar
bvar
IMF
III
Central Interval Coverage − alpha = 50
IV
Central Interval Coverage − alpha = 80
absolute
directional
CI Coverage − 50
CI Coverage − 80
Forecast Horizon
V
Point Forecast − Squ. Error
Squared Error
VI
Point Forecast − Abs. Error
Absolute Error
VII
CRPS − expanding window
VIII
CRPS − LOO
CRPS by Sample

4 Coverage, by target, methods and source

I



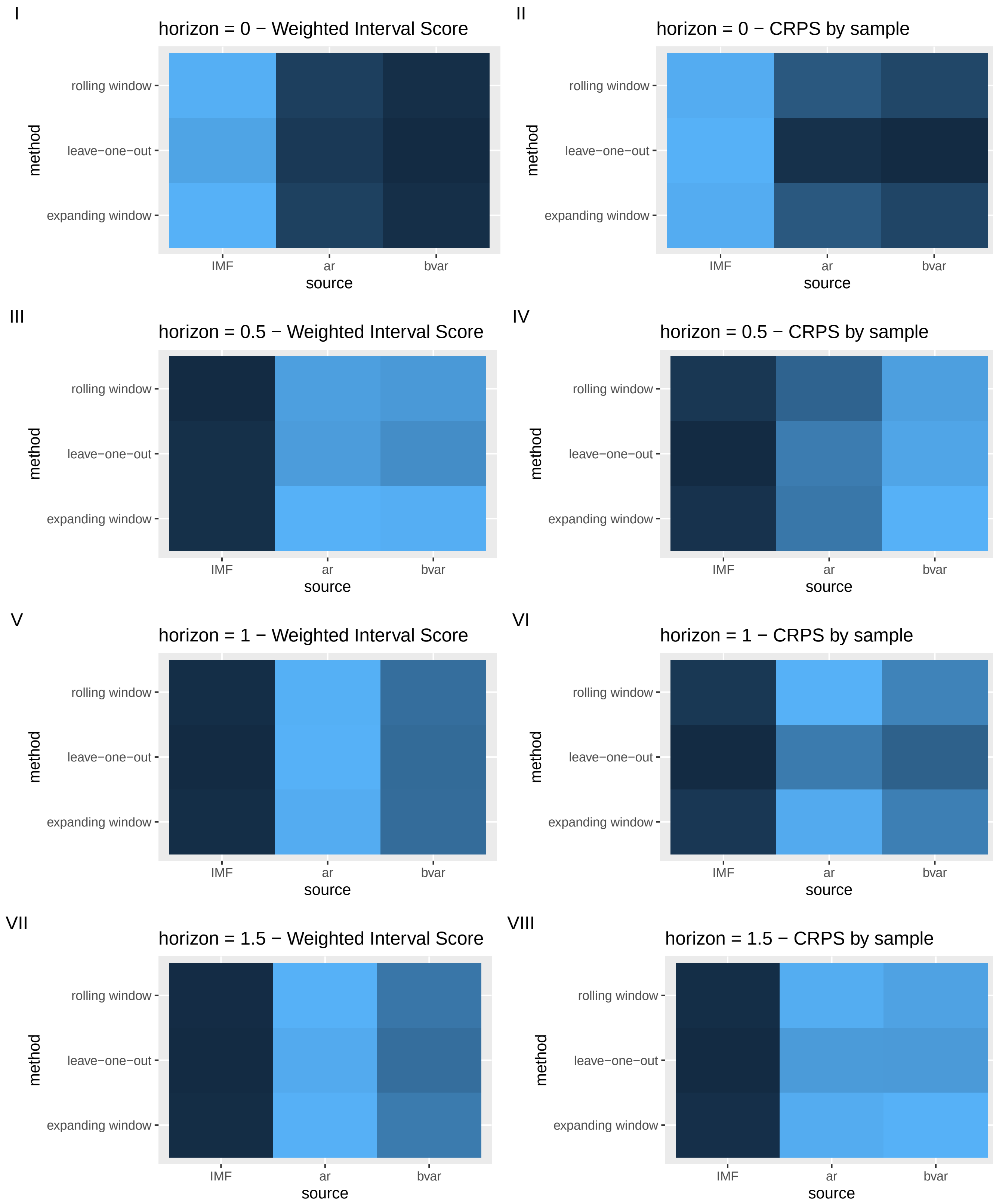
II



	IMF	ar	bvar
horizon = 0			
expanding window__interval__score	0.142	0.088	0.078
expanding window__sample__crps	0.105	0.068	0.059
leave-one-out__interval__score	0.136	0.084	0.076
leave-one-out__sample__crps	0.107	0.049	0.046
rolling window__interval__score	0.141	0.087	0.078
rolling window__sample__crps	0.105	0.068	0.060
horizon = 0.5			
expanding window__interval__score	0.228	0.275	0.274
expanding window__sample__crps	0.164	0.190	0.210
leave-one-out__interval__score	0.228	0.268	0.263
leave-one-out__sample__crps	0.161	0.192	0.206
rolling window__interval__score	0.226	0.269	0.267
rolling window__sample__crps	0.166	0.183	0.204
horizon = 1			
expanding window__interval__score	0.420	0.757	0.595
expanding window__sample__crps	0.300	0.530	0.448
leave-one-out__interval__score	0.412	0.768	0.591
leave-one-out__sample__crps	0.272	0.440	0.388
rolling window__interval__score	0.420	0.765	0.600
rolling window__sample__crps	0.301	0.542	0.456
horizon = 1.5			
expanding window__interval__score	0.465	1.037	0.819
expanding window__sample__crps	0.323	0.603	0.613
leave-one-out__interval__score	0.456	1.013	0.765

leave-one-out_sample_crps	0.312	0.567	0.566
rolling window_interval_score	0.463	1.039	0.799
rolling window_sample_crps	0.319	0.604	0.582

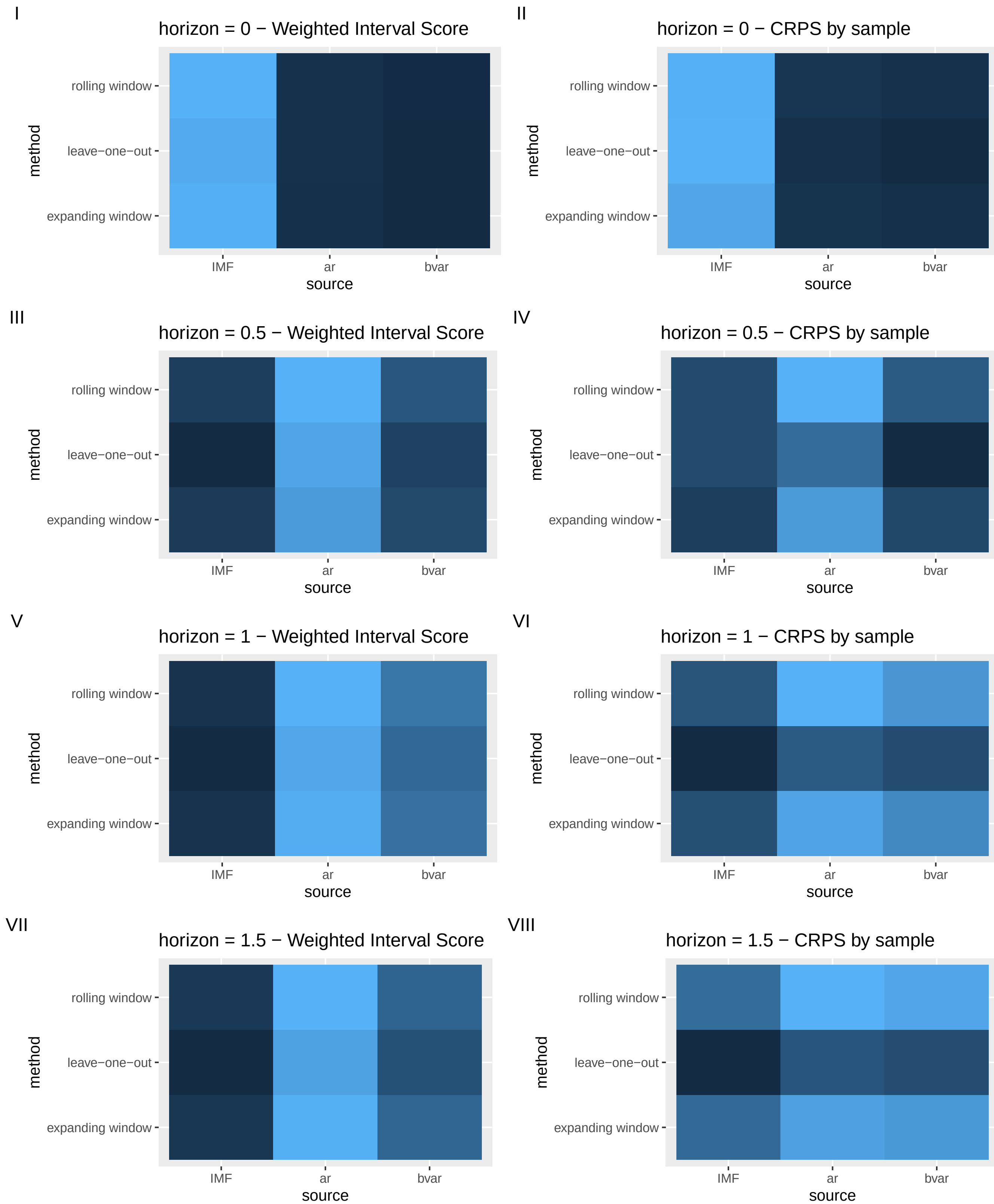
4.1 Inflation



4.2 GDP

	IMF	ar	bvar
horizon = 0			
expanding window__interval__score	0.365	0.110	0.099
expanding window__sample__crps	0.254	0.083	0.076
leave-one-out__interval__score	0.356	0.112	0.099
leave-one-out__sample__crps	0.269	0.077	0.068
rolling window__interval__score	0.368	0.112	0.100
rolling window__sample__crps	0.267	0.086	0.079
horizon = 0.5			
expanding window__interval__score	0.446	0.506	0.455
expanding window__sample__crps	0.324	0.369	0.329
leave-one-out__interval__score	0.435	0.512	0.451
leave-one-out__sample__crps	0.331	0.348	0.314
rolling window__interval__score	0.448	0.519	0.464
rolling window__sample__crps	0.331	0.379	0.338
horizon = 1			
expanding window__interval__score	0.922	1.198	1.070
expanding window__sample__crps	0.709	0.882	0.828
leave-one-out__interval__score	0.902	1.187	1.051
leave-one-out__sample__crps	0.627	0.732	0.703
rolling window__interval__score	0.922	1.208	1.082
rolling window__sample__crps	0.720	0.906	0.854
horizon = 1.5			
expanding window__interval__score	1.163	1.424	1.271
expanding window__sample__crps	0.894	0.998	0.985
leave-one-out__interval__score	1.135	1.400	1.227

leave-one-out_sample_crps	0.765	0.855	0.838
rolling window_interval_score	1.170	1.431	1.270
rolling window_sample_crps	0.900	1.028	1.008


I
horizon = 0 − Weighted Interval Score
II
horizon = 0 − CRPS by sample
III
horizon = 0.5 − Weighted Interval Score
IV
horizon = 0.5 − CRPS by sample
V
horizon = 1 − Weighted Interval Score
VI
horizon = 1 − CRPS by sample
VII
horizon = 1.5 − Weighted Interval Score
VIII
horizon = 1.5 − CRPS by sample
method
rolling window
leave−one−out
expanding window
source
IMF
ar
bvar

67