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.

2 A short note on error handling

In almost all 72 cases, absolute error handling gives lower scores than directional error handling. The only exception is the inflation series for the IMF forecasts and horizon 0, where the expanding window and rolling window method give *slightly* lower scores for the directional methodology. We thus decide to focus on the absolute errors in this document.

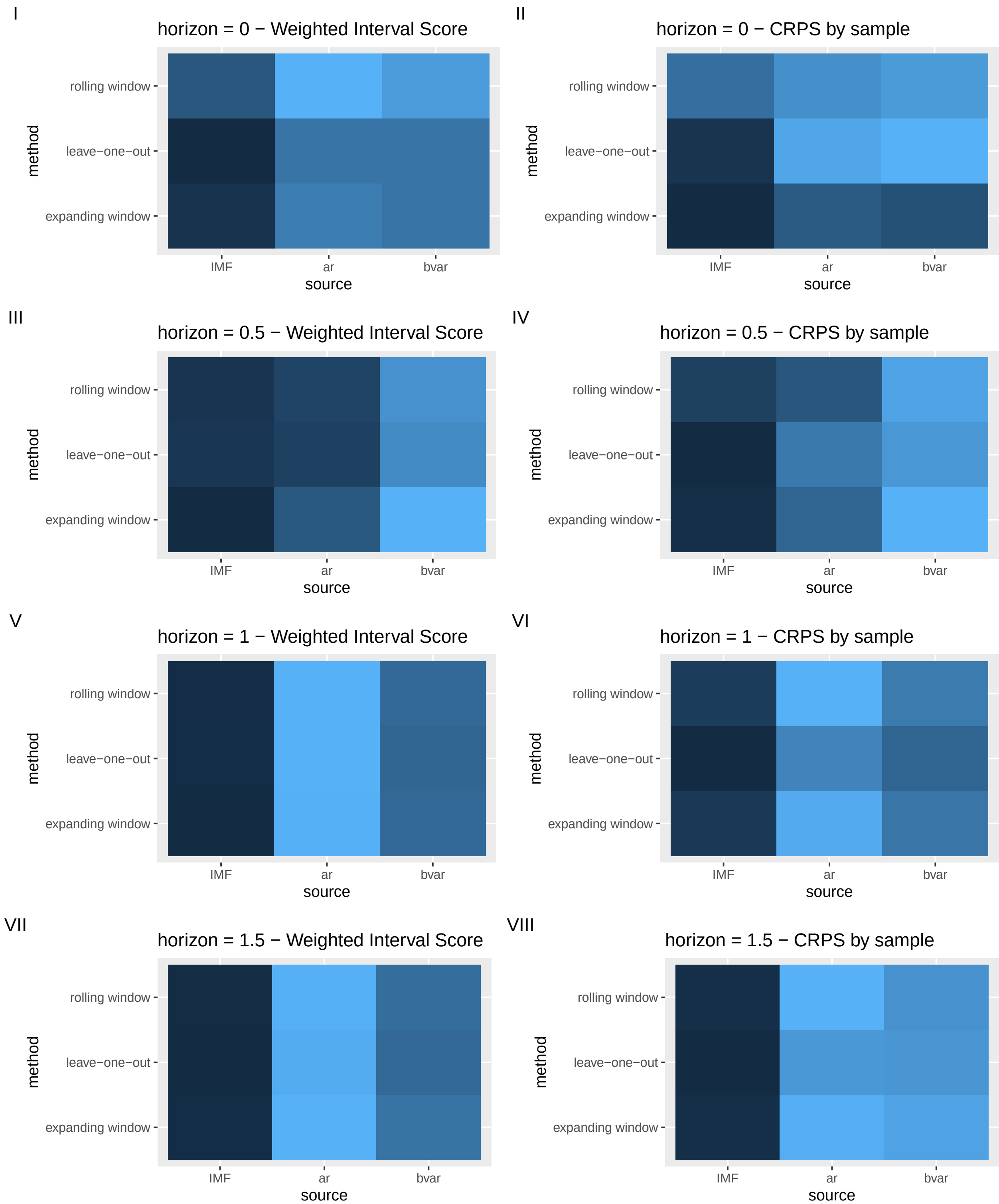
3 Scores, by estimation method, Horizon and forecast source

	IMF	ar	bvar
horizon = 0			
expanding window_interval_score	0.115	0.123	0.122
expanding window_sample_crps	0.087	0.092	0.091
leave-one-out_interval_score	0.114	0.122	0.122

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

leave-one-out__sample__crps	0.088	0.099	0.100
rolling window__interval__score	0.119	0.128	0.126
rolling window__sample__crps	0.094	0.097	0.098
horizon = 0.5			
expanding window__interval__score	0.258	0.272	0.296
expanding window__sample__crps	0.182	0.208	0.241
leave-one-out__interval__score	0.262	0.265	0.286
leave-one-out__sample__crps	0.180	0.217	0.230
rolling window__interval__score	0.261	0.266	0.288
rolling window__sample__crps	0.191	0.201	0.235
horizon = 1			
expanding window__interval__score	0.448	0.737	0.590
expanding window__sample__crps	0.327	0.504	0.426
leave-one-out__interval__score	0.453	0.739	0.584
leave-one-out__sample__crps	0.302	0.448	0.400
rolling window__interval__score	0.451	0.739	0.591
rolling window__sample__crps	0.333	0.514	0.434
horizon = 1.5			
expanding window__interval__score	0.495	1.044	0.800
expanding window__sample__crps	0.346	0.627	0.602
leave-one-out__interval__score	0.486	1.025	0.760
leave-one-out__sample__crps	0.337	0.583	0.577
rolling window__interval__score	0.494	1.041	0.779
rolling window__sample__crps	0.347	0.632	0.570

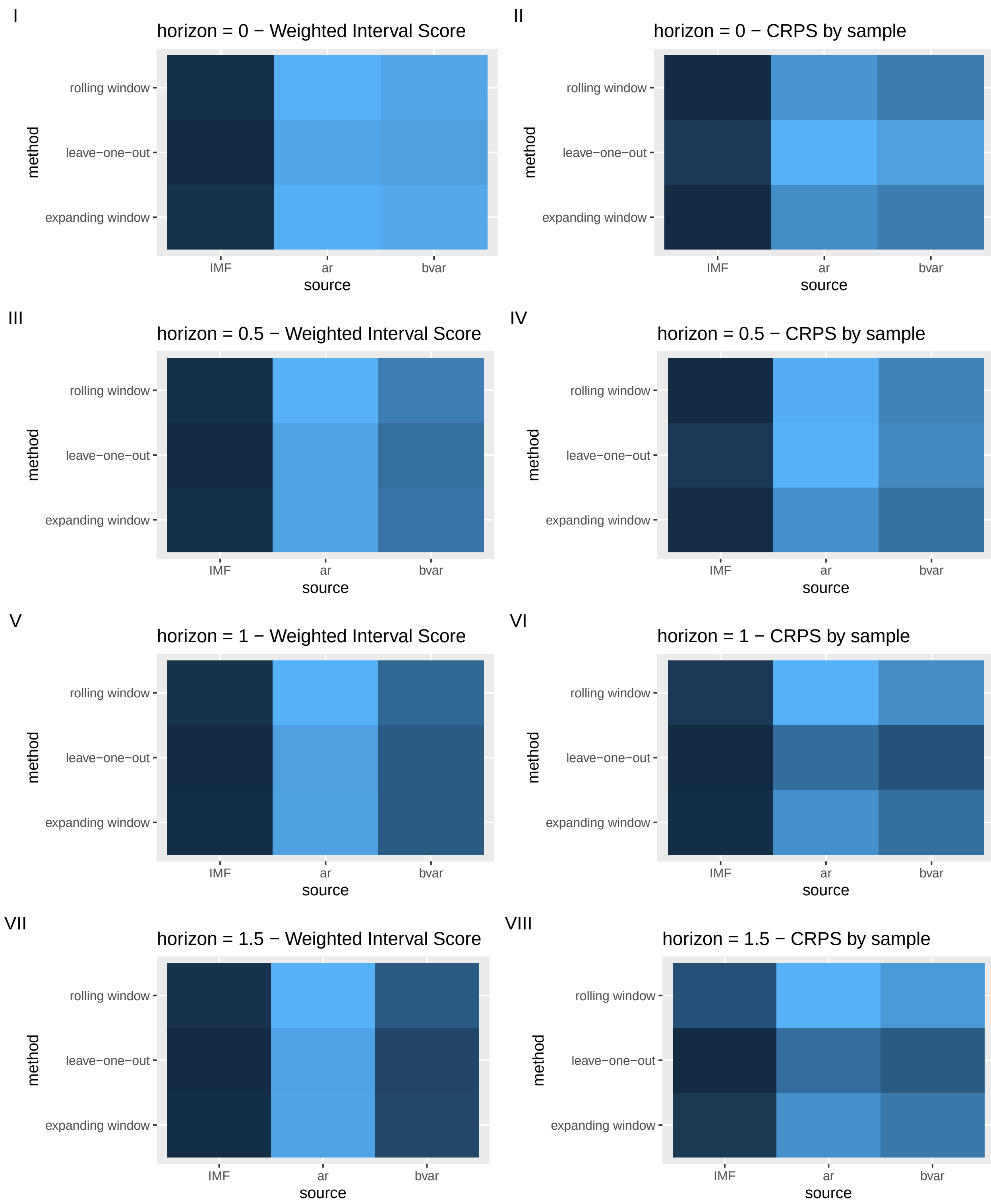
3.1 Inflation



3.2 GDP

	IMF	ar	bvar
horizon = 0			
expanding window__interval__score	0.241	0.301	0.298
expanding window__sample__crps	0.178	0.209	0.204
leave-one-out__interval__score	0.237	0.297	0.294
leave-one-out__sample__crps	0.183	0.219	0.214
rolling window__interval__score	0.240	0.302	0.297
rolling window__sample__crps	0.178	0.211	0.204
horizon = 0.5			
expanding window__interval__score	0.416	0.540	0.493
expanding window__sample__crps	0.298	0.383	0.358
leave-one-out__interval__score	0.411	0.540	0.489
leave-one-out__sample__crps	0.310	0.408	0.377
rolling window__interval__score	0.416	0.554	0.504
rolling window__sample__crps	0.297	0.405	0.373
horizon = 1			
expanding window__interval__score	0.837	1.090	0.942
expanding window__sample__crps	0.640	0.822	0.763
leave-one-out__interval__score	0.831	1.086	0.943
leave-one-out__sample__crps	0.634	0.759	0.709
rolling window__interval__score	0.851	1.122	0.970
rolling window__sample__crps	0.663	0.875	0.815
horizon = 1.5			
expanding window__interval__score	1.045	1.288	1.102
expanding window__sample__crps	0.790	0.937	0.897

leave-one-out_interval_score	1.038	1.284	1.099
leave-one-out_sample_crps	0.765	0.882	0.849
rolling window_interval_score	1.056	1.312	1.143
rolling window_sample_crps	0.831	0.985	0.949


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

4 Coverage, by target, methods and source