

Metric	Explanation
WIS (Weighted) interval score	<p>The weighted interval score (smaller values are better) is a proper scoring rule for quantile forecasts. It converges to the continuous ranked probability score (which itself is a generalisation of the absolute error to probabilistic forecasts) for an increasing number of intervals. The score can be decomposed into a dispersion (uncertainty) component and penalties for over- and underprediction. For a single interval, the score is computed as</p> $IS_{\alpha}(F, y) = (u - l) + \frac{2}{\alpha} \cdot (l - y) \cdot 1(y \leq l) + \frac{2}{\alpha} \cdot (y - u) \cdot 1(y \geq u),$ <p>where <math>1()</math> is the indicator function, <math>y</math> is the true value, and <math>l</math> and <math>u</math> are the <math>\frac{\alpha}{2}</math> and <math>1 - \frac{\alpha}{2}</math> quantiles of the predictive distribution <math>F</math>, i.e. the lower and upper bound of a single prediction interval. For a set of <math>K</math> prediction intervals and the median <math>m</math>, the score is computed as a weighted sum,</p> $WIS = \frac{1}{K + 0.5} \cdot \left( w_0 \cdot  y - m  + \sum_{k=1}^K w_k \cdot IS_{\alpha}(F, y) \right),$ <p>where <math>w_k</math> is a weight for every interval. Usually, <math>w_k = \frac{\alpha_k}{2}</math> and <math>w_0 = 0.5</math>.</p> <p>Its proximity to the absolute error means that when averaging across multiple targets (e.g. different weeks), it will be dominated by targets with higher absolute values.</p>
Interval coverage	<p>Interval coverage is a measure of marginal calibration and indicates the proportion of observed values that fall in a given prediction interval range. Nominal coverage represents the percentage of observed values that should ideally be covered (e.g. we would like a 50 percent prediction interval to cover on average 50 percent of the observations), while empirical coverage is the actual percentage of observations covered by a certain prediction interval.</p>
Bias	<p>(Relative) bias is a measure of the general tendency of a forecaster to over- or underpredict. Values are between -1 and 1 and 0 ideally. For continuous forecasts, bias is given as</p> $B(F, y) = 1 - 2 \cdot (F(y)),$ <p>where <math>F</math> is the CDF of the predictive distribution and <math>y</math> is the observed value.</p> <p>For quantile forecasts, <math>F(y)</math> is replaced by a quantile rank. The appropriate quantile rank is determined by whether the median forecast is below or above the true value. We then take the innermost quantile rank for which the quantile is still larger (under-prediction) or smaller (over-prediction) than the observed value.</p> <p>In contrast to the over- and underprediction penalties of the interval score it is bound between 0 and 1 and represents a general tendency of forecasts to be biased rather than the absolute amount of over- and underprediction. It is therefore a more robust measurement.</p>