Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments

Discussion

(Matei Demetrescu)

universitätbonn

Calibration testing

- The distribution of forecast errors should match the assumptions of the forecasting model.
- One should therefore check for mismatches.
- In this respect, density forecasts are more demanding than point forecasts.

In particular,

- the probability integral transformations (PITs) should be independent uniformly distributed,
- and inverse normal transformations thereof (INTs) should be Gaussian.

Checking normality

- Existing tests for normality focus on particular sample moments[1] such as
  - mean and variance (Berkowitz 2001)
  - skewness and kurtosis (Jarque/Bera 1980, Bai/Ng 2005)
- No consistency in general, unless "many" moments considered,
- but χ^2 asymptotics.

- In the case where serial dependence is allowed for (say h-step ahead forecasts),
- estimation of a long-run covariance matrix is required.

[1] Classical goodness-of-fit tests are not a popular choice.

Main contributions

- Existing procedures test for normality, but INTs are *standard normal*.
- So relevant information could be found in the first two moments as well!
- Moreover, uncentered sample moments are easier to work with.

Along these lines,

- some of the involved long-run covariances are zero;
- and using standardized PITs (S-PITs) instead of INTs are a further improvement.

Open issues

- Understand finite-sample behavior
- Power against certain nonstationarities
- Effect of model estimation?

Additional Monte Carlo experiments deliver some answers.

Size distortions

Size (5% nominal), testing $y_t \sim \mathcal{N}id\,(0,1)$, simple covariance matrix estimator.

T	$\widehat{\mu}_{34}$	$\widehat{\alpha}_1^{int}$	$\widehat{\alpha}_{12}^{int}$	$\widehat{\alpha}_{123}^{int}$	$\widehat{\alpha}_{1234}^{int}$	$\widehat{\alpha}_1^{s\cdot pit}$	$\widehat{\alpha}_{12}^{s\cdot pit}$	$\widehat{\alpha}_{123}^{s\cdot pit}$	$\widehat{\alpha}_{1234}^{s\cdot pit}$
50	4.0	4.7	9.8	16.9	42.7	4.7	5.4	7.2	9.2
100	8.6	5.3	7.8	13.5	33.0	5.0	5.7	6.5	7.5
250	10.3	5.4	6.7	9.6	21.8	5.7	5.9	5.9	6.0
500	9.7	4.8	5.5	6.7	14.9	4.6	4.9	5.2	5.0
1000	8.6	4.3	4.7	6.0	11.0	4.6	4.8	5.1	4.7

The problem is likely the high variance of the (long-run) covariance matrix estimator.[2]

[2] Cf. the improved behavior of $\widehat{\alpha}^0_{\cdot}$.

Variance breaks

- Changes in the volatility of the series,
- not captured by the forecasting model,
- affect density (if not point) forecasts.

Power (5% nominal), testing $y_t \sim \mathcal{N}id\,(0, 1/2.5)$ for $t < T/2$ and $y_t \sim \mathcal{N}id\,(0, 4/2.5)$ for $t \geq T/2$.

T	$\widehat{\mu}_{34}$	$\widehat{\alpha}_1^{int}$	$\widehat{\alpha}_{12}^{int}$	$\widehat{\alpha}_{123}^{int}$	$\widehat{\alpha}_{1234}^{int}$	$\widehat{\alpha}_1^{s\cdot pit}$	$\widehat{\alpha}_{12}^{s\cdot pit}$	$\widehat{\alpha}_{123}^{s\cdot pit}$	$\widehat{\alpha}_{1234}^{s\cdot pit}$
50	2.9	4.7	10.7	20.4	30.5	4.8	10.4	12.6	15.6
100	5.3	4.9	8.4	15.4	21.9	4.8	13.2	14.0	18.5
250	23.2	5.3	5.4	8.8	30.0	5.1	26.9	24.3	42.3
500	67.8	5.6	4.8	7.1	64.4	5.3	49.4	44.3	75.5
1000	97.4	5.3	5.4	6.6	95.7	5.4	83.2	77.9	97.5

Stationarity tests (Nyblom/Makelainen 1983, KPSS 1992) may give even better results.

Using residuals

The case more relevant in practice

Size (5% nominal), testing $y_t = \widehat{\varepsilon}_t / \widehat{\sigma}_\varepsilon$ with $\widehat{\varepsilon}_t$ OLS residuals from $z_t = 1 + 0.5\, z_{t-1} + \varepsilon_t$, $\varepsilon_t \sim (0, \sigma^2)$

T	$\widehat{\mu}_{34}$	$\widehat{\alpha}_1^{int}$	$\widehat{\alpha}_{12}^{int}$	$\widehat{\alpha}_{123}^{int}$	$\widehat{\alpha}_{1234}^{int}$	$\widehat{\alpha}_1^{s\cdot pit}$	$\widehat{\alpha}_{12}^{s\cdot pit}$	$\widehat{\alpha}_{123}^{s\cdot pit}$	$\widehat{\alpha}_{1234}^{s\cdot pit}$
50	3.8	0.0	0.0	3.4	32.3	0.0	0.0	0.7	1.2
100	8.4	0.0	0.0	2.3	21.0	0.0	0.0	0.8	0.9
250	10.0	0.0	0.0	1.1	11.0	0.0	0.0	0.7	1.1
500	9.9	0.0	0.0	0.8	7.0	0.0	0.0	0.4	1.0
1000	8.9	0.0	0.0	0.8	4.9	0.0	0.0	0.5	0.9

- Results robust to changes in AR coefficient.
- Using a long-run covariance matrix estimator does not change the essential message.
- Use KPSS-type statistic or correct critical values.

To sum up

- Raw moments are often more informative than just skewness and kurtosis.
- The approach is not restricted to INTs.
- Long-run covariance matrix estimation is an issue.
- The residual effect appears to be negligible,
  with one important exception (can be accounted for)