Testing Quantile Forecast Optimality

Marc-Oliver Pohle, Jack Fosten and Daniel Gutknecht

Discussion by Laura Coroneo (University of York)

12th ECB Conference on Forecasting Techniques

12-13 June 2023

Introduction

- Very interesting paper that proposes a misspecification test for multiple quantile forecasts at multiple horizons
- The test builds on multiple Mincer-Zarnowitz quantile regressions cast in a moment equality framework.
- Main test for the null hypothesis of autocalibration (optimality with respect to the information contained in the forecasts themselves).
- Two extensions:
  - Test for optimality with respect to larger information sets
  - Joint test for autocalibration for multiple series

Main idea: Quantile Mincer-Zarnowitz Test

- Definition: an h-step ahead forecast $\widehat{y}_{\tau,t,h}$ is autocalibrated if

  $$\widehat{y}_{\tau,t,h} = q_{y_t}(\tau|\mathcal{F}_{t-h})$$

  where $q_{y_t}(\tau|\mathcal{F}_{t-h})$ is the conditional τ-quantile of y_t given $\mathcal{F}_{t-h}$

- Quantile Mincer-Zarnowitz Test for autocalibration based on:

  $$q_{y_t}(\tau|\mathcal{F}_{t-h}) = \alpha_h^{\dagger}(\tau_k) + \widehat{y}_{\tau,t,h}\beta_k^{\dagger}(\tau_k) + \varepsilon_{t,h}(\tau_k)$$

  where $E[1\{\varepsilon_{t,h}(\tau_k) \leq 0\} - \tau_k] = 0$.

- The composite null hypothesis of autocalibration is given by

  $$H_0^{QMZ} : \{\alpha_h^{\dagger}(\tau_k) = 0\} \cap \{\beta_k^{\dagger}(\tau_k) = 1\}, \forall h \in \mathcal{H} \text{ and } \tau_k \in \mathcal{T}$$

  against the alternative $H_1^{QMZ} : \{\alpha_h^{\dagger}(\tau_k) \neq 0\}$ and/or $\{\beta_k^{\dagger}(\tau_k) \neq 1\}$ for at least some $h \in \mathcal{H}$ and $\tau_k \in \mathcal{T}$

Main Idea: Quantile Mincer-Zarnowitz Test

- Given P out-of-sample observations, estimate the coefficients by QR
- Under H_0^{QMZ}

$$\sqrt{P}\left(\begin{pmatrix} \widehat{\alpha}_1(\tau_1) \\ \widehat{\beta}_1(\tau_1) \\ \vdots \\ \widehat{\alpha}_H(\tau_K) \\ \widehat{\beta}_H(\tau_K) \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 1 \end{pmatrix}\right) = \sqrt{P}\widehat{\boldsymbol{m}} \to^d N(0, \boldsymbol{\Sigma}) \qquad (1)$$

- Proposed test statistic for the null hypothesis of autocalibration

$$\widehat{U}_{QMZ} = \sum_{s=1}^{S} \left(\sqrt{P}\widehat{m}_s\right)^2$$

- Use bootstrap critical values using the moving block bootstrap directly from the forecasts (given the assumption that $P/R \to 0$)
- Two extensions proposed:
  1. Augmented Quantile Mincer-Zarnowitz test (optimality with respect to larger information sets)
  2. Multivariate Quantile Mincer-Zarnowitz test (autocalibration for multiple series)

General Comments

- Very interesting paper that addresses an important issue (how to assess a forecasting method jointly for different quantiles and horizons)
- When considering multiple quantiles and multiple horizons, the number of conditions is large, so a Wald test that requires an estimate of the variance covariance matrix of the moment conditions can be unfeasible
- The proposed testing approach is feasible and has good properties

Comments: Bootstrap

1. **Joint resampling**:
   - In the autocalibrations test, the bootstrap is performed resampling from $\{y_t, \widehat{y}_{\tau,t,h}\}_{t=R+1}^{T}$ jointly for each $\tau = \tau_1, \ldots, \tau_K$ and $h = 1, \ldots, H$
   - In the multivariate extension, the bootstrap is performed by resampling from $\{y_t, (\widehat{\boldsymbol{y}}_{\tau,t,h})_{\tau=\tau_1,\ldots,\tau_K, h=1,\ldots,H}\}_{t=R+1}^{T}$
   - Why is it not performed jointly for all the quantiles and horizons also in the univariate tests?
2. **Conditioning variables**: how are the additional regressors treated in the bootstrap of the Augmented Quantile Mincer-Zarnowitz test?

Comments: Monte Carlo Simulations

1. **Autocorrelation**: the simulations currently only use one AR(1) coefficient value in the data generating process ($b = 0.6$) and two values for the forecasts ($\widetilde{b} = \{0.6, 0.8\}$).
   - How does the finite sample size of the test depend on b?
   - Does the finite sample power of the test depend only on $b - \widetilde{b}$, or also on b?
   - Does the block length selection l depend on b?
2. **Number of Conditions**: the simulations use 3 quantile levels and 4 horizons. How do the properties of the test change as we change the total number of conditions tested?
3. **Sample Size**: the smallest sample size considered is $P = 120$. What about smaller sample sizes?

Comments: Empirical Applications

- **Recalibration**: the paper mentions at various points the possibility of recalibration (i.e. to use the conditional quantiles obtained from the Mincer Zarnowitz regressions as recalibrated forecasts). This is not performed in the paper. It would be interesting to investigate the performance of the recalibrated forecasts.

- Minor points

  1. Clarification about how block lengths are selected. The Monte Carlo uses $l = 4, 8, 12$, the application to financial returns $l = 10$ and the application to macro variables $l = 4$.

  2. The possibility of looking at each horizon/quantile is quite interesting, but for the macro application only the contributions to the test statistic are reported. It would be interesting to see the p-values as well.

Conclusion

- The paper addresses an important issue in forecast evaluation that has been overlooked in the literature but is extremely important for practitioners and policymakers.
- Proposes feasible test statistics for the null hypotheses of
  - autocalibration
  - optimality with respect to a larger information sets
  - autocalibration for multiple series
- Hidden Gem: the appendix contains a Horizon Monotonicity Test that exploits the fact that under optimality, the expected quantile loss is monotonically non-decreasing in the forecast horizon.