Discussion of “Tails of inflation forecasts and tales of monetary policy” by Andrade, Ghysels and Idier

James Mitchell (Warwick Business School, University of Warwick)

“Inflation: Drivers and Dynamics Conference”, Federal Reserve Bank of Cleveland

30 September 2016
The authors introduce I@R measures of risk/uncertainty and density forecast asymmetry

- and test their utility (point forecasting inflation and interest rates) using individual SPF data

Forecast asymmetries important (especially post ZLB)

- central banks routinely talk about inflation risks and the balance of these
  - even when the statistical evidence for asymmetries in their density forecasts appears small; see Mitchell & Hall (2005, OBES); Knüppel & Schultefrankenfeld (2012, IJCB)

- the authors find time-varying asymmetries in the (US) SPF inflation densities - and that these explain inflation and interest rate outturns
Risk versus (Knightian) uncertainty

- Distinction too often blurred
  - *Imprecise probabilities*: Do we really think respondents 100% believe their (rounded?) forecasts in the outer bins?
  - MPC at BoE don’t even quantify 10% outer tail risks (see Mitchell & Weale, 2016)

- But I focus discussion on
  1. econometrics of how we aggregate and use the individual SPF densities
  2. how we might/should evaluate and then improve upon the authors’ I@R measures
Should we average probabilities or quantiles? I.e., should we extract the quantiles then aggregate? Or aggregate then extract the quantiles?

It matters theoretically and empirically... relates to the broader literature (incl. in management science) on forecast (or expert) density combination.
Should we average probabilities or quantiles? II

- Linear Opinion Pool commonly used to aggregate, including by FRB Philadelphia when publishing aggregate SPF density

\[ G(y)^{LOP} = \sum_{i=1}^{N} w_i G_i(y) \]

where \( G_i(y) \) are the cumulative distribution forecasts of forecaster \( i \),

where \( \sum_{i=1}^{N} w_i = 1 \)

- Option 1: could compute \( I@R(p) \) from \( G(y)^{LOP} \)

- But authors, in effect, choose to aggregate the \( N \) forecast densities using the “Quantile Opinion Pool” (Galvao et al., 2016; Lichtendahl et al., 2013 Management Science; Busetti, 2016)
QOP combines the forecasters’ quantile functions

\[ G(y)^{-1} = \sum_{i=1}^{N} w_i G_i^{-1}(p) \]

where \( p \) is the probability level

- Authors’ Option 2: \( I@R(p) = \sum_{i=1}^{N} w_i G_i^{-1}(p) \), with \( w_i = 1/N \)

QOP always delivers sharper densities than LOP (Lichtendahl et al. 2013); QOP can preserve distributional form of individual densities, unlike LOP
Why should $I@R(p) / ASY$ explain the conditional *mean* of inflation or the target interest rate?

Because of misspecification (poor calibration) of $I@R(p)$? Or asymmetric central bank preferences - as authors discuss?

- Looks like $I@R(p)$ is poorly calibrated, as 30% of outturns fall in the 10% tails
  - Test (conditional) coverage using Christoffersen (1998, IER) type LR tests?
  - Use *proper* scoring rules? Gneiting & Ranjan (2011, JBES) decompositions of CRPS; Diks et al. (2011, JoE) censored likelihood

Shouldn’t $I@R(p)$, if well-calibrated, explain the $p$-th quantile not the mean inflation outturn?

- It is “optimal” to take the $p$-th quantile as your forecast if you have an asymmetric piecewise linear loss function (Gneiting 2011 JASA)
- Does this make sense for central banks?
Knüppel & Schultefrankenfeld (2012, IJCB) evaluate “risk” (Pearson mode skewness) forecasts via a modified Mincer-Zarnowitz regression

\[
\left( \frac{\pi_{t+h} - \text{mode}_{t+h|t}}{\sigma_{t+h|t}} \right) = \alpha + \beta \left( \frac{\text{mean}_{t+h|t} - \text{mode}_{t+h|t}}{\sigma_{t+h|t}} \right) + \epsilon_{t+h}
\]

with forecast “optimality” (but under quadratic loss) implying \( \alpha = 0 \) and \( \beta = 1 \)

- i.e., if the point forecast is the mode, then upside (downside) risks should on average be followed by outturns that are greater (less) than the point forecasts

- What’s the evaluation test for the authors’ Bowley-based risk measure, \( ASY \)?
Can we improve $i@R$, ASY...?

- Use measures of “risk” that can be directly evaluated (ex post) using calibration tests?

- And/or should, like the threshold rather than quantile based risk measure of Kilian and Manganelli (2007, JMCB), the authors’ risk measure relate to the user’s risk preferences?
Given the errors frequently made when forecasting second moments, can forecasters in practice produce meaningful forecasts of third moments?

Consider alternative combination strategies, including trimming?

- Overconfidence. Isn’t this an argument for use of the LOP rather than the QOP? Galvao et al. (2016) find a Beta Opinion Pool works well...

Consider alternative weights on the individual forecasters?

- Opschoor et al. (2014) choose $w_i$ to maximise the accuracy of tail forecasts
Implications for macro models

- Need increasingly to move beyond (approximately) Gaussian models that deliver (approximately) symmetric forecast densities

Mitchell ("Inflation: Drivers and Dynamics Conference", Federal Reserve Bank of Cleveland)