Discussion of “Tests of Equal Predictive Ability with Real-Time Data” by T. E. Clark and M.W. McCracken

Juri Marcucci
Bank of Italy

5^{th} ECB Workshop on Forecasting Techniques
Forecast Uncertainty in Macroeconomics and Finance
Frankfurt, December 1, 2007

Forecasting with real-time data may be troublesome

- RTD affect forecasting in many dimensions:
  - Data inputs
  - Estimated parameters $\widehat{\beta}$
  - Model (data revision may change model specification)
- Which Actual should we use for evaluation?
  Many choices:
  - 1^{st} release (2^{nd} , 3^{rd})
  - 4 quarters later
  - Last benchmark (last snapshot before benchmark revision)
  - Latest available

Overview

- **Important question**:
  What are the asymptotic and finite-sample properties of tests of Equal Predictive Ability (EPA) when forecasters use Real-Time Data (RTD)?
- **Direct Multi-step predictions**:
  - Non-Nested models
  - Nested models
- **Main point of the paper**:
  Suggest how to adjust tests of EPA to make correct out-of-sample inference in a framework with data revisions.
- **Main results:**
  - **ignoring data revision** can lead to **incorrect inference**
  - adjusted test with reasonable size and power properties
  - application to real-time predictive content of GDP/GAP for inflation

Some (not easy) notation

- If observables are subject to a *revision process*
  $\Rightarrow$ their statistical properties might **differ** from those of final
- Want to predict $y_{t+\tau}(t')$ with $x'_{i,s}(t)$ where $t' = t + \tau + r$ and $t = R, \ldots, R + P - \tau + 1$
- vintage horizon is fixed ($r \ll R$)
- Model i: $y_{s+\tau}(t') = x'_{i,s}(t)\beta^*_i + u_{i,s+\tau}(t)$
- Forecast errors $\Rightarrow \widehat{u}_{i,t+\tau}(t') = y_{t+\tau}(t') - x'_{i,t}(t)\widehat{\beta}^*_{i,t}$
  - Do you consider revisions only in y_t?
- $\overline{d} = (P - \tau + 1)\sum_{t=R}^{T} \widehat{u}^2_{1,t+\tau}(t') - \widehat{u}^2_{2,t+\tau}(t')$

Non-Nested Models

- $MSE - t(S_{dd}) = (P - \tau + 1)^{1/2} \frac{\overline{d}}{\sqrt{\widehat{S}_{dd}}}$
- However, with parameter estimation uncertainty $P^{1/2}\overline{d} \xrightarrow{d} N(0, \Omega)$ where $\Omega = S_{dd} + 2(1 - \pi^{-1} \ln(1 + \pi))(FBS_{dh} + FBS_{hh}BF')$ and $F = (-2Eu_{1,t+\tau}x'_{1,t}, -2Eu_{2,t+\tau}x'_{2,t})$
- $MSE - t(S_{dd})$ is asymptotically valid only when $F = 0$ (few special cases when forecast error is uncorrelated with predictors, same loss for estimation and evaluation)
- with **data revisions** $\Rightarrow$ population residuals $[y_{s+\tau} - x'_{i,s}\beta^*_i]$ and forecast errors $[y_{t+\tau}(t') - x'_{i,t}(t)\widehat{\beta}^*_{i,t}]$ **do not have the same covariance structure**.
- Thus if $E[y_{s+\tau} - x'_{i,s}\beta^*_i]x_{i,s} = 0$ we cannot say anything about $E[y_{t+\tau}(t') - x'_{i,t}(t)\widehat{\beta}^*_{i,t}]x_{i,t}(t) = 0$
- Must use $MSE - t(\Omega)$

Non-Nested Models (cont.)

- $\pi = 0 \Rightarrow \Omega = 0$
- $F = 0$ when x is unrevised and
  - y unrevised
  - revisions of y uncorrelated with x
  - final revised vintage of y used for evaluation
  - vintages of y redefined so that data releases are used both for estimation and evaluation
- Revision process modeled as a combination of **news** and **noise**
  - if data revision contain news, revisions are not forecastable
  - if data revision reduce to noise, revisions are forecastable

Non-Nested Models (cont.)

- DGP for Final data:

$$
\begin{aligned}
y_t &= \beta x_{1,t-1} + \beta x_{2,t-1} + e_{y,t} + v_{y,t} \\
x_t &= e_{x,t} + v_{x,t}
\end{aligned}
$$

- Initial estimates

$$
\begin{aligned}
y_t(t) &= y_t - v_{y,t} + w_{y,t} \\
x_t(t) &= x_t - v_{x,t} + w_{x,t}
\end{aligned}
$$

- Noise component w creates a non-zero correlation between real-time forecast errors and predictors

Nested models

- $MSE - t$ and $MSE - F = (P - \tau + 1)\frac{\overline{d}}{MSE_2}$.
- Clark and McCracken (2005) and McCracken (2006) show that with nested models these tests have **non-standard distributions** (CVs are suggested and bootstrap for multi-step ahead forecasts).
- With data revisions and nested models the asymptotic properties of these tests **change dramatically** (for example in the t-test, average loss differentials $\overline{d}$ is re-scaled by $\sqrt{P}$ instead of P)
- Thus, in contrast to previous finding, *can construct* a t-test that is *asymptotically standard normal* under H_0

Nested models (cont.)

- **Pros and cons:**
  - Both $MSE - t$ and $MSE - F$ **diverge** under H_0
  - no assumptions of correct specification of the model
  - can conduct asymptotically valid tests using neither the bootstrap nor non-standard tables
  - Need estimate of Ω
- With *data revisions* the asymptotic distribution of $MSE - t$ can differ from *Theorem 2* and can be **highly non standard** with complications due to nuisance parameters. (left for further research!)
- **Strong belief** that the approximations developed in the previous papers should **reasonably approximate** the true asymptotic distribution.

Monte Carlo and empirical application

- Monte Carlo for non-nested and nested case (3 DGPs)
- 1-step and 4-step ahead forecasts
- Assume a single revision published with 1 or 4 period delay.
- Revisions (news and noise) and DGPs calibrated and tailored to the inflation data used in empirical section (Aruoba, 2006)
- All long-run variances and covariances are estimated with Newey-West ($h = 2\tau$)
- $R = (40, 80)$ and $P = (20, 40, 80, 160)$

Some comments

- The **non-nested** case works. Promising results!
- In the **nested** case $MSE - t\,(\Omega)$ should be preferred, but the **motivations behind** are **not completely clear**. I think that some further work should be done.
- DGP for the revision process: news-to-noise ratio for DGP1 seems quite high (large Ω)

DGP1 $\frac{\sigma^2_{v,y}}{\sigma^2_{w,y}} = \frac{.9}{.2} = 4.5$ while $\frac{\sigma^2_{v,x}}{\sigma^2_{w,x}} = \frac{.3}{2} = .15$

DGP2 $\frac{\sigma^2_{v,y}}{\sigma^2_{w,y}} = \frac{.2}{.2} = 1$ while $\frac{\sigma^2_{v,x}}{\sigma^2_{w,x}} = \frac{.3}{.5} = .6$

DGP3 $\frac{\sigma^2_{v,y}}{\sigma^2_{w,y}} = \frac{.2}{.2} = 1$ while $\frac{\sigma^2_{v,x}}{\sigma^2_{w,x}} = \frac{.3}{.5} = .6$

Monte Carlo results: Non-nested models

Size:

- $MSE - t(S_{dd})$ seems **always oversized** for 1-step ahead especially for low π. **Well sized** for larger π. For 4-step ahead highly oversized except when $\pi \geq 1$.
- $MSE - t(\Omega)$ seems **correctly sized** in all cases and at both forecasting horizons. The only exception: for low π slightly oversized.

Power:

- Here at 1-step ahead the two tests have **similar power properties**. Surprisingly, at 4-step horizons the $MSE - t(S_{dd})$ seems **almost always more powerful** than $MSE - t(\Omega)$.
- Here I really don't see any gain in terms of power in using Ω instead of S_{dd}!
- In sum, in the non-nested case, your tests seem to work fine.

Monte Carlo results: Nested models

Size:

- $MSE - F[CM]$: **well sized** in almost all cases. Slightly oversized for small π.
- $MSE - t(\Omega)[N]$ always *oversized*
- $MSE - t(S_{dd})[N]$: *oversized*
- $MSE - t(S_{dd})[CM]$: *oversized* but not for low π

Power:

- $MSE - F[CM]$: **good power properties** especially for high π
- $MSE - t(\Omega)[N]$: **best power** $\forall\pi$
- $MSE - t(S_{dd})$: good power only in DGP1 with noise

Some comments on Monte Carlo results

- Why is $MSE - F$ working so well? (technically it's invalid with data revisions)
- Can you try also the DM-adjusted test (Clark and West, 2007)?
- Can't use $ENC - t$ or $ENC - F$? Good properties in Clark and McCracken (2005)
- In nested case some size distortion remains even with Ω. Only exception: DGP1 (e.g. P=80). How do you explain it?
- Why increasing P the size distortion does not reduce so much? Is this related to the truncation lag?

Some comments on Monte Carlo results (cont.)

- If predictors x_t are without noise there shouldn't be a problem and you could use unadjusted tests. What about **qualitative regressors**? (E.g. Consumer confidence or Business confidence indicators; not revised).
- **Power worsens** from 1-step to 4-step. This is in contrast with Clark and McCracken (2005) and some results I found (not RTD). With multi-step ahead power doesn't decrease so much. I would expect that if there are revisions 4-step ahead forecasts should be less affected by the noise due to the revision process. How do you explain this intuitively? Does this depend on the lag truncation parameter that you use to estimate the LR variance?

Empirical results with RT inflation forecasting

Non-Nested models:

- Adjusted and unadjusted tests have same size. However $MSE - t(S_{dd})$ is *more powerful* and *rejects* more often the null of EPA. $MSE - t(\Omega)$ always fails to reject.

Nested models:

- $MSE - t(\Omega)[N]$ is more powerful in Monte Carlo than $MSE - t(S_{dd})[N]$ but empirically **you fail to reject** in **3 cases** with sample 1970-2003 at 1-step ($t + 2$,$t + 5$ and $t + 13$) while $MSE - t(S_{dd})[N]$ rejects. Should it not be the reverse?
- Similar size properties but more powerful test fails to reject when the less powerful rejects. This is strange.
- AR(4) with GDP growth outperforms the naive AR(2).
- Interestingly $MSE - F$ shows **good properties**. If it rejects, either or both $MSE - t(S_{dd})[N]$ and $MSE - t(\Omega)[N]$ reject.
- However I might be wrong with the slanted font (not easy to see).

Features of the paper

- **Strengths:**
  - **First paper** to deal with asymptotic properties of tests of EPA with *real-time data*
  - Well done theoretical analysis with *good insights* for future empirical applications
- Weaknesses:
  - *Full asymptotic distribution* of tests not developed with **nested models** and data revision
  - Use of approximations of previous work to approximate a complicated unknown distribution in nested case
  - There is some room for some improvements (especially for the nested case).
  - Excellent paper!