Tests of Conditional Predictive Ability: Some Simulation Evidence

Michael W. McCracken

Discussion by Minchul Shin (FRB Philadelphia)

System Committee on Econometrics Meeting, 2019

This paper ...

[1] Provide data-generating-processes (DGPs) that are consistent with Giacomini and White (2006, *Econometrica*)'s theoretical framework

[2] Analyze finite-sample properties of Giacomini and White (GW, hereafter)'s testing procedures for

1. Unconditional predictive ability
2. Conditional predictive ability

where forecasts are generated with

1. Fixed-window estimation scheme
2. Rolling-window estimation scheme

In this discussion

I will talk about this paper in terms of ...

1. Relevance
2. Challenges

Then, I will list my comments on the paper

Relevance

Among 50 regular articles published over the last year in the *International Journal of Forecasting*:

- 25 papers perform statistical testing on equal-predictive-ability
- 20 papers report results based on Diebold-Mariano (DM)-type statistic with
  - Recursive-window estimation (11 papers)
  - **Rolling-window estimation (8 papers)**
  - **Fixed-window estimation (1 paper)**

Challenges

Forecast target: y_{t+h}

Two forecasts made at time t: $f_{1,t}$ and $f_{2,t}$

Define a loss differential

$$d_{t+h} = L(y_{t+h}, f_{1,t}) - L(y_{t+h}, f_{2,t})$$

where $L(y, f)$ is a loss function

Consider two types of null hypotheses for equal predictive ability

$$E[d_{t+h}] = 0 \quad \text{or} \quad E[d_{t+h}|\mathcal{F}_t] = 0$$

Challenges

Conditional predictive ability test:

$$E[d_{t+h}|\mathcal{F}_t] = 0$$

GW's theory is based on a specific form of $\mathcal{F}_t$

They assume that $f_{1,t}$ and $f_{2,t}$ are measurable with respect to $\mathcal{F}_t$

For 1-step-ahead prediction,

- (d_{t+1}) becomes a martingale difference sequence (MDS)
- $(d_{t+1} z_t)$ becomes a MDS where $z_t \in \mathcal{F}_t$
- Their asymptotic theory boils down to LLN/CLT for MDS

For h-step-ahead prediction, d_{t+h} is serially correlated only up to $(h-1)$ displacements

Challenges

Is this assumption realistic?

Assume a quadratic loss function,

$$d_{t+1} = (y_{t+1} - f_{1,t})^2 - (y_{t+1} - f_{2,t})^2$$

Simple arrangement shows that GW's null hypothesis holds when

$$E[y_{t+1}|\mathcal{F}_t] = \frac{1}{2}(f_{1,t} + f_{2,t})$$

For example, if $f_{1,t}$ is unbiased, then $f_{2,t}$ has to be unbiased.

It is hard to come up with such an example at least for me. But, the author came up with such an example

Challenges

Is this assumption realistic?

Assume a quadratic loss function,

$$d_{t+1} = (y_{t+1} - f_{1,t})^2 - (y_{t+1} - f_{2,t})^2$$

Simple arrangement shows that GW's null hypothesis holds when

$$E[y_{t+1}|\mathcal{F}_t] = \frac{1}{2}(f_{1,t} + f_{2,t})$$

For example, if $f_{1,t}$ is unbiased, then $f_{2,t}$ has to be unbiased.

It is hard to come up with such an example at least for me. But, the author came up with such an example

Victory!

Comment 1: HAC-Bandwidth

Comment 1: If the paper was really about documenting a finite sample properties of GW's test procedures, then it is much more natural to **use a rectangular kernel rather than a Bartlett kernel** to estimate the long-run variance for "conditional" EPA test

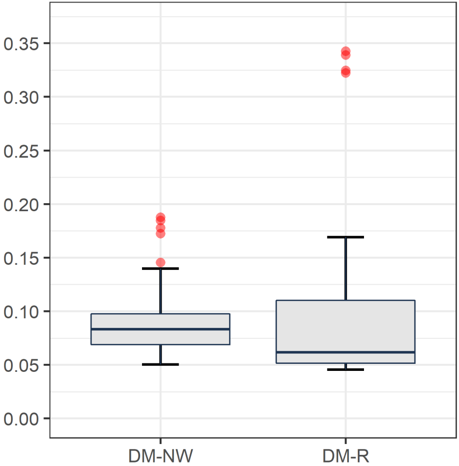
- Giacomini and White (2006) emphasized this in the paper. "Use HAC and truncate at $(h-1)$ for the conditional testing."
- Also emphasized in Giacomini (2010)

Under the conditional null hypothesis, it is better because ...

- Smaller size distortion due to less biased LRV estimation
- More powerful if the alternative entails a serial correlation

Comment 1: HAC-Bandwidth

Empirical distribution of actual sizes for all experiments considered in the author's paper (all 80 specifications), **nominal size = 5%**



- **DM-NW (This paper)**: Bartlett kernel with Newey and West's automatic bandwidth rule
- **DM-R (GW's original rule)** : Rectangular kernel truncated at $(h - 1)$

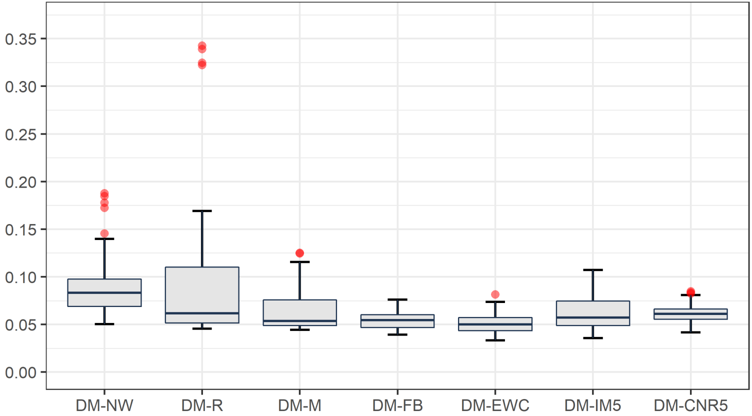
Comment 2: On size distortion

Comment 2: Most of size distortions that we saw from the author's exercises can be eliminated by using different approaches

- Recent tools from HAR-inference literature: Fixed-b asymptotics, t-statistic-based testing, randomization test using asymptotic symmetry, etc

- Accurate estimation of the long-run variance may not be an issue here (e.g., fixed-b asymptotics)

Comment 2: On size distortion



- DM-NW: This paper
- DM-R: GW's original rule
- DM-M: Harvey et al. (1997)
- DM-FB: Kiefer and Vogelsang (2005)
- DM-EWC: Muller (2004)
- DM-IM: Ibragimov and Muller (2010)
- DM-CNR: Canay, Romano, Shaikh (2017)

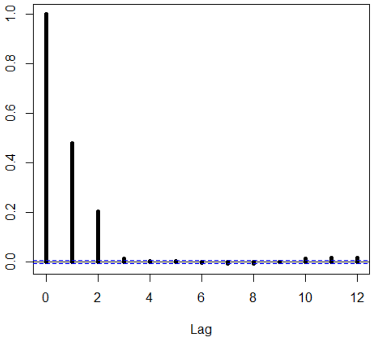
Comment 3: Interesting property of the DGP

Comment 3: Conditional-Rolling DGP generates a very strong conditional heteroscedasticity in the loss differential

Comment 3: Interesting property of the DGP

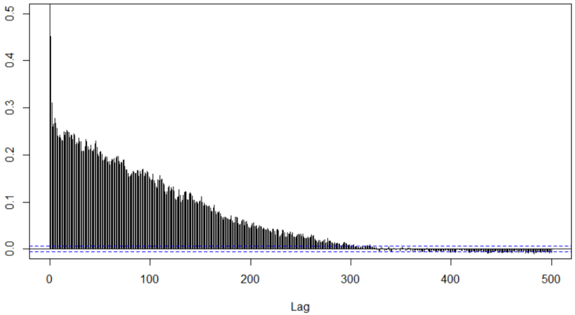
For example, R = 175; Rbar = 175; h = 3; with P = 1000000;

ACF of d_{t+h} is MA(2) as expected



Comment 3: Interesting property of the DGP

ACF of d_{t+h}^2



Fitting MA(2)-GARCH(1,1) gives me the following estimates

$$\omega = 0.00, \quad \beta_1 = 0.23, \quad \beta_2 = 0.81$$

where the GARCH equation is defined as

$$\sigma_t^2 = \omega + \beta_1 e_{t-1}^2 + \beta_2 \sigma_{t-1}^2$$

Comment 4: My personal view about GW

Conditional EPA testing:

$$E[d_{t+h}|\mathcal{F}_t] = 0$$

GW's theory is based on a specific form of $\mathcal{F}_t$

- For example, $f_{1,t}$ and $f_{2,t}$ are measurable with respect to $\mathcal{F}_t$
- Theoretically convenient, more powerful test, etc.
- It makes sense if you want to know whether d_{t+h} is predictable
- However, for this null to hold we have to impose a very strong restrictions on $f_{1,t}$ and $f_{2,t}$

Comment 4: My personal view about GW

GW's conditional testing also used for explanatory analysis:

- Is d_{t+1} statistically different from zero during recession?
- Is d_{t+1} is associated with some economic variables (z_t)?

One could perform a test based on

$$d_{t+h} = \beta_0 + \beta_1 z_t + \varepsilon_{t+h}$$

$E[\varepsilon_{t+h}] = 0$ and $E[\varepsilon_{t+h} \times z_t] = 0$

Then, just test $\beta_0 = 0$ and $\beta_1 = 0$ with HAR inference **WITHOUT “truncation at (h-1)”** admitting that ε_{t+h} can exhibit higher-order serial correlation

Conclusion

- This paper is very relevant for forecasters
- This paper attacks a challenging problem
- This paper offers several useful lessons for users of GW's tests