Discussion of 'Does modeling a structural break improve forecast accuracy?' by Tom Boot and Andreas Pick

Discussant: Ching-Wai (Jeremy) Chiu

10th ECB workshop on Forecasting Techniques

19 June 2018

Disclaimer

The views expressed in this discussion are mine and do not necessarily reflect the views of the Bank of England or its policy committees. All errors are mine.

Summary of the paper

- The paper provides a formal test to check whether modeling a break improves forecast accuracy
- Test the null hypothesis that h-step ahead MSFE using the full sample estimates is not bigger than the MSFE using the post-break sample
  - If the break date is known: Wald test with a Chi-squared distribution
  - If the break date is unknown: contribution of this paper

Summary of the paper

- Considering local breaks, the paper computes the difference between expected asymptotic MSFE of the partial sample forecast and that of the full sample forecast
- Under a couple of assumptions, the paper derives critical values which are (i) dependent (ii) independent of the break date
  - The test is near optimal although the break date is not consistently estimable
- The test is extended to testing against the combination of post-break and full sample forecasts

Summary of the paper

- A break in parameter does not necessarily imply a break in forecast
- Apply this procedure for 130 macroeconomic and financial time series from FRED-MD. The paper concludes that **few breaks are relevant for forecasting**
- The paper makes a nice contribution to the literature

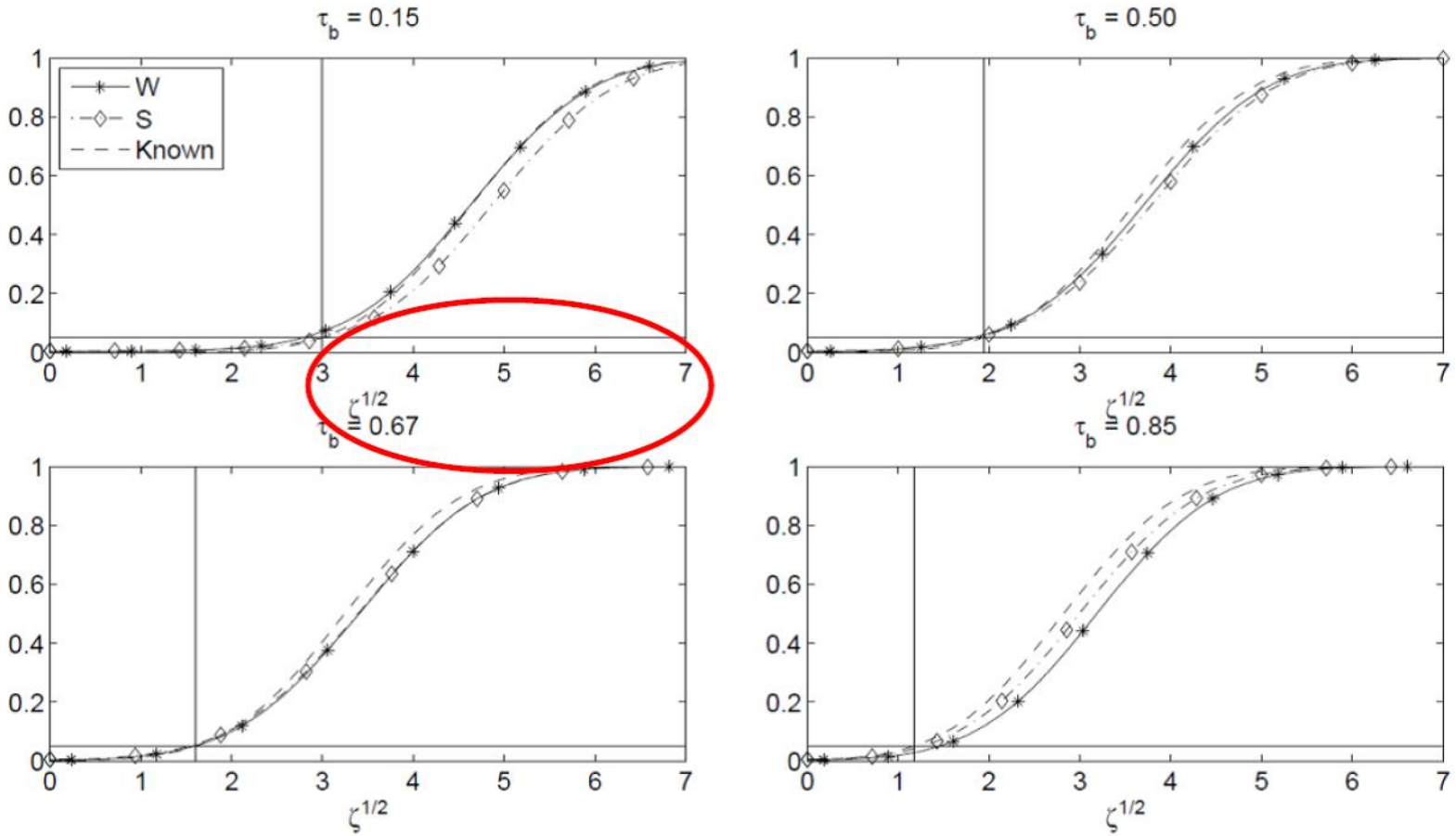
My comments

1. Power of the test
2. Are we missing some breaks in the empirical exercise?
3. Robustness of the test

Comment 1: Power of the test

- Asymptotic power increase with 'tail size' breaks

Figure 2: Asymptotic power when testing between a post-break and full-sample forecast at $\alpha = 0.05$



Comment 1: Power of the test

- Finite sample analysis shows low power

Table 3: Finite sample analysis: size and power when testing between post-break and full-sample forecast

		$T = 120$					$T = 240$					$T = 480$				
ρ	$\lambda \backslash \tau_b$	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
Wald-test (18)																
0.0	0	0.05	0.05	0.06	0.05	0.03	0.06	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.17	0.20	0.22	0.21	0.17	0.21	0.22	0.23	0.21	0.16	0.24	0.24	0.23	0.21	0.16
	2	0.43	0.48	0.52	0.53	0.47	0.52	0.54	0.55	0.53	0.48	0.57	0.56	0.56	0.55	0.49
0.3	0	0.04	0.05	0.06	0.05	0.03	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.13	0.17	0.21	0.21	0.17	0.18	0.20	0.22	0.20	0.16	0.22	0.23	0.22	0.21	0.16
	2	0.33	0.40	0.47	0.50	0.46	0.46	0.50	0.53	0.52	0.47	0.54	0.54	0.55	0.55	0.48
0.6	0	0.03	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.05	0.03	0.05	0.06	0.06	0.05	0.03
	1	0.08	0.12	0.19	0.20	0.16	0.13	0.17	0.20	0.20	0.15	0.18	0.20	0.22	0.21	0.15
	2	0.19	0.26	0.39	0.46	0.43	0.33	0.40	0.47	0.50	0.45	0.47	0.49	0.52	0.53	0.47
0.9	0	0.02	0.05	0.10	0.09	0.06	0.02	0.04	0.08	0.07	0.04	0.03	0.05	0.06	0.06	0.04
	1	0.04	0.07	0.17	0.24	0.20	0.04	0.08	0.16	0.21	0.16	0.07	0.11	0.17	0.20	0.15
	2	0.09	0.12	0.24	0.44	0.44	0.09	0.14	0.28	0.43	0.39	0.16	0.24	0.37	0.46	0.41

Comment 1: Power of the test

Issue of pinning down the break date

- Pesaran, Pick and Pranovich (2013): construct model weights robust to uncertainty about ...
  - ... break date: T^{-1}
  - ... break size in slope: T^{-2}
  - ... break size in variance: T^{-3}

Comment 1: Power of the test

Issue of pinning down the break date

- Pesaran, Pick and Pranovich (2013): construct model weights robust to uncertainty about ...
  - ... break date: T^{-1}
  - ... break size in slope: T^{-2}
  - ... break size in variance: T^{-3}

My thoughts:

1. Clements and Hendry (2005) argues that, to predict breaks, one needs an extra information set which describes factors in shifting relationship
   - Legislation / financial innovation / political factors
   - Expanding the information set may help pinning down the break date, hence improving the power
2. Exploit the correlations between time series
   - A break in real consumption/ investment may signal a break in output

Comment 1: Power of the test

Choice of nominal size

- Low test power is also attributable to the low nominal size, reflecting the econometrician's aversion towards Type I error

Comment 1: Power of the test

Choice of nominal size

- Low test power is also attributable to the low nominal size, reflecting the econometrician's aversion towards Type I error

My thoughts:

1. What is the basis of setting the size as 5%? The test will only pick up tail breaks, limiting its usefulness
2. Situations where we may be more tolerant of Type I error
   - Small sample
   - Cost of missing a break is high, especially when external evidence points to a break but the test says otherwise.
3. A decision framework to trade off Type I with Type II errors (Chiu, Hayes, Kapetanios and Theodoridis 2018)

Comment 1: Power of the test

Choice of MSFE as the objective function

- Pesaran and Timmermann (2007): trade-off between bias and variance
  - Short post-break samples substantially increase the variance of forecasts, even though they are unbiased

Comment 1: Power of the test

Choice of MSFE as the objective function

- Pesaran and Timmermann (2007): trade-off between bias and variance
  - Short post-break samples substantially increase the variance of forecasts, even though they are unbiased

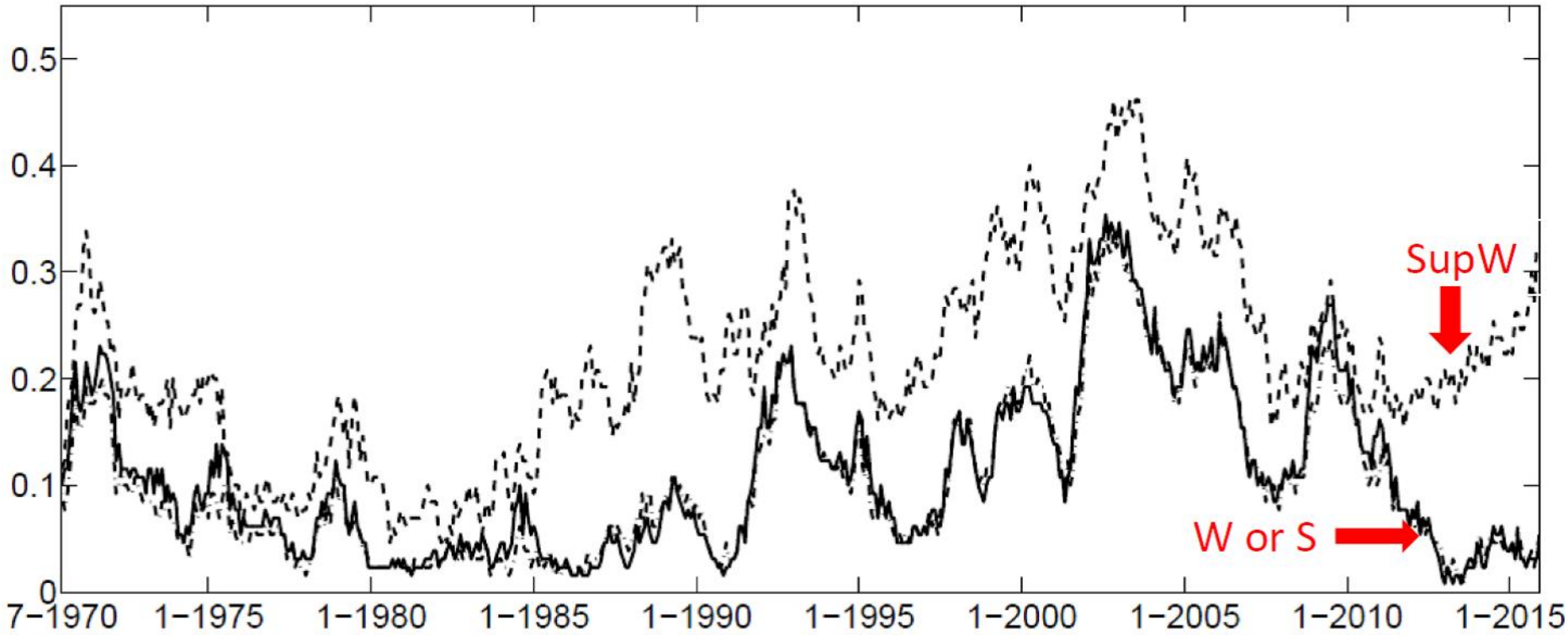
My thoughts:

1. From a policy perspective: pinning down the impact from a recent scenario
2. How will the power change if we only care about bias only?

Comment 2: Are we missing some breaks?

- Divergence between the two tests after 2010

Figure 8: Fraction of significant structural break test statistics over estimation samples AR(1)



Comment 2: Are we missing some breaks?

- The fraction of series with a significant break falls substantially after 2010
  - The paper explains on p.29: *as demonstrated in Figures 1 and 3, breaks in the early sample are less likely to be relevant for forecasting. However, Andrews SupW test does not use this information*.
- The two methods move in opposite direction
- SupW test indicates increasing number of series with breaks, but the current test says otherwise

Comment 2: Are we missing some breaks?

- The fraction of series with a significant break falls substantially after 2010
  - The paper explains on p.29: *as demonstrated in Figures 1 and 3, breaks in the early sample are less likely to be relevant for forecasting. However, Andrews SupW test does not use this information.*
- The two methods move in opposite direction
- SupW test indicates increasing number of series with breaks, but the current test says otherwise

My thoughts:

1. Could you provide further explanation? It looks to me that the test is discounting the SupW breaks
   - Does it reflect the quick fall in power towards the end of sample?
2. The productivity puzzle occurs during this period
   - It will be useful to see the results of sub-categories

Comment 3: Robustness of the test

- What causes a structural break?
  - One-off big shock
  - Breaks which are continuous and slowly changing
  - Or a combination of both
- It is possible to find more than one break in the sample

Comment 3: Robustness of the test

- What causes a structural break?
  - One-off big shock
  - Breaks which are continuous and slowly changing
  - Or a combination of both
- It is possible to find more than one break in the sample

My thoughts:

1. Is this test robust against different types of structural breaks? My impression is that this is built on (a), not (b).
   - It helps the reader if the paper can be more upfront about this issue
2. What should we do if there are multiple breaks?
3. How about testing against forecasts generated by AR models estimated on rolling windows?
   - Does a rolling window estimation provide an easier life to practitioners?

Conclusion

- A nice and interesting contribution to the literature
- Some thoughts about
  1. Power of the test
  2. Are we missing some breaks in the empirical exercise?
  3. Robustness of the test