BANK FOR INTERNATIONAL SETTLEMENTS

Concluding Conference of the Macro-prudential Research (MaRs) Network of the European System of Central Banks

Discussion of Session 5, Early Warning Models

Mathias Drehmann

Bank for International Settlements

Paper 1: News and Narratives in Financial Systems: Exploiting Big Data for Systemic Risk Assessment (David Gregory, et al)

- Extract "market sentiment" (excitement and anxiety), by counting words eg

Excitement	Anxiety
amaze	anxiety
attract	avoid
beneficial	bother

- Extract consensus using information entropy
- Show that text analysis is helpful

Paper 2: Identifying Excessive Credit Growth and Leverage (Alessi and Detken)

- Build binary classification trees
  - split data at each node into two sub-samples by finding best indicator/ threshold that minimizes wrong classifications
  - good in sample but not necessarily robust
- Grow random forest that bootstraps a large number of trees
  - Allows to measure importance of indicators by permutating the n-th indicator in out-of sample (out of bag) cases
- Use random forest to determine key indicators and then grow one tree
  - Very important variables: debt service ratio, bank credit to GDP, house prices

Comparing Different EWIs: Results from a Horse Race Competition Among Members of the MaRs Network (Alessi… Zigraiova)

- Compare 9 different EWI methods by running a horse race
  - Same banking crisis dates
  - Same window around crises (20 to 4 q before crises)
  - Same data set for explanatory variables (+ publically available data)
  - Authors were asked to use as many EU countries as possible and do out of sample test for early 1990s crises
  - Use same benchmark to compare models (AUROC, T1 and T2 errors,…)
    - (explain usefulness indicator better)
- Conclusion:
  - Multivariate models outperform simple models
  - Different models have different strength

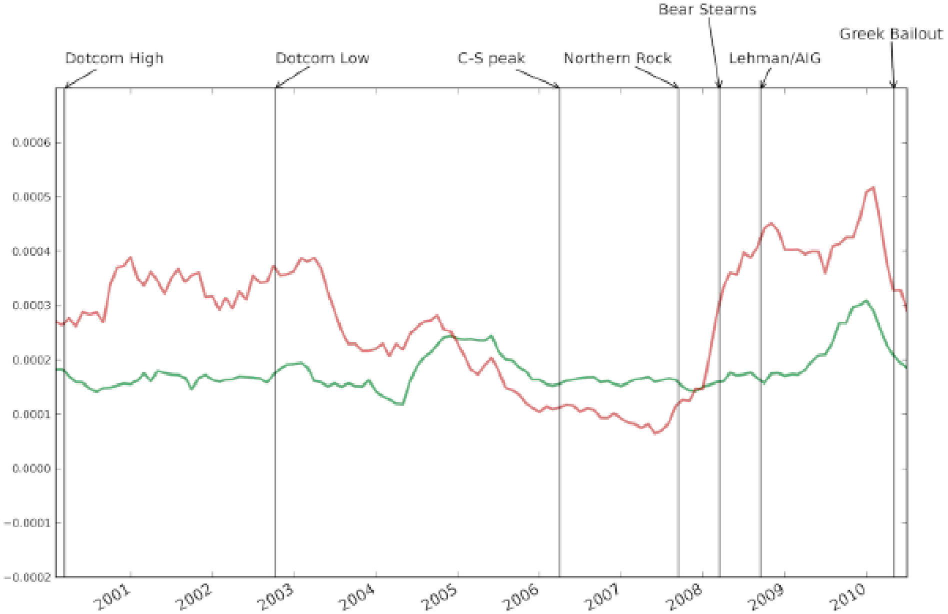
Discussion takes policy perspective

- What do policy makers want from EWIs?
  - Gauge build-up of vulnerabilities
  - Help to set macroprudential tools

    → Discussion assumes that aim is to steer economy wide tool (eg CCB) as other tools (eg sectoral tool CCBs) need different focus
- In Drehmann and Juselius (2014) we set out 5 criteria
  - EWI has a right timing (i.e. signals vulnerability ahead of crises)
  - EWI is stable (i.e. doesn't fluctuate from on-off and does not decrease ahead of crises)
  - Chosen EWI is should outperform (statistically) other indicators
  - EWI should be robust
  - EWI should be intuitive

David Gregory, et al

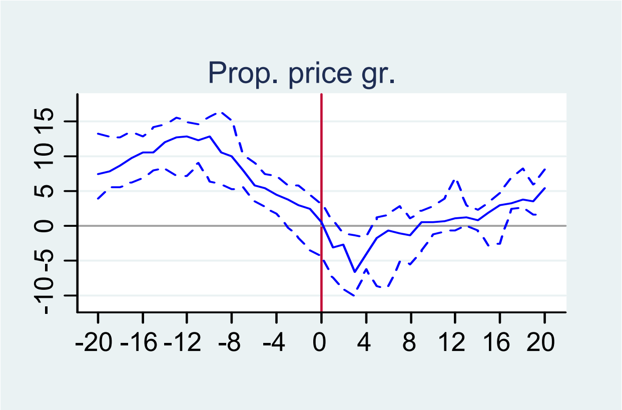
- Intuitive, yes but please more discussion for the novice
- Timing: excitement odd movement in 2010? anxiety – lagging?



- Looks stable and is robust to different sources
- Outperform not asked

Alessi and Detken

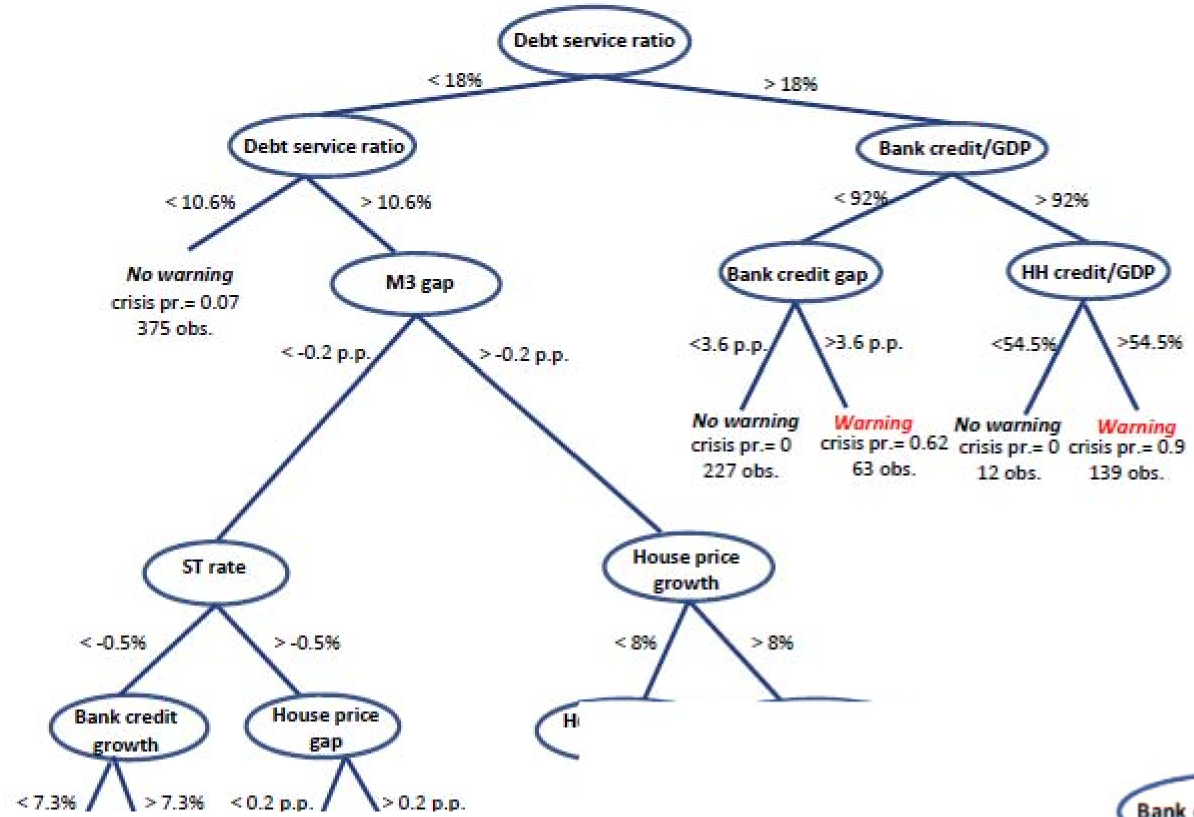
- Intuitive: Tree yes, but forest no (hence authors use tree)
- Timing:
  - Yes
- Stable, unclear as
  - Relevant periods are all lumped together
  - Year before crisis is not considered



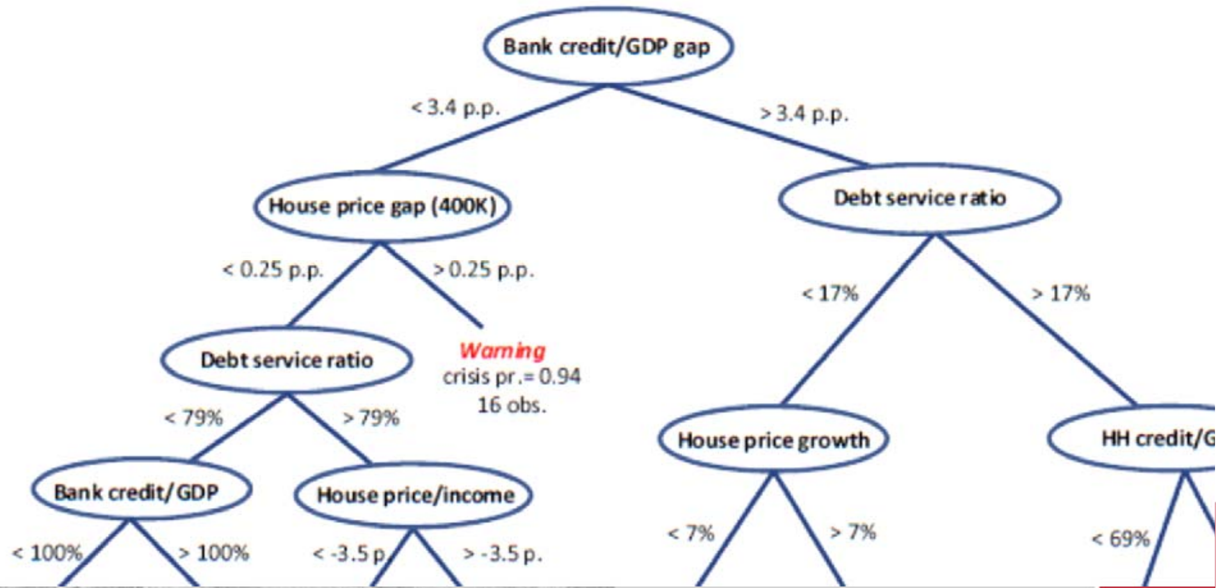
Alessi and Detken (II)

- Outperform
  - By design
- Robust, but
  - Level parameters (eg Credit /GDP) capture time trend and heavily reliant on tax codes and hence country specific
  - Odd country selection (eg many CEE countries missing but also Norway)
  - Why start in 1970 as big structural change since then?
  - Indicator importance changes
    - if sample is cut in 2006. Tree seems to be very different with eg DSR at root in full sample and M3 gap in pre 2006 (show tree)
    - If there are different preferences , big problem as cost\benefits of macropru interventions are not known
- General question: How robust is the forest with small sample of crises?

Preferred tree



Balanced preferences



Alessi… Zigraiova

- Intuitive
  - Some models more, some less
  - Paper itself unclear as big task (and 27 authors)
    - Idea: Short annex for each model discussing methodology and results
  - Add discussion on which indicators are (not) important across models
- Timing and Stability
  - As Alessi and Detken

Alessi… Zigraiova II

Performance seems very impressive

4-20 quarters Horizon	*AUROC*	*Type I error (%)*	*Type II error (%)*	*Absolute usefulness*
Baltussen et al.	0.889	6.2	31.1	0.314
Bush et al.	0.720	50.0	23.0	0.135
Antunes et al.	0.974	16.0	2.1	0.4097
Neudorfer, Sigmund	0.9928	5.5	2.93	0.395
Kauko	0.893	88.75	1.74	0.047
Behn et al.	0.931	7.3	22.0	0.354
Babecký et al.	0.856	7.9	59.9	0.161
Joy et al.	0.8416	0.0	42.5	0.288
Alessi, Detken	0.928	48.0	3.0	0.245

Table 4: In-sample performance statistics over the total horizon

Alessi… Zigraiova (II)

- "Outperformance" main question of the paper, but
  - How do models compare to benchmarks?
    - CCB guide, Shin's non-core liability ratio, Taylor and Schularick…
  - Show which model predicts\misses which crisis, given your cut-off criteria
    - In sample and out-of sample
  - Which model is statistically the best (see below)?
- Robustness should be strength, but
  - Require models to be estimated on the same sample not only same data
    - How does this affect individual models?
    - Homogenous sample allows for statistical comparison
  - How results change with different preferences
  - Are models robust to different crisis dating (eg Laeven and Valencia,..)
    - Is Spanish crisis included?