Improving early warning indicators for banking crises – satisfying policy requirements

by Mathias Drehmann and Mikael Juselius
Bank for International Settlements

Presentation at: 'Understanding Macro-prudential Regulation'
Norges Bank, Oslo, 29–30 November 2012

Discussion by Fabio Fornari
(ECB – DG-E, Capital Markets and Financial Structure Division)

What is the paper about and some thoughts

- Difficult task of anticipating events, crises
- At least one year ahead
- Consistently across episodes, which may be different while having a similar nature
- Also episodes likely to have different costs, either to policy makers but more importantly social costs
- Not even obvious what instruments policy makers have to change events, how long it takes for them to act, how long it takes for their actions to produce effects
- Difficult to control for events between quarter t-4, when actions should have been taken, and t
- All in all….crises continue to occur….it would mean something…..or not?

In fact early warning indicators (EWI) should

- … be preferences free
- … have right timing
- … be persistent and consistent
- … be understood by policymakers
- … be robust
- (a hard task, as you can guess)

ROC (take recessions as 'event')

y(t) is a real valued scalar and together with the threshold λ defines a binary prediction recession when $y(t) \geq \lambda$ and an expansion when it is below λ.

So define

True positive rate $TP(\lambda)=P(y(t)\geq\lambda|S(t)=1)$

False positive rate $FP(\lambda)=P(y(t)\geq\lambda|S(t)=0)$

The ROC curve is the plot of all the possible combinations $TP(\lambda)$ and $FP(\lambda)$ when λ varies on the real line (simple and effective tool…)

ROC

In general, there may be different benefits and costs associated with making accurate predictions and errors; hence the overall utility of the classification can be expressed as (see Baker and Kramer, 2007):

$$\begin{aligned} U(r) &= U_{11} ROC(r)\pi + U_{01}(1 - ROC(r))\pi + \\ &\quad U_{10} r(1-\pi) + U_{00}(1-r)(1-\pi) \end{aligned} \tag{1}$$

$$U_{11}\frac{dROC}{dr}\pi - U_{01}\frac{dROC}{dr}\pi + U_{10}(1-\pi) - U_{00}(1-\pi) = 0$$

or, rearranging

$$\frac{dROC}{dr} = \frac{U_{00} - U_{10}}{U_{11} - U_{01}}\frac{(1-\pi)}{\pi}. \tag{2}$$

That is, the optimum is that point where the slope of the ROC curve equals the expected marginal rate of substitution between the net utility of accurate expansion and recession prediction.

ROC

Underlying the classification problem is the view that the observations of Y_t reflect a mixture of two distributions. Specifically, let Z_t denote the observations of Y_t for which $S_t = 1$, with probability density function (pdf) given by f, and cumulative probability distribution (cdf) given by F. Similarly, let X_t denote the observations of Y_t for which $S_t = 0$ and with pdf given by g and cdf given by G. Then, the ROC curve can also be seen as a plot of $ROC(r) = 1 - G(F^{-1}(1-r))$ versus r, $r \in [0,1]$, so that the slope of the ROC curve is

$$\frac{dROC}{dr} = \frac{g(F^{-1}(1-r))}{f(F^{-1}(1-r))}$$

that is, the slope of the ROC curve is the likelihood ratio between f and g. Hence, expression (2) relates the likelihood ratio between the expansion and recession distributions with the expected marginal relative utility from correct classification.

ROC

Given U_{ij}, $i, j \in \{0, 1\}$, one can therefore determine the *optimal operating point* as the threshold c^* that meets the equilibrium condition (2). Under the assumption $U_{ii} = 1$ and $U_{ij} = -1$ and $\pi = 0.5$, the optimal operating point maximizes the distance between $TP(c)$ and $FP(c)$, which is the well-known Kolmogorov-Smirnov statistic (Kolmogorov, 1933; Smirnov, 1939). Clearly the assumption $\pi = 0.5$ is

ROC

A summary of all the trade-offs contained in the ROC curve and a commonly used measure of overall classification ability is the area under the ROC curve $(AUROC)$:

$$AUROC = \int_0^1 ROC(r)dr; \qquad AUROC \in [0.5, 1], \tag{3}$$

where it is clear that a perfect classifier has $AUROC = 1$ whereas a coin-toss classifier has $AUROC = 0.5$. A perverse classifier can generate an $AUROC < 0.5$ but then, by reversing the interpretation of the classifier's predictions from $S_t = 1$ when $Y_t > c$ to $S_t = 0$ (and vice versa when $Y_t < c$) the classifier would generate an $AUROC > 0.5$ so that for practical purposes an $AUROC = 0.5$ is the benchmark lower bound. This issue crops up in Section 5 and we show how it can be handled in practice there.

AUROC and its computation

- **It has also been shown that AUROC = P(Z>X)**
- **Hence a simple estimator for AUROC is**

$$\widehat{AUROC} = \frac{1}{n_0 n_1} \sum_{j=1}^{n_0} \sum_{i=1}^{n_1} \left\{ I(Z_i > X_j) + \frac{1}{2} I\left(Z_i = X_j\right) \right\}$$

- **where I(A) is the indicator function and n_k is the number of observation for state k. The last term is a tie-breaking rule.**
- **It has been also shown that AUROC above is a two-sample rank-sum statistics that can be cast as a Wilcoxon-Mann-Whitney U statistics.**

AUROC and its computation

- **Also important to know that**

$$\sqrt{n_1}\left(\widehat{AUROC} - P[Z > X]\right) \xrightarrow{d} N(0, \sigma^2)$$

$$\sigma^2 = \frac{1}{n_0 n_1}\left[AUROC(1 - AUROC) + (n_1 - 1)(Q_1 - AUROC^2) + (n_0 - 1)(Q_2 - AUROC^2)\right]^{1/2}$$

$$Q_1 = \frac{AUROC}{2 - AUROC}; Q_2 = \frac{2AUROC^2}{1 + AUROC}.$$

- **so that inference can be conducted according to standard Wald-type tests.**

Authors' main findings

- Look at a broad range of indicators
- And find that:
  - Credit-to-GDP gap best indicator for predicting crises 2-5 years in advance
  - Debt service ratios highly successful indicator for predicting crises 1-2 years in advance

Recall: why they choose ROC

- **To fully evaluate quality of a signal would need to know preferences of policymakers, which are unknown**
  - **What are costs of wrong signals (false positives)?**
  - **What are the benefits of correct signals (true positives)?**
- **Need to evaluate signalling quality independent of preferences**

Variables

- **Construct and test a range of potential early warning indicators building on Drehmann et al (2011).**
- **Select indicator variables from…**

Credit measures: Credit-to-GDP gap and real credit growth

Asset prices: Real property and equity price gaps and real property and equity price growth

Non-core bank liabilities (Hahm, Shin, and Shin (2012)): claims on banks held by foreign creditors relative to M2

GDP growth

History of financial crises

Variables

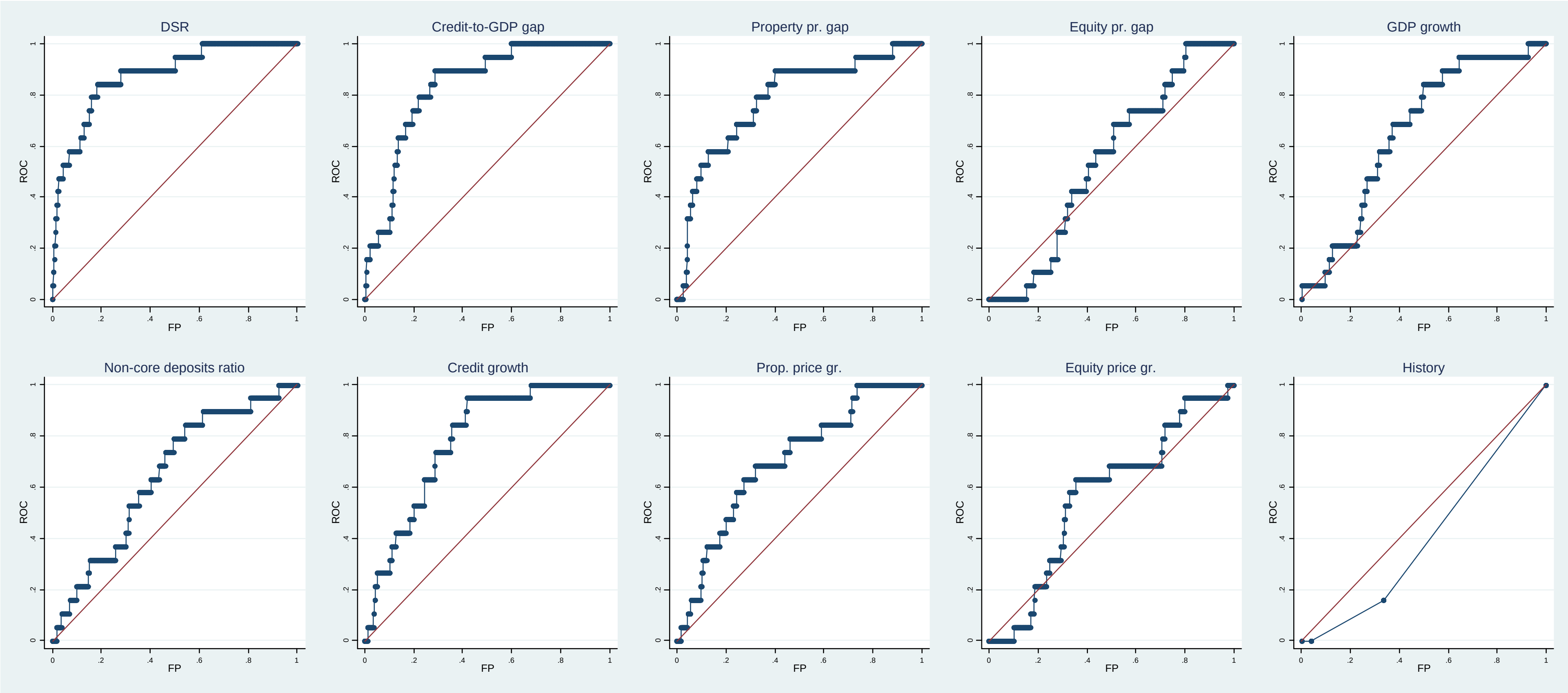
- **...and add one new measure:**

  Debt service ratio (DSR) (Drehmann and Juselius (2012)): interest payments and repayments on debt divided by income

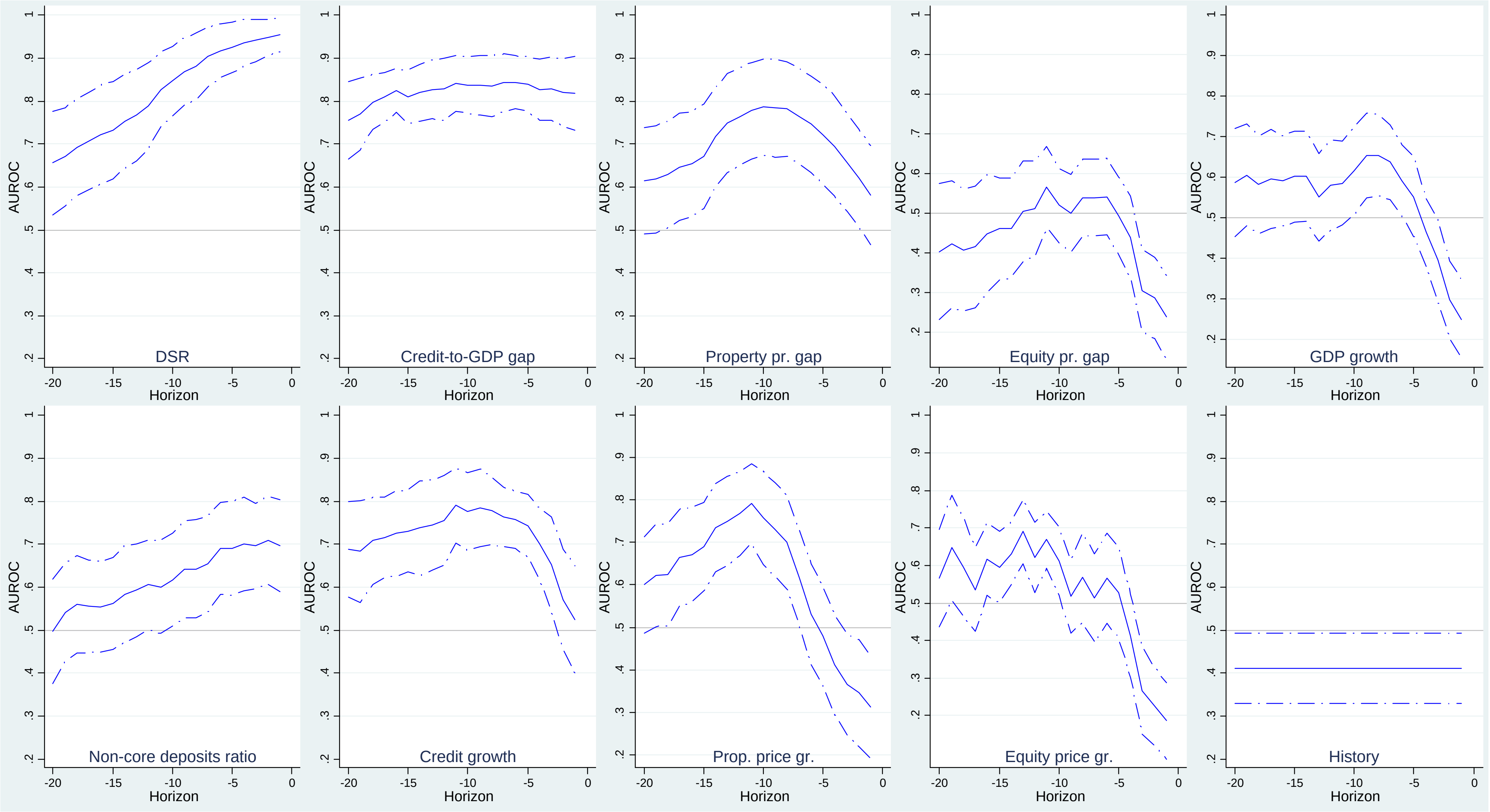
Samples, countries

- **Quarterly time-series data from 27 countries.**
  - The sample starts in 1980 for most countries and series, and at the earliest available date for the rest
  - Use balanced sample
- **Dating of systemic banking crises in Laeven and Valencia (2012)**
  - ignore crises which are driven by cross-border exposures
  - adjust dating for some crisis after discussions with CBs
- **Discard signals in the two years after the beginning of a crisis into account to avoid bias (Bussiere and Fratzscher (2006)).**

ROC curves for 2 year forecast horizon

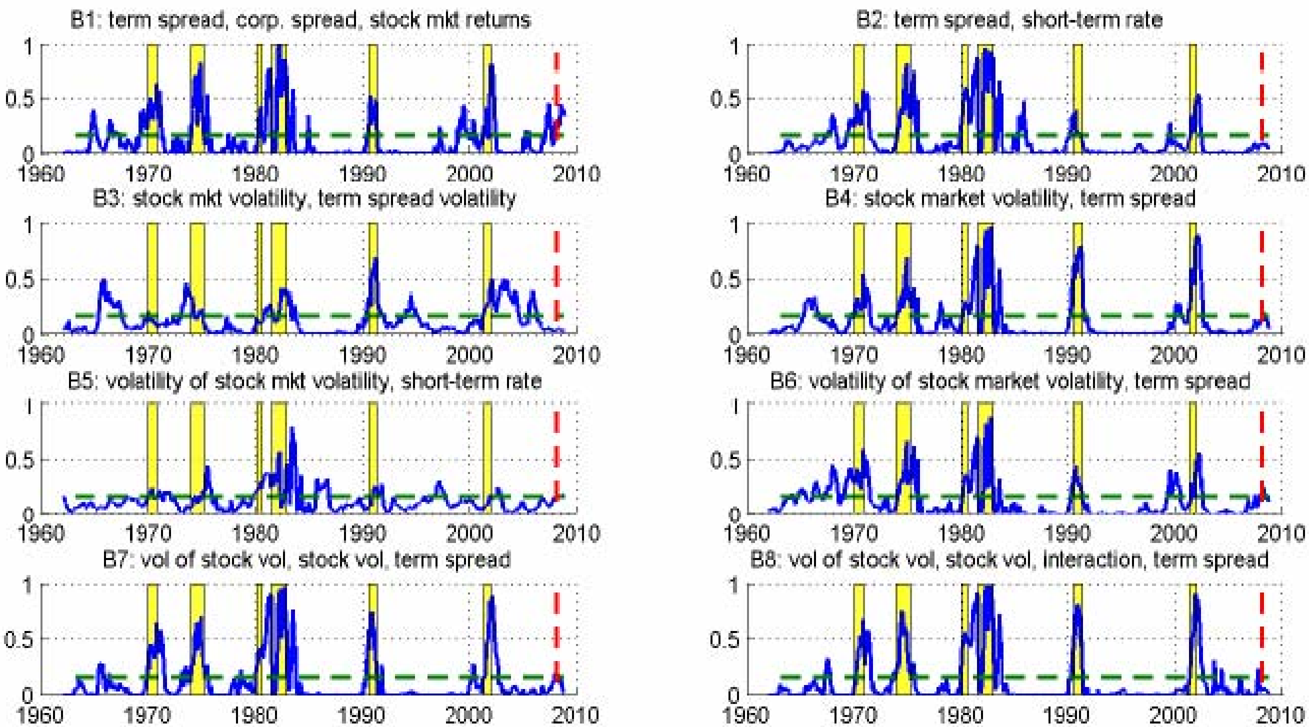


ROC curves over time: some indicators do ROC(K)!



Why not comparing to parametric models

Probabilities of recession



Choosing out of a number of models

- **Conditional predictive ability**. We use Giacomini-White testing strategy. Let $\Delta\epsilon_{t,k}^{i,j} \equiv |\epsilon_{t,k}^{i}| - |\epsilon_{t,k}^{j}|$. Regress $\Delta\epsilon_{t+1,k}^{i,j}$ on some vector of variables $h_{t,k}^{i,j}$, deemed to explain the failure of equal conditional predictive ability stemming from any two blocks:

$$\Delta\epsilon_{t+1,k}^{i,j} = \delta_k^{i,j} \cdot h_{t,k}^{i,j} + u_{t+1,k}^{i,j}, \quad t = M-1, \cdots, N-k, \tag{1}$$

where for any two predicting blocks i and j, and predicting horizon k, $\delta_k^{i,j}$ is a vector of constants, and, finally, $u_{t+1,k}^{i,j}$ is a residual term.

Adaptive decision rule

- In Monte Carlo experiments, Giacomini and White show that test has both reasonable size and power, once $h_{t,k}^{i,j} = [1 \ \Delta\epsilon_{t,k}^{i,j}]^\top$. We make this choice.

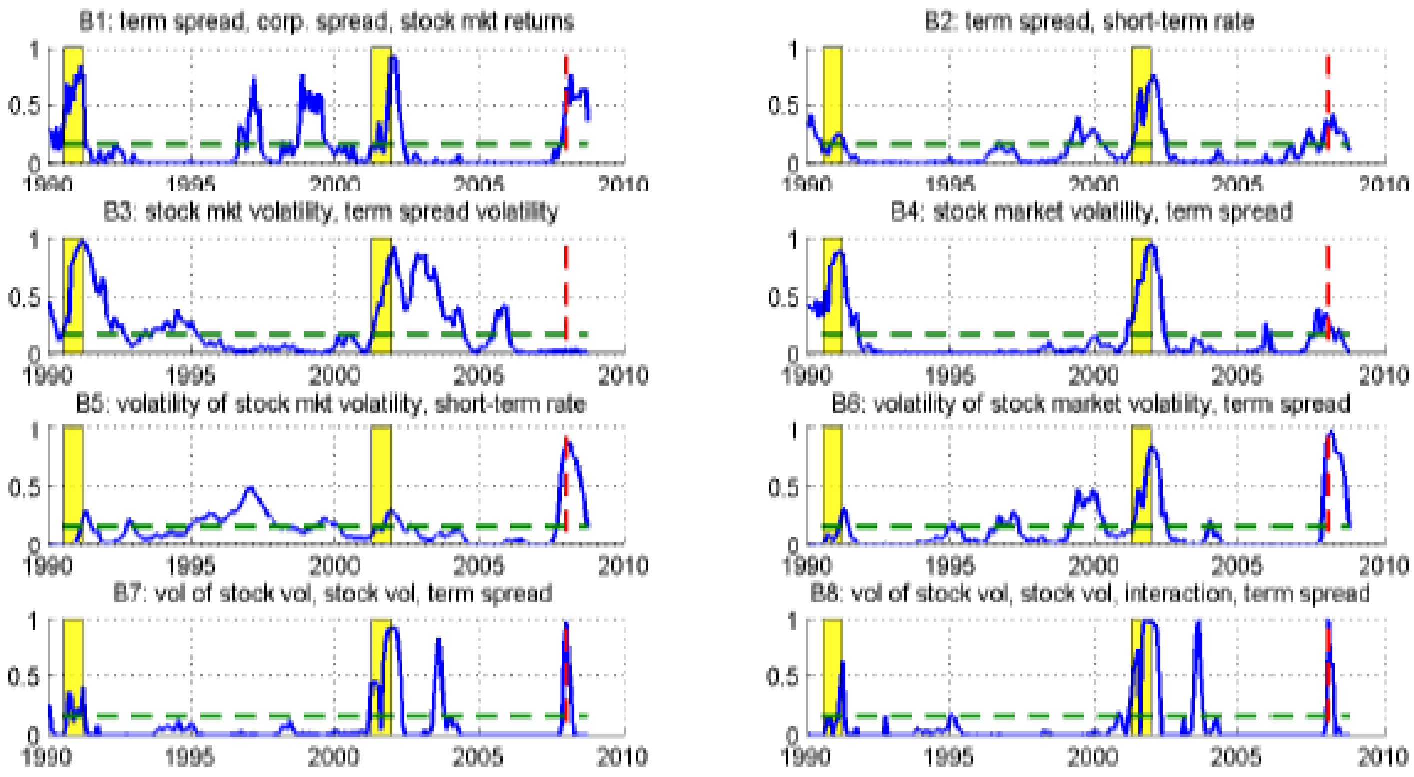
- The GW test can be used to implement an adaptive decision rule for selecting a predictive block over the others, thus exploiting the best conditional predictive power of any block. We report the frequency at which we reject block i for block j, over the entire out-of-sample period,

$$\frac{1}{N-M-1}\sum_{t=M-1}^{N-1}\mathbb{I}(E_N(\Delta\epsilon_{N+1,k}^{i,j})>0) \approx \frac{1}{N-M-1}\sum_{t=M-1}^{N-1}\mathbb{I}(\hat{\delta}_k^{i,j}\cdot h_{t,k}^{i,j}>0),$$

where $\mathbb{I}$ is the indicator function.

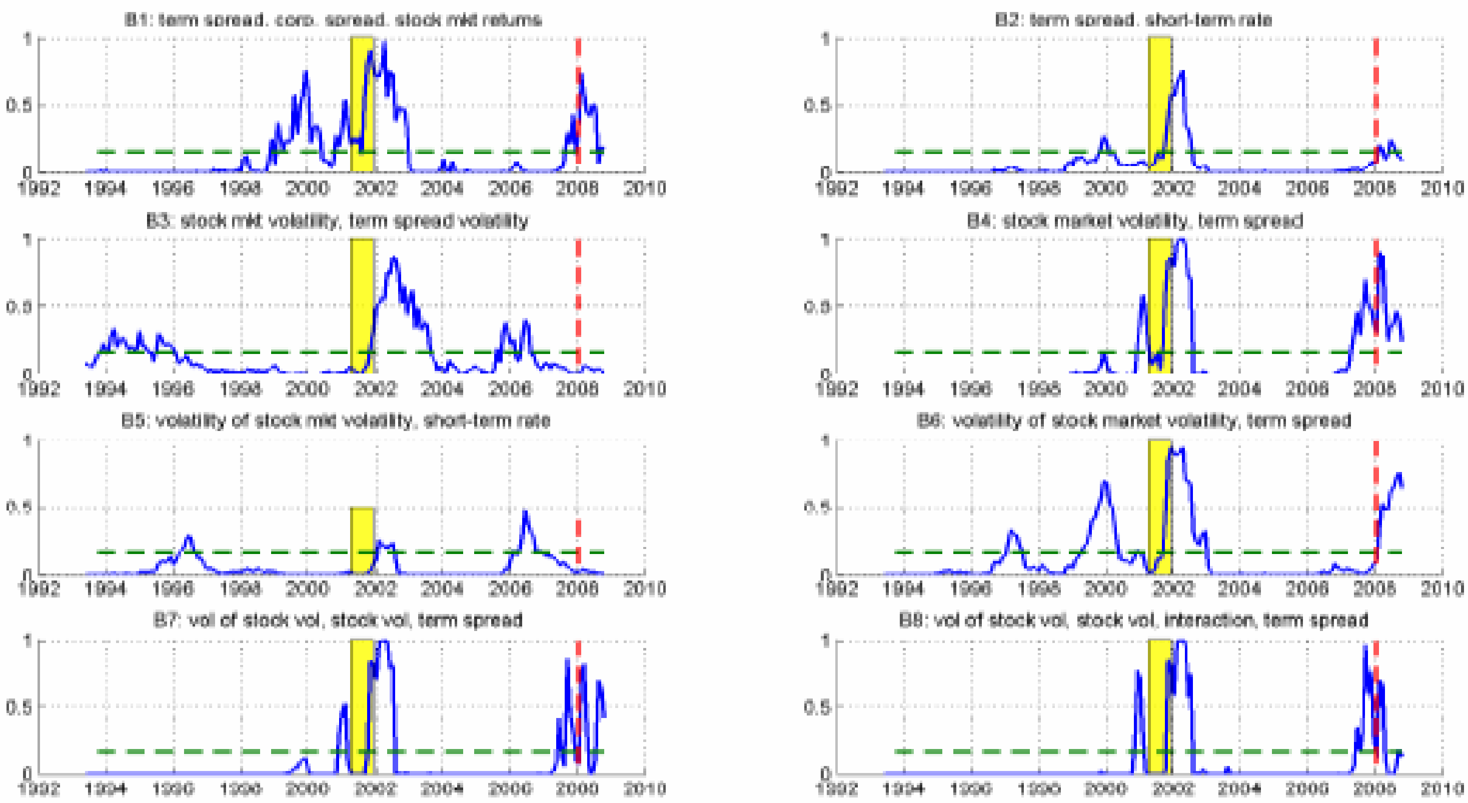
Out of sample coincident

Probit, out-of-sample, coincident



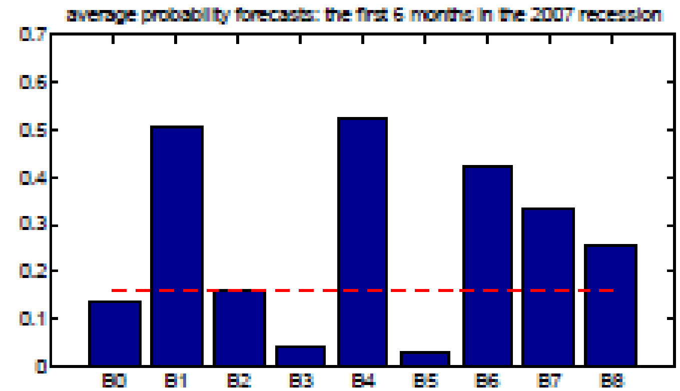
Out of sample, leading

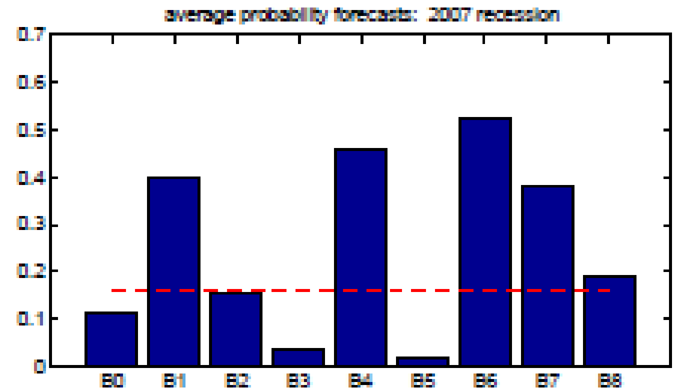
Probit, out-of-sample, six-month projections

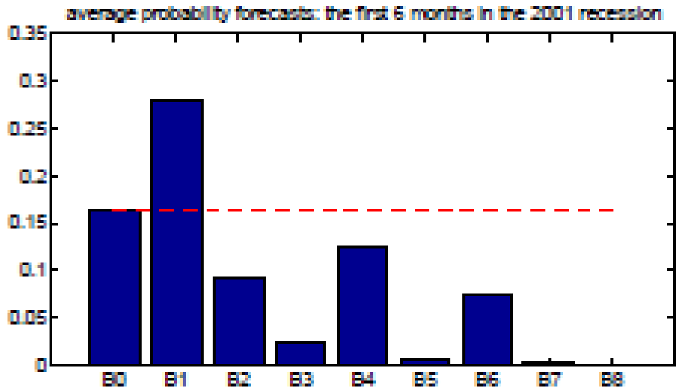


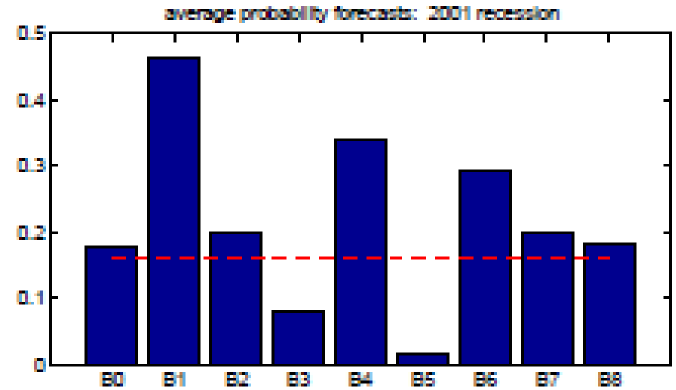
Comparing episodes

A tale of two recessions









Wrap up

- … very interesting paper
- … also with challenging evidence
- … I'd add
- … some comparison to parametric methods
- … some attempt to quantify costs for actions
- … in the end knowing that something may happen does not mean we are able to avoid it…depends too much on contingencies