International Journal of Forecasting 30 (2014) 759–780

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Evaluating early warning indicators of banking crises: Satisfying policy requirements

Mathias Drehmann, Mikael Juselius *

Bank for International Settlements, Centralbahnplatz 2, CH-4002 Basel, Switzerland

ARTICLE INFO

Keywords:
Early warning indicator
ROC
Area under the curve
Macroprudential policy

ABSTRACT

Ideally, early warning indicators (EWI) of banking crises should be evaluated on the basis of their performance relative to the macroprudential policy maker's decision problem. We translate several practical aspects of this problem — such as difficulties in assessing the costs and benefits of various policy measures, as well as requirements for the timing and stability of EWIs — into statistical evaluation criteria. Applying the criteria to a set of potential EWIs, we find that the credit-to-GDP gap and a new indicator, the debt service ratio (DSR), consistently outperform other measures. The credit-to-GDP gap is the best indicator at longer horizons, whereas the DSR dominates at shorter horizons.

1. Introduction

Early warning indicators (EWIs) are an essential component in the implementation of time-varying macroprudential policies, such as countercyclical capital buffers, which can help reduce the large losses associated with banking crises. EWIs in this context must not only have sound statistical forecasting power, but also need to satisfy several additional requirements. For instance, the signals need to arrive early enough that the resulting policy measures have time to be effective, and they need to be stable, as policy makers tend to react to trends. In general, deriving optimal empirical models for forecasting requires detailed knowledge of the underlying decision problem (e.g., Granger & Machina, 2006). However, such knowledge is currently not available in the context of macroprudential policies, as there is limited experience from which the expected costs and benefits could be estimated (CGFS, 2012). Nevertheless, it is still possible to incorporate the qualitative aspects of the policy maker's decision problem into the estimation and evaluation procedures for EWIs. Laying down such an approach and applying it to a range of EWIs is the main objective of this paper.

Given the difficulty of estimating the costs and benefits of macroprudential policies, the second-best option is to evaluate EWIs over a range of possible utility functions. As the optimal decision under a specific utility function implies a specific trade-off between Type I and Type II errors, one way to achieve this is to consider the full mapping between such trade-offs that a given EWI generates. This mapping is called the receiver operating characteristic (ROC) curve.[1] Going back to World War II, the ROC curve has a long tradition in other sciences (e.g., Swets & Picket, 1982), but its applications to economics are more scarce. Recent exceptions include the work of Berge and Jorda (2011), Cohen, Garman, and Gorr (2009), Gorr and Schneider (2011) and Jorda et al. (2011).

The ROC curve has several useful properties (e.g., Hsieh & Turnbull, 1996). In particular, the area under the curve (AUC) is a convenient and interpretable summary measure

* Corresponding author.
E-mail addresses: mathias.drehmann@bis.org (M. Drehmann), mikael.juselius@bis.org (M. Juselius).

[1] The ROC curve is a mapping of the false positive rate (Type II errors) to the true positive rate (the complement of Type I errors). The somewhat awkward name goes back to its original use in trying to differentiate noise from radar signals. A parallel way of expressing the signalling quality of an EWI is the correct classification frontier (e.g., Jorda, Schularick, & Taylor, 2011), which is more intuitive for optimal choice problems.

0169-2070/$ – see front matter © 2014 Published by Elsevier B.V. on behalf of International Institute of Forecasters.
http://dx.doi.org/10.1016/j.ijforecast.2013.10.002

of the signalling quality of a binary signal. AUCs can also be estimated easily. Parametric and non-parametric estimators are available as well as confidence bands and Wald statistics for comparing the AUCs of two signals, together with confidence bands and Wald statistics (e.g., Janes, Longton, & Pepe, 2009; and Pepe, Janes, & Longton, 2009).

Following this literature, we adopt the AUC as the primary metric for assessing and comparing the classification abilities of EWIs, and use it to embed macroprudential policy requirements in the evaluation process. In particular, we specify three additional criteria related to the timing, stability and interpretability of ideal EWIs of banking crises.

The appropriate timing is a crucial requirement for EWIs. On the one hand, macroprudential policies need time before they become effective (e.g., Basel Committee, 2010). On the other hand, signals which arrive at very early stages can also be problematic as policy measures are costly (e.g., Caruana, 2010). We therefore require that signals should arrive at least one and a half years but no more than five years ahead of a crisis. The stability of the signal is a second, largely overlooked, requirement. For one thing, policy makers tend to base their decisions on trends rather than reacting to changes in signalling variables immediately (e.g., Bernanke, 2004). A gradual implementation of policy measures may also allow policy makers to affect market expectations more efficiently, and to deal with uncertainties in the transmission mechanism (CGFS, 2012). Since EWIs that issue stable and persistent signals reduce the uncertainty regarding trends, they allow for more decisive policy actions. The final, less tangible, requirement is that EWI signals should be easy to interpret, as any forecasts, including EWIs, that do not "make sense" are likely to be ignored by policy makers (e.g., Lawrence, Goodwin, O'Connor, & Onkal, 2006; Onkal, Thomson, & Pollock, 2002).

In the empirical part of this paper, we use our approach to assess the performances of 10 different EWIs. We mainly look at the EWIs individually, but at the end of the paper we also consider how to combine them. Our sample consists of 26 economies, covering quarterly time series starting in 1980. The set of potential EWIs includes more established indicators such as real credit growth, the credit-to-GDP gap, growth rates and gaps of property prices and equity prices (e.g., Drehmann, Borio, & Tsatsaronis, 2011), as well as the non-core liability ratio proposed by Hahm, Shin, and Shin (2013). We also test two new measures: a country's history of financial crises, and the debt service ratio (DSR). The DSR was first suggested in this context by Drehmann and Juselius (2012), and is defined as the proportion of interest payments and mandatory repayments of principal to income. An important data-related innovation of our analysis is that we use total credit to the private non-financial sector, obtained from a new BIS database (Dembiermont, Drehmann, & Muksakunratana, 2013).

We find that the credit-to-GDP gap and the DSR are the best performing EWIs in terms of our evaluation criteria. Their forecasting abilities dominate those of the other EWIs at all policy-relevant horizons. In addition, these two variables satisfy our criteria pertaining to the stability and interpretability of the signals. As the credit-to-GDP gap reflects the build-up of leverage of private sector borrowers and the DSR captures incipient liquidity constraints, their timings are somewhat different. While the credit-to-GDP gap consistently performs well, even over horizons of up to five years ahead of crises, the DSR becomes very precise two years ahead of crises. Using and combining the information from the two indicators is therefore ideal from a policy perspective. Of the remaining indicators, only the non-core liability ratio fulfils our statistical criteria, but its AUC is always statistically smaller than that of either the credit-to-GDP gap or the DSR. These results are robust with respect to different aspects of the estimation, such as the particular sample or the specific crisis classification used.

The remainder of the paper is organised as follows. Section 2 relates the procedures for evaluating EWIs to the decision problem of the macroprudential policy maker. In particular, it introduces ROC curves and translates various additional policy requirements into statistical evaluation criteria. Section 3 discusses the data and introduces the potential EWIs. Section 4 evaluates and compares the signalling quality of the EWIs based on the criteria laid down in the previous sections and undertakes robustness checks. Finally, Section 5 concludes.

2. Evaluating EWIs based on policy requirements

When the purpose of a forecast is to guide a policy decision in an uncertain environment, the policy maker's preferences and constraints are important for the ex post evaluation of its forecasting performance, as these components define the loss function (e.g., Granger & Machina, 2006; and Pesaran & Skouras, 2002; and references therein). Equally, the preferences that implicitly correspond to standard statistical evaluation criteria rarely make sense in a specific policy context. For example, a comparison of alternative forecasts based on the squared error loss generally will not capture the economically relevant trade-off, even as an approximation, and therefore such a comparison is likely to be sub-optimal (Granger & Machina, 2006; Granger & Pesaran, 2000).

The close link between decisions under uncertainty and forecasts suggests that there may be substantial benefits from specifying the constraints and preferences of the policy maker explicitly. For example, Elliott and Lieli (2013) construct a utility-based forecast for binary outcomes and show that it leads to large gains over other existing methods. However, the main difficulty of such an approach is that it is more information intensive, and may require knowledge about preferences that are not observable directly.

In this section, we discuss the problem of evaluating EWIs of banking crises – i.e., forecasts of the likelihood that a banking crisis will occur, given a set of covariates – from the perspective of macroprudential policy. We begin by discussing the difficulties of assessing the costs and benefits of such policies. In light of these difficulties, we introduce an evaluation metric that is consistent with the underlying decision problem but is nevertheless robust