Working Paper Series

No 45 / June 2017

Use of unit root methods in early warning of financial crises

by
Timo Virtanen
Eero Tölö
Matti Virén
Katja Taipalus

ESRB
European Systemic Risk Board
European System of Financial Supervision

Abstract

In several recent studies unit root methods have been used in detection of financial bubbles in asset prices. The basic idea is that fundamental changes in the autocorrelation structure of relevant time series imply the presence of a rational price bubble. We provide cross-country evidence for performance of unit-root-based early warning systems in ex-ante prediction of financial crises in 15 EU countries over the past three decades. We find especially high performance for time series that are explicitly related to debt, which issue signals a few years in advance of a crisis. Combining signals from multiple time series further improves the predictions. Our results suggest an early warning tool based on unit root methods provides a valuable accessory in financial stability supervision.

Key words: Financial crises; unit root; combination of forecasts

JEL codes: G01, G14, G21

1. Introduction

Given the costs[1] sustained by many countries in the aftermath of the 2007–2009 Great Recession, the current interest in developing early warning indicators for financial crises is hardly surprising. A good starting point for that is identifying the elements of a prototype crisis. A recent literature survey by Kauko (2014), drawing on several influential articles, concludes that typical banking crisis emerges after hefty increases in asset prices and indebtedness. In retrospect, respective time series look like a bubble. Thus, a basic issue for policymakers becomes one of predicting asset price bubbles.

Predicting an asset price bubble and the ensuing financial crisis is in many respects different from other prediction or forecasting exercises, especially from the point of view of public authorities. There is no generally accepted definition of financial crisis and the crisis may take many forms. Thus, prediction must rely on different indicators and different models. From the practical point of view of monitoring, priority has to be given to analytical frameworks that are easy to interpret and use information with minimum time lags. This does not mean that we should adopt a completely atheoretical approach, quite the opposite. In this paper we propose early warning indicators building on the main findings of asset-bubble theory elaborated by Campbell, Lo and McKinlay (1997), Campbell and Shiller (1988), Craine (1993), and Koustas and Serletis (2005).

The Campbell and Shiller (1988) paper notes that the behavior of dividend-price ratio together with dividend growth can reveal an existing rational asset bubble. Although the early literature deals with stocks, it is relevant for other assets as well. For example, in examining house prices, we can substitute rents for dividends. This approach can even be extended to the debt-to-GDP ratio, where income growth in a macro-setting plays a similar role to aggregate dividend growth. This is obviously true in a world where the functional distribution of income is constant. Moreover, we know from conventional government debt accounting models (see e.g. Wilcox, 1989) that stationary growth rate of taxable income is incompatible with a continuously increasing (nonstationary) debt-to-GDP ratio.

Empirical tests that deal with the presumed bubble properties of financial time series include Elliot (1996) and Elliot *et al.* (1999), which deal with the power of unit root tests (specifically) with different initial observations; Kim *et al.* (2002), Busetti and Taylor (2004),

---

[1] Lawrence Ball (2014), for example, estimates that the cost to 23 OECD economies in the aftermath since 2008 has averaged about 8.4% of GDP.

Leybourne (1995), and Leybourne *et al.* (2004), which deal with testing changes in the persistence of time series; and Homm and Breitung (2012), whose method considers stock market applications of unit root tests. Phillips *et al.* (2011 and 2013) use a "right-tailed" unit root test for detecting bubble-type behavior in time series, as well as develop a sup ADF (SADF) test statistic and derive its (limiting) distribution. The authors apply their testing procedure to several financial times series, and demonstrate reasonably good *ex-post* prediction performance. A similar approach relying on standard ADF tests with a rolling window (henceforth denoted RADF test) is found in Taipalus (2006a) and Taipalus and Virtanen (2016).

This study investigates the *ex-ante* predictive performance of the SADF test and RADF test in signaling the risk of financial crisis. Our testing sample consists of 15 EU countries over the period 1980–2014. The data include a broad set of financial and macroeconomic variables (see the next section for details) and a dataset of financial crises. The performance evaluation is based on the standard measures of the early warning literature that rely on the number of type I and type II errors.

The method would work perfectly if bubbles and their collapse were linked one-to-one with financial crises and if we could detect all such bubbles with our method. In reality this link is indicative at best since neither bubbles nor financial crises are binary occurrences. Also, we may not detect all bubbles and we may detect bubbles that do not lead to financial crises. The financial crisis prediction could likely be improved by further studying the size of the presumed bubble and resiliency of the financial system.

The results in this paper show that the unit root based methods are successful in predicting financial crises both in full sample and out-of-sample evaluation. This suggests strong links between financial crises and bubbles. The two methods yield quite similar results, on average beating benchmark conventional signaling method. The credit-to-GDP ratios and debt-servicing costs are the best-performing indicators in our "usefulness" metric, but indicators based on house prices and stock prices are also found to be useful. Furthermore, policymakers benefit from combining single-value indicators into composite indices. Using two variables, the obvious choices are the credit-to-GDP ratio and debt servicing costs. As the number of components is increased, including predictors such as house and stock prices adds to the usefulness of a composite index.

The findings are consistent with Jordá et al. (2015) who find that find that credit fueled asset price bubbles are more dangerous than others. Our results suggests that only slightly more than half of the crises in our data were preceded by a clear bubble in the residential real estate market, but for over two-thirds of the crises there was a long period of explosive growth of the total credit-to-GDP ratio ahead of the crisis. Moreover, results show that credit and house price based indicators typically signal crises many years in advance.

In summary, it is worthwhile to include unit root methods as a tool in identifying bubbles and the emerging risk of a financial crisis.

The rest of this paper is organized as follows. The data are elaborated in section 2.A. Sections 2.B and 2.C review the unit-root-based early warning methods. Section 2.D introduces the performance evaluation setup. Sections 3.A and 3.B present the empirical results for single variables and multiple variables, respectively. Section 3.C presents results on signal timing and section 3.D provides robustness checks. Section 4 concludes.

2. Empirical analysis

A. Data

The empirical analysis makes use of data from the following EMU countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain. We also include Denmark, Great Britain, and Sweden in our sample. Available data extend well back into the 1970s, when the regulatory environment was quite different from today. Administrative credit rationing was still common in many countries and the functioning of financial markets was markedly different from the present system. To limit possible bias due to structural changes in banking, we start our sample from 1980. In addition, robustness checks are reported for a short-sample starting from 2003 to capture only the most recent financial crises.

The dataset is based on a quarterly data series compiled by the ECB and shared within the macro-prudential analysis group (MPAG). The variables include credit-to-GDP ratios, credit aggregates, debt-servicing costs, residential and commercial real estate prices, stock prices, and other macroeconomic variables, see Table 1 for details. Most of the time series are based on publicly available data from BIS, ECB, and OECD. The exceptions are the ECB's debt-servicing costs and commercial real estate price data, which are not publicly available. We have also amended the dataset with higher frequency monthly observations of stock market from Bloomberg and house prices from BIS.

Acknowledging that previous research (e.g. Peltonen *et al.*, 2014; Eidenberger *et al.*, 2014; Ferrari and Pirovano, 2015: and Schularick and Taylor, 2012) shows numerous variables related to labor markets, government finance, and output growth are poor predictors of crises, we nevertheless include the entire dataset into our first round of analysis for demonstration purposes. In the subsequent analysis we concentrate on the most relevant variables.

Evaluating the quality of the warning signals requires data on financial crises. Choosing the right period is critical in such analysis as wrong choices automatically invalidate the results.[2] We use the crisis classification scheme of the ECB and ESRB (Detken *et al.*, 2015), which is based extensively on expert opinion from individual central banks. As an alternative to the defined crisis period, we use loan losses of banks (in relation to total lending) to illustrate stressed periods. The loan loss data was collected by Jokivuolle *et al.* (2015).

B. Stationarity and rational bubbles

A constantly demising dividend yield may be a sign of worsening overpricing. Rising prices should at some point be realized as higher dividends. If they are not, the price rise is not based on fundamentals. This becomes apparent from the log dividend-price ratio model derived by Campbell and Shiller (1988):[3]

$$d_t - p_t = \mathbb{E}_t\left[\sum_{j=0}^{\infty} \rho^j \left(r_{t+j} - \Delta d_{t+j}\right)\right] - \frac{c-k}{1-\rho}, \tag{1}$$

where d is log dividend, p is log price, and r is the discount rate. Constants ρ, c and k are parameters from log-linear approximation. The stationarity of and r and Δd implies that the log dividend yield must also be stationary (for details, see Cochrane, 1992; and Craine, 1993).

Presence of a unit root in the log dividend yield means that agents or their expectations are not rational (assuming no other fundamental market failures). A possible interpretation is that there is a rational bubble. Indeed, this view has spawned several studies where the stationarity properties of stock prices are examined using unit root testing procedures. Although some studies (e.g. Corsi and Sornette, 2014) attempt a general model of bubbles, the empirical work usually relies on unit root testing.[4]

---

[2] Recent evidence is provided in Ristolainen (2017).

[3] The formula is a generalization of the Gordon (1962) growth model for the case where dividend growth and rates of return change over time.

[4] Banerjee *et al.* (2013) and Franses (2013) offer novel techniques in bubble detection. Banerjee *et al.* (2013) use a random coefficient autoregressive model, while Franses (2013) tests the feedback between first and second differences of time series, causing the time series to explode.

For our purposes, the asset pricing equation defines a "market fundamental" price for an asset, i.e. the price justified by future dividends and possible other returns. Any deviations from this fundamental price could possibly constitute an asset price bubble. Denoting this fundamental price by p_t^f, we can formulate the asset price at time t as

$$p_t = p_t^f + b_t \ , \tag{2}$$

where b_t is the price bubble component. Assuming all agents (buyers and sellers of the asset) are rational and that they have the same information about the fundamental price, the bubble component must either be zero or follow a submartingale process (Phillips *et al.*, 2013):

$$\mathbb{E}_t(b_{t+1}) = (1 + r_f)b_t \ , \tag{3}$$

where r_f is the risk-free interest rate used to discount future earnings. Thus, when a bubble is present, the asset price process changes from I(1) (or even I(0)) to an explosive process.[5] Methods used in this paper are designed to detect this change in the time series dynamics.

C. Unit root tests for rational bubbles

This section introduces the RADF test (Taipalus, 2006a and Taipalus and Virtanen, 2017) and the SADF test of Phillips *et al.* (2013). Both methods are based on the familiar ADF test equation

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \varepsilon_t \qquad . \tag{4}$$

The difference between the two methods is in the application of time windows in the testing procedure.

For the RADF test, equation (4) is estimated for all possible subsamples of length W. Starting at $t = W$, the first estimation window includes observations $[0, W]$. The estimation window is rolled forward one observation at a time so that at each point t, the window includes observations $[t - W, t]$. The estimated parameter γ is tested for H$_0$: mild unit root ($\gamma = 0$)

[5] The term "bubble" is rather vague characterization of historical phenomena that have preceded financial or economic crises. It is thus hardly surprising that there is no consensus on the mathematical representation of a bubble. For our purposes, Eq. (3) is probably a good approximation.

against the alternative hypothesis of explosive behavior ($\gamma > 0$). Because of this hypothesis specification, the right tail of the Dickey-Fuller distribution is used (i.e. null is rejected if the test statistic exceeds the 95% critical value). Note that this differs from the usual application of the ADF test, where H_0 is I(1) and the alternative hypothesis is I(0) (i.e. the left tail of the Dickey-Fuller distribution is used).

Phillips *et al.* (2013) use the same empirical regression model (2), but their test estimates the ADF statistic repeatedly on a backward expanding sample sequence. The critical value of test, which differs from the normal ADF critical value, is compared to the sup value of this sequence – hence the name backward sup ADF test.[6]

Is there theoretical grounds for favoring either windowing system? The procedure of Phillips *et al.* (2013) no doubt appeals conceptually,[7] but the fixed rolling window corresponds with the practical objective of simply spotting the fundamental change in the data-generating process of the underlying time series, i.e. the point at which the “normal” regime in the time series process shifts to a different regime. In this respect, both Monte Carlo evidence and tests with actual time series suggest that standard unit root tests perform *ex post* about as well as the PSY test (see Taipalus, 2006a, and Taipalus and Virtanen, 2017).

Both tests have three parameters that can be adjusted: the length of the window (W), [8] the AR-lag length of the ADF equation (p), and the significance level of the test (α). The window length of the test should be set considering the sampling frequency of the data. Based on earlier simulation studies, the window length should correspond to three to five years of data (see Taipalus and Virtanen 2017).[9] The AR-lag length can be fixed or determined based on information criteria. Significance level remains a free parameter that can be used to adjust the sensitivity of the test.

To make the tests comparable, we initially use the same fixed set of parameters throughout the data. Therefore, we report the results using fixed window length of 12, 24 and 36 observations, an AR-lag length of one, and α of 0.05. As a second step we optimize the parameters in-sample and observe the out-of-sample performance.

---

[6] See Phillips *et al.* (2013) for details on how critical values are calculated.

[7] Phillips *et al.* (2011), for example, argue that standard unit root tests are inappropriate tools for detecting bubble behavior, because they fail to distinguish effectively between a stationary process and a periodically collapsing bubble model. In their view, the latter look more like data generated from a unit root (or even a stationary autoregression process) than a potentially explosive process.

[8] The SADF method uses the respective window size as the minimum window size.

[9] With quarterly data, three years amounts to only twelve observations of data, a length generally considered too short for a normal (left-tailed) ADF test. Based on our experience with real data and Monte Carlo simulations, however, we note that short windows do produce reasonably good results with right-tailed hypothesis testing

Another important point concerning both tests is that short-run explosive contractions in time series ("negative bubbles") are identified as warning signals in the test statistics. In practice, this means that the methods issue warning signals also from e.g. stock market crashes or contractions in house prices. In the context of early warning methods, however, such signals are not considered relevant as they typically occur after the crisis has started. Thus, in our evaluation, we remove all warning signals that occur when the corresponding time series is decreasing in value.

D. Performance evaluation

A bubble warning signal is issued at the point t if the null hypothesis presented in the previous section, H_0: "mild unit root", is rejected.[10] To assess the predictive performance of the unit root methods in predicting financial crises, we measure how frequently the bubble signals correctly precede known crises in the data. To this end, we define a pre-crisis window starting three years and ending one year before each crisis. We then employ a "relative usefulness" measure to come up with a single number that we can use to rank the different variables. To account for publication lags, all quarterly data are lagged by one quarter.

The relative usefulness measure draws upon the policy loss functions of Demirgüc-Kunt and Detragiache (2000) and Bussière and Fratzscher (2008), and the usefulness measure proposed by Alessi and Detken (2011) and later supplemented by Sarlin (2013). The loss function of Alessi and Detken (2011) is defined as follows:

$$L(\theta) = \theta T_2 + (1-\theta)T_1 = \theta \frac{C}{A+C} + (1-\theta)\frac{B}{B+D} \ , \qquad (5)$$

where the right-hand side is a weighted average of the type I and type II error rates, T_1 and T_2, respectively.[11] The weights θ and (1- θ) in the loss function reflect the policy maker's presumed preferences for type I and type II errors. A parameter value θ higher than 0.5 means that

[10] A possible simplification here is to just compare the estimate of γ to a threshold value to extract the warning signals as in Taipalus and Virtanen (2017). Extensive Monte Carlo testing gives at least partial justification for procedure. We could even use a simpler AR(1) equation, where the AR parameter $\rho_1 \geq 1$. This is because the "t-statistic" is defined as the ratio of the estimated coefficient and its standard error. Moreover, the test statistic for zero corresponds to a p-value that is close to 0.958. Thus, when the estimated coefficient is zero or positive, the p-value will always be above 0.95. This partly explains why the right-tailed test works so well with small window sizes. Since we are looking for test statistic values close to zero, variance is not a big issue as the estimated coefficient will be close to zero.

[11] In the formula, the order of T_1 and T_2 differs a bit from the earlier literature. This is largely a matter of convention in forming the null hypothesis. Here, a type I error (false positive) is the incorrect rejection of a true null hypothesis H_0. We thus set H_0, i.e. "no crisis within the next 3 years" so that a false positive indicates a false

the policymaker is more averse to missing a signal of an upcoming crisis than to receiving a false alarm. For the most part, as commonly done in the literature, we set $\theta = 0.5$, but also try $\theta = 0.6$ as a robustness check. *A* is the number of periods in which an indicator provides a correct signal (crisis starts within 1 to 3 years of issuing the signal), and *B* the number of periods in which a wrong signal is issued. *C* is the number of periods in which a signal is not generated during a defined period from the onset of the crisis (1 to 3 years). Finally, *D* denotes the number of periods in which a signal is correctly not provided. In other words, *A*=*TP*, number of true positives; *B*=*FP*, number of false positives; *C*=*FN*, number of false negatives; and *D*=*TN*, number of true negatives.

The relative usefulness statistic is defined as

$$U_r = \frac{\min(\theta, 1-\theta) - L}{\min(\theta, 1-\theta)} \ . \tag{6}$$

A relative usefulness of 1 would mean that the indicator is able to perfectly forecast all the pre-crisis windows and produces no false positives. A relative usefulness of 0 or less means that the indicator is not useful.

Table 2 demonstrates the relative usefulness concept for the full set of variables using three different window lengths in the RADF test. The top performers with $U_r \gg 0$ are the various credit-to-GDP ratios, large credit aggregates, and debt-servicing ratios. As pointed out above, some variables were not expected to perform well as reflected by close to zero or negative U_r. This set of weak predictors includes the level of GDP, level of interest rates, the unemployment rate, household disposable income and is dropped from the subsequent analysis. Regarding the window length used in the ADF test, a three-year window length performs best in line with the simulation evidence (Taipalus and Virtanen, 2017). The results are robust to different values of the window length in the sense that same variables have the best predictive performance with all window lengths. As shorter window is also conceptually more appealing from the early-warning point of view, we adopt the 3 year window size for the rest of our study.

---

alarm. A type II error (false negative) incorrectly retains a false null hypothesis. Thus, our false negative here means failure to detect a crisis.

3. Empirical results

Section 3.A provides the basic result demonstrating the usefulness of unit root methods in signaling financial crises. In section 3.B, we show that combining signals from multiple indicators enhances our crisis predictions. Thereafter, in section 3.C we return to single indicators and the timing of signals. Specifically, we consider the average time it takes for crisis to start after a signal is issued and demonstrate the added benefit in crisis prediction of signals aggregated over time. Finally, section 3.D provides robustness checks of model parameters and crisis variable.

A. Signals from single indicators

To illustrate our approach, we offer a set of RADF signal graphs for three predictor variables in Figures 1–3 (credit-to-GDP ratios, debt service ratio, and real house prices). The graphs show the development of the underlying variable, the signals from the unit roots, and the pre-crisis/crisis periods. In a visual inspection, the indicators all perform quite well in terms of signaling alarms well in advance of crisis onset. For example, the signals from debt service ratio and credit-to-GDP gap (Figures 1–2) successfully signal the most recent financial crisis in Denmark, Spain, France, Greece, Ireland, Portugal, and Sweden. However, the indicator also alarms for Austria, Belgium, and Italy, none of which experienced systemic crises according to the crisis dataset.[12]

Table 3 reports the relative usefulness for the RADF and SADF tests using the fixed parameter values (W=3 years, p=1, α=0.05). With θ=0.5 (first two columns), the best-performing indicators are the various types of credit-to-GDP ratios, real credit stocks, and debt servicing ratios. Real-estate and stock price based indicators have positive usefulness but they generally rank lower than the debt based variables. We find that using monthly data does not improve the usefulness values significantly relative to corresponding quarterly data.[13] Overall, the usefulness values of different variables are similar with the two methods but the average is not quite as high for RADF as it is for SADF. With a relatively higher share of missed crises and lower share of false alarms (last four columns in Table 3), RADF appears to be too insensitive compared to SADF so the performance difference could be an artefact of the choice of

[12] The banking sectors in these countries suffered considerable losses during 2008–2012, though.
[13] Of course, having early access to the data is in itself an advantage for the policymaker.

the significance level parameter. The performance difference is aggravated when the policymaker is more averse to missing a crisis (θ=0.6 in columns 3–4).

The results in Table 3 are already effectively out-of-sample because none of the model parameters is based on the data. One might want to select the sensitivity parameter α in such a way that it conforms to the policymaker's preference. For instance, a policymaker who wishes to have relatively few false alarms would have relatively lower α. In Table 4, we consider an out-of-sample evaluation, where the policymaker decides on the parameter α based on 1980–1999 training data, and the performance is subsequently measured in the following period (2003 to 2014).[14] The results are also compared with a conventional signaling method.[15]

It turns out that optimizing α with a training period, does not have a consistent effect on the performance in the later evaluation. Comparing columns (1) and (2) with columns (3) and (4), respectively in Table 4, the usefulness values do not change on average whether ~~the~~ α is fixed to 0.05 or optimized. This means that the initial guess of α=0.05 was a good one, and there is not enough past data to improve on that. Compared with the conventional signaling method, the RADF and SADF method produce higher number of warning indicators with relatively high usefulness. This result holds whether α is optimized or not. Between the two unit root tests, a small performance advantage remains for the SADF even when α is optimized. This suggests that its windowing system benefits some of the indicators. Especially indicators based on real-estate prices or relative real-estate prices seem to be largest gainers.

B. Aggregating signals from multiple indicators

We can form signal composite by taking a weighted sum of signals from *N* indicators:

$$A_N = w_1 I_1 + w_2 I_2 + \cdots + w_N I_N \,, \tag{7}$$

where w_i are the weights and each I_i is either 0 (no alert) or 1 (alert). A warning signal is obtained when the signal composite is higher than some threshold. For simplicity, we only consider uniform weights.

---

[14] Years 2000–2002 are not used in the training, because in 2003 the policymaker would not know whether they should be classified as pre-crisis or tranquil periods.

[15] The signaling method issues a crisis signal when the indicator moves about a predetermined threshold value. Here we use trend gap transformation of the indicator, and the threshold value is optimized based on the training sample. The trend is calculated with one-sided Hodrick-Prescott filter with a smoothing parameter of 400,000. This value, originally proposed by Borio (2012), is today widely used in official contexts. See Gerdrup *et al.* (2013) and Repullo and Saurina (2011) for analysis and discussion.

$$A_N = I_1 + I_2 + \cdots + I_N\ , \tag{8}$$

Now a warning signal is obtained if at least k of the indicators alert. Below, we will consider cases N=2 and N=3. To further limit the number of possible choices, we set I_1 to be the signals from total credit-to-GDP ratio as it was among the best performing indicators in section 3.A. For our 17 variables, the number of combinations is then 120. Figure 4 depicts as an example the signal composite that we call "SUM 3.1". It is composed of total credit-to-GDP, NFC debt service ratio and house price-to-rent ratio. In what follows we do not for sake of brevity show results for all the 120 combinations but only those 11 combinations listed in Table 5a. The choice is motivated by what follows.

We proceed to evaluate the usefulness just as in section 3.A. Table 5b presents the full sample results for fixed parameter values (W=3 years, p=1, α=0.05). For N=2, we include four combinations SUM 2.1–2.4. The usefulness values are relatively higher in the case that signal from at least one indicator constitutes an alert. The corresponding highest usefulness for both RADF and SADF is obtained with combination of signals from credit-to-GDP ratio and household debt-servicing cost ratio (SUM 2.3 in Table 5b). The performance of this indicator is better performance than the credit-to-GDP indicator alone, which means that combining indicators is useful at least in sample.

For N=3 the usefulness improves further, but the optimal combination depends on whether we use RADF or SADF. For RADF it is SUM 3.5 in Table 5b with "at least one alerts" condition (i.e. credit-to-GDP ratio combined with both HH and NFC debt service ratio). For SADF, the best are SUM 3.6 and SUM 3.7 but with "at least two alerts" condition. Former combines credit-to-GDP ratio with total credit to households and the debt-servicing ratio. Latter combines the credit-to-GDP ratio with total credit to households and the real stock price index.

So far we have shown that combining signals can be useful in sample. Table 5c demonstrates that the performance may prevail out of sample as well**.** Composites that perform well during their 1980–1999 training, continue to perform relatively well in the 2003–2012 out-of-sample estimation. Due to random variation, the performance differences between the best-performing composites are small, though. Generally, there is no winning composite index that would have highest relative usefulness in all sub-samples.

In summary, the results suggest that even in the simplest form, it is useful to consider the warning signals from several indicators at the same time. If a policymaker considers only a few indicators, more weight should go to those indicators that have higher relative usefulness when evaluated alone (e.g. credit-to-GDP ratios and debt-servicing costs). However, when the number of considered indicators increases, the relatively weaker and noisy indicators such as the alerts derived from house price ratios and stock price developments add positively to the aggregate useful information.

C. Signal timing

The unit root indicators as such do not tell how far away in the future a crisis may be. At best it can be said that the probability of a crisis increases as the number of consecutive warning signals grows. We can study the typical alerting lead using standard OLS regression. To this end we define a dummy variable $D_{i,t}$ which flags the starting period of each financial crisis. We consider a set of models explaining the start of crisis with the lagged early warning signal parameterized by the lag L:

$$D_{i,t} = \alpha + \beta I_{i,t-L} \,, \tag{9}$$

where $I_{i,t}$ are the warning signals. Given that β is positive and statistically significant, the typical alerting lead is taken to be the lag length L of the model with highest log-likelihood. Larger alerting lead means a longer time on average between indicator alerts and the onset of the financial crisis.

Table 6 shows the alerting lead for RADF and SADF. Debt-servicing costs and equity prices alert relatively late. Earliest warning are obtained with the house price-based indicators with lag lengths up to four and five years. The credit-based indicators fall somewhere within a two- to three-year lag. Hence, it appears that the credit-based indicators may benefit having near optimal lag length when compared to the one- to three-year windows used in the relative usefulness measure. The relative usefulness of equity prices and debt servicing ratios would increase if the prediction horizons were shorter, while the house price-based indicators would benefit from a longer prediction horizon. E.g. Drehmann *et al.* (2011) addressed this issue by recommending flexibility in forecast horizons.

It may also be informative how consecutive signals behave. For instance, there could be two completely different patterns, i.e. warning signals arrive in a *steady* cumulative manner, or they are *sporadic*, with one warning signal rarely followed by another. However, relative usefulness is not amenable to measure such benefit. To allow greater flexibility in the timing and pattern of signals, we construct an additional measure: the *success rate* of each variable to predict a forthcoming crisis within a window of five years before the crisis starts.

Here, the criterion for "predicted crisis" is that a warning is signaled for at least six consecutive quarters within the five-year pre-crisis window, allowing a break of at most one quarter. As such "number of crises predicted" does not consider that an indicator that signals all the time would predict every crisis, we construct yet another measure – the number of crises falsely predicted. As above, a pattern of six consecutive alarms is treated as a false alert. If there is no data on the variable during the pre-crisis window, the crisis is omitted from the calculation.

We now turn to the *number* of crises correctly predicted and false alarms based on our metric that counts consecutive signals within five years of crisis onset (see Table 7). While this criterion is laxer than our usefulness measure, it still provides insight into the behavior of different variables. For example, only slightly more than half of the crises in our data were preceded by a clear bubble in the residential real estate market as evaluated from the price-to-rent variable. On the other hand, for over two-thirds of the crises in our data, there was a long period of explosive growth of the total credit-to-GDP ratio ahead of the crisis.

Recalling Paul Samuelson's famous observation that "the stock market has forecast nine of the last five recessions," we must concede the same problem arises in early warning models for banking crises. Here, the model performance is similar to earlier studies. The ratio of correct crisis prediction to false crisis prediction is at best 2.5 for the debt-to-disposable income ratio, i.e. our best indicator predicted seven of the past five crises. Similarly, the prediction-to-false-alarm ratios for debt-servicing ratios and real-estate price ratios rank relatively high on this metric, about 1–2 and close to 1, respectively. By this metric, the credit-to-GDP ratio, which had the highest relative usefulness, predicted about *eleven* of the past five crises. The bottom line is that one should be careful when ranking the indicators by relative usefulness (or any other metric) as the results can be quite sensitive.

D. Robustness checks

Contrary to most other prediction models, our unit root indicator does not crucially depend on the choice of the threshold values of the trend gaps, but rather the choice of lag lengths, window sizes, and significance levels. To investigate the sensitivity to different window lengths and the significance level parameter, Figure 5 presents a plot of the trade-off between shares of type I and type II errors for the crisis alerts computed with RADF test.[16] Window lengths vary from 12 to 48 observations and confidence levels vary from 0.8 to 0.99. The figure also shows the ROC curve for the conventional signaling method.[17] Here too credit and debt service ratios generally outperform other variables, but specific performance depends on the parameter choices for the model. Some parameter choices perform better than the choices used to obtain our main results. A notable example is the real estate price-to-rent ratio, where a longer window length produces a much better result.

Finally, a look at the results with banks' loan losses as the crisis indicator concur with our earlier finding on the ability to predict banking system distress and financial crises, see Figure 6. In all cases, the big peaks in the loans losses/lending ratios can be predicted with the early warning indicator (in Figure 6, the credit-to-GDP ratio and RADF test is used in computing the warning signals). Thus, the results do not seem to be overly sensitive to the choice of crisis definition. The definition of crisis periods is, of course, something that is important from the point of view of evaluation of ex post performance of different methods, but it is far more important to traditional prediction methods (logit models, neural networks) than to unit root indicators, which do not necessarily require historical data for estimates or training (for additional discussion, see Drehmann and Juselius, 2013).

4. Concluding remarks

This study found that an early warning indicator or set of indicators based on unit root testing can help in predicting financial crises – assuming, of course, that the indicators are based on relevant time-series information and are computed using an appropriate set of parameter values e.g. windows length and number of lags. Although the choice of these values is determinative and creates a certain amount of specification uncertainty, unit root indicator approaches have several unquestionable advantages compared e.g. with conventional probit/logit-model model based prediction systems. They are easy to compute and flexible, and they may be used with different time frequencies. There is no upper limit in principle for the size of the data in terms of the number of indicators. They also allow repeated tests and accumulation of information.

[16] We expect the results to be qualitatively similar for the SADF test.

[17] The receiver operating characteristic (ROC) curve is the graph connecting all possible pairs of type I and type II errors obtained by altering the signaling threshold.

For example, using weekly data, one can scrutinize the pattern of repeated warnings rather than treat the test procedure as one-shot experiment.

As for extensions, a compelling issue is how to combine the alerts from different variables into a risk measure. For instance, one could formulate an additive model, searching for the optimal combination of variables and optimize the weights with which each variable is added to the risk measure. The model might also consider the length of the alert, e.g. explosive growth of the credit-to-GDP ratio continuing over several years certainly predicts a higher risk of a crisis than a few isolated alerts.

Performing such optimization using historical data means one is betting on the probability that future crises will unfold in the same way as past crises, an assumption that may not hold. The underlying irony of Reinhart and Rogoff's "this time is different" policymaker excuse is that historical regularities seem to be strikingly persistent.

References

Alessi, L. and Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy*, 27(3): 520–533.

Ball, L. (2014). Long-term damage from the Great Recession in OECD countries. NBER Working Paper 20185.

Banerjee, A., Chevillon, G., and Kratz, M. (2013). Detecting and forecasting large deviations and bubbles in a near-explosive random coefficient model. ESSEC Business School, ESSEC Working Paper 1314.

Betz, F., Oprică, S., Peltonen, T., and Sarlin, P. (2013). Predicting distress in European banks. ECB Working Paper 1597.

Borio, C. (2012). The financial cycle and macroeconomics: What have we learnt? Working Paper 2012/395, Bank for International Settlements.

Busetti, F. and Taylor, A. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics*, 123: 33–66.

Bussière, M., and Fratzscher, M. (2008). Low probability, high impact: Policy making and extreme events. *Journal of Policy Modeling*, 30(1): 111–121.

Campbell, J. Y., Lo, A.W., and McKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, NJ.

Campbell, Y. and Shiller, R. (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3): 195–228.

Campbell, Y. and Shiller, R. (1988b). Stock prices, earnings and expected dividends. *Journal of Finance*, 43(3): 661–676.

Cochrane, J. (1992). Explaining the variance of price-dividend ratios. *Review of Financial Studies*, 5(2): 243–280.

Corsi, F. and Sornette, D. (2014). Follow the money: The monetary roots of bubbles and crashes. *International Review of Financial Analysis*, 32: 47–59.

Craine, R. (1993). Rational bubbles – A test. *Journal of Economic Dynamics and Control*, 17: 829–846.

Demirgüc-Kunt, A. and Detragiache, E. (2000). Monitoring banking sector fragility: A multivariate logit approach. *World Bank Economic Review*, 14: 287–307.

Drehmann, M., Borio, M., and Tsatsaronis, K. (2011). Anchoring countercyclical capital buffers: The role of credit aggregates. *International Journal of Central Banking*, 7: 189–240.

Drehmann, M. and Juselius, M. (2013). Evaluating early warning indicators of banking crises: Satisfying policy requirements. BIS Working Papers No. 421.

Eidenberger, J., Neudorfer, B., Sigmund, M., and Stein, I. (2014). What predicts financial (in)stability? A Bayesian approach. Discussion Paper Deutsche Bundesbank No. 36/2014.

Elliot, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review*, 40: 767–783.

Elliot. G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64: 813–836.

Ferrari, S. and Pirovano, M. (2015). Early warning indicators for banking crises: A conditional moments approach. MPRA Discussion Paper No. 62406. Available at https://mpra.ub.uni-muenchen.de/62406/.

Franses, P. (2013). Are we in a bubble? A simple time-series-based diagnostic. Erasmus School of Economics, Econometric Institute Report 2013-12.

Gerdrup, K., Bakke Kvinlog, A., and Schaanning, E. (2013). Key indicators for a countercyclical capital buffer in Norway – Trends and uncertainty. Staff memo. Norges Bank.

Gordon, H. (1962). *The Investment, Financing and Valuation of the Corporation*. Irwin, Homewood, IL.

Jokivuolle, E., Pesola, J., and Viren, M. (2015). Why is credit-to-GDP a good measure for setting countercyclical capital buffers? *Journal of Financial Stability*, 18: 118–126.

Jorda, O., Schularick, M., and Taylor, A. (2015). Leveraged bubbles. *Journal of Monetary Economics*, 76, Supplement, pp. S1-S20.

Homm, U. and Breitung, J. (2012). Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics*, 10(1): 198–231.

Kim, T-H., Leybourne, S., and Newbold, P. (2002). Unit root tests with a break in innovation variance. *Journal of Econometrics*, 109: 365–387.

Koustas, Z. and Serletis, A. (2005). Rational bubbles or persistent deviations from market fundamentals? *Journal of Banking and Finance*, 29: 2523–2539.

Leybourne, S. (1995). Testing for unit roots using forward and reverse Dickey-Fuller regression. *Oxford Bulletin of Economics and Statistics*, 57: 559–571.

Leybourne, S., Kim, T., and Taylor, A. (2006). Regression-based test for a change in persistence. *Oxford Bulletin of Economics and Statistics*, 68(5): 595–621.

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics*, 12: 167–176.

Peltonen, T., Piloiu, A., and Sarlin, P. (2014). Tail-dependence measures to predict bank distress. Unpublished mimeo.

Phillips, P., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52: 201–226.

Phillips, P., Shi, P., and Yu, J. (2013). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P500. Cowles Foundation Discussion Paper 1914.

Reinhart, C. and Rogoff, K. (2009). The aftermath of financial crises. *American Economic Review*, 99(2): 466–472.

Repullo, R. and Saurina, J. (2011). The countercyclical capital buffer of Basel III: A critical assessment. CEPR Discussion Paper No. 8304.

Ristolainen, K. (2017). *Essays on early warning indicators of banking crises*. University of Turku, Series E, No. 14. (Published doctoral dissertation).

Sarlin, P. (2013). On policymakers' loss functions and the evaluation of early warning systems. ECB Working Paper No. 1509.

Schularick, M. and Taylor, A. (2012). Credit booms gone bust: Monetary policy, leverage cycles and financial crises, 1870–2008. *American Economic Review*, 102(2): 1029–1061.

Taipalus, K. (2006a). Bubbles in the Finnish and US equities markets. Bank of Finland Studies, Series E, No. 3.

Taipalus, K. (2006b). A global house price bubble? Evaluation based on a new rent-price approach. Bank of Finland Research Discussion Papers 29/2006.

Taipalus, K. (2012). Detecting asset price bubbles with time-series methods. Bank of Finland Scientific Monographs E47: 2012.

Taipalus, K., and Virtanen, T. (2016). Predicting asset bubbles with unit root methods. Bank of Finland, unpublished mimeo.

Wilcox, D. (1989). The sustainability of government deficits: Implications for present value borrowing constraints. *Journal of Money, Credit and Banking*, 54: 1837–1847.

Table 1. Names and definition of all tested variables

Variable	F(*)	L(**)	Content
Total credit-to-GDP	Q	x	Ratio of (nominal) total credit to the private non-financial sector to (nominal) GDP
Household debt service ratio	Q	x	Debt service to income ratio, households
Debt service ratio	Q	x	Debt service to income ratio, households and non-financial corporations
Total credit	Q	x	Total credit to private non-financial sector, in billion local currency (real)
Total NFC credit (nominal)	Q	x	Total credit to non-financial corporations, in billion local currency (nominal)
Total NFC credit (real)	Q	x	Total credit to non-financial corporations, in billion local currency (real)
Bank credit-to-GDP	Q	x	Ratio of (nominal) bank credit to the private non-financial sector to (nominal) GDP
Bank credit	Q	x	Bank credit to private non-financial sector, in billion local currency (real)
Total credit (nominal)	Q	x	Total credit to private non-financial sector, in billion local currency (nominal)
Commercial RE price (real)	Q	x	Commercial property price index (real, source: ECB)
NFC credit-to-GDP	Q	x	Ratio of (nominal) total credit to non-financial corporations to (nominal) GDP
Bank credit (nominal)	Q	x	Bank credit to private non-financial sector, in billion local currency (nominal)
NFC debt service ratio	Q	x	Debt service to income ratio, non-financial corporations
Gross domestic product (real)	Q	x	Gross domestic product, quarterly levels, in million local currency (real)
Commercial RE price (nominal)	Q	x	Commercial property price index (nominal, source: ECB)
Commercial RE price (nominal)	Q	x	Commercial property price index (nominal, source: ECB)
M3 (real)	Q	x	Monetary aggregate M3 (real)
M3 (nominal)	Q	x	Monetary aggregate M3 (nominal)
Residential RE price (OECD, real)	Q	x	Residential property price index (real, source: OECD)
Residential RE price (ECB, nominal)	Q	x	Residential property price index (nominal, source: ECB)
Nominal interest rate	Q		Three-month money market interest rates (nominal)
Consumer price index	Q	x	Consumer price index
Stock price (nominal)	Q		Stock price index (nominal)
Residential RE price-to-rent	Q	x	Residential real estate price to rent index. (Source: OECD)
Gross domestic product (nominal)	Q	x	Gross domestic product, quarterly levels, in million local currency (nominal)
Government debt to GDP	Q	x	General government consolidated gross debt
Stock price	Q		Stock price index (real)
Residential RE price (OECD, nominal)	Q	x	Nominal real estate price index (OECD)
Gross disposable income	Q	x	Gross disposable income of households in million local currency
Loans to income	Q	x	Ratio of household loans to gross disposable income
Residential RE price-to-income	Q	x	Residential real estate price to income index. (Source: OECD)
Total household credit (nominal)	Q	x	Total credit to households, in billion local currency (nominal)
Household credit-to-GDP	Q	x	Ratio of (nominal) total credit to households to (nominal) GDP
Total household credit (real)	Q	x	Total credit to households, in billion local currency (real)
Housing loans (real)	Q	x	Loans for house purchases, in million Euros, outstanding amounts (real)
Housing loans (nominal)	Q	x	Loans for house purchases, in million Euros, outstanding amounts (nominal)
Unemployment rate	Q	x	Unemployment rate
Government bond yield	Q		Long-term government bond yield (nominal)
Rents	Q	x	Actual rentals paid by tenants including other actual rentals, index
Residential RE price (nominal)	M		Residential real estate nominal price index (various sources)
Residential RE price (real)	M		Residential real estate real price index (various sources, only certain countries)
Stock price (nominal)	M		Stock price index (nominal, source: Bloomberg)

Table 2. Usefulness values of RADF test for all variables and 3 window sizes.

	Window length=12			Window length=24			Window length=36		
Variable	Ur	FPR	FNR	Ur	FPR	FNR	Ur	FPR	FNR
Total credit-to-GDP	0.43	0.19	0.38	0.29	0.22	0.49	0.39	0.23	0.38
Household debt service ratio	0.39	0.17	0.44	0.34	0.13	0.52	0.28	0.14	0.57
Debt service ratio	0.37	0.10	0.52	0.26	0.08	0.65	0.26	0.05	0.69
Total credit	0.40	0.34	0.26	0.17	0.39	0.44	0.17	0.40	0.43
Total NFC credit (nominal)	0.35	0.42	0.23	0.07	0.40	0.52	-0.19	0.38	0.81
Total NFC credit (real)	0.34	0.29	0.37	0.10	0.33	0.57	0.03	0.37	0.60
Bank credit-to-GDP	0.34	0.20	0.46	0.39	0.22	0.38	0.45	0.24	0.31
Bank credit	0.34	0.33	0.33	0.30	0.37	0.33	0.30	0.40	0.31
Total credit (nominal)	0.33	0.43	0.24	0.13	0.44	0.43	0.06	0.41	0.53
Commercial RE price (real)	0.27	0.16	0.57	0.17	0.12	0.70	0.17	0.10	0.73
NFC credit-to-GDP	0.26	0.17	0.57	0.14	0.16	0.70	0.07	0.18	0.75
Bank credit (nominal)	0.27	0.46	0.27	0.21	0.45	0.34	0.20	0.43	0.37
NFC debt service ratio	0.25	0.13	0.62	0.17	0.09	0.74	0.13	0.10	0.78
Gross domestic product (real)	0.23	0.35	0.42	0.21	0.38	0.42	0.02	0.30	0.68
Commercial RE price (nominal)	0.23	0.26	0.51	0.20	0.20	0.59	0.18	0.20	0.63
Commercial RE price (nominal)	0.23	0.26	0.51	0.20	0.20	0.59	0.18	0.20	0.63
M3 (real)	0.19	0.29	0.52	0.23	0.32	0.44	0.23	0.32	0.45
M3 (nominal)	0.21	0.38	0.42	0.25	0.31	0.43	0.08	0.27	0.65
Residential RE price (OECD, real)	0.15	0.23	0.62	0.11	0.21	0.68	0.17	0.17	0.66
Residential RE price (ECB, nominal)	0.14	0.33	0.53	0.07	0.36	0.57	0.15	0.38	0.47
Nominal interest rate	0.14	0.03	0.82	0.00	0.01	0.99	0.00	0.01	0.99
Consumer price index	0.12	0.39	0.49	-0.50	0.35	0.70	-0.07	0.28	0.79
Stock price (nominal)	0.12	0.15	0.74	0.01	0.14	0.85	-0.02	0.16	0.87
Residential RE price-to-rent	0.13	0.19	0.68	0.10	0.17	0.73	0.13	0.14	0.72
Gross domestic product (nominal)	0.14	0.38	0.48	0.12	0.29	0.58	-0.07	0.25	0.81
Government debt-to-GDP	0.11	0.05	0.85	0.07	0.10	0.83	0.17	0.00	0.83
Stock price	0.10	0.11	0.79	-0.02	0.10	0.92	-0.01	0.08	0.93
Residential RE price (OECD, nominal)	0.10	0.34	0.56	0.05	0.35	0.60	0.19	0.31	0.49
Gross disposable income	0.07	0.25	0.68	0.11	0.32	0.57	-0.12	0.37	0.75
Loans to income	0.07	0.40	0.53	0.38	0.28	0.33			
Residential RE price-to-income	0.07	0.16	0.77	0.12	0.16	0.72	0.17	0.16	0.67
Total household credit (nominal)	0.05	0.48	0.47	0.06	0.48	0.47	0.11	0.44	0.46
Household credit-to-GDP	0.03	0.32	0.65	0.23	0.31	0.46	0.35	0.29	0.36
Total household credit (real)	0.01	0.42	0.56	0.18	0.45	0.37	0.27	0.41	0.31
Housing loans (real)	0.00	0.42	0.58	0.06	0.46	0.47	0.32	0.43	0.25
Housing loans (nominal)	0.00	0.43	0.58	0.14	0.44	0.42	0.18	0.44	0.38
Unemployment rate	-0.01	0.06	0.96	-0.01	0.04	0.96	-0.03	0.03	1.00
Government bond yield	-0.04	0.04	1.00	-0.01	0.01	1.00	0.00	0.00	1.00
Rents	-0.16	0.50	0.66	-0.06	0.45	0.61	0.26	0.20	0.54

Ur denotes the relative usefulness value, FNR (FPR) the proportion (%) of false negative (positive) predictions. The policymaker's preference parameter is θ=0.5. Data cover 1980 to 2012. A one-quarter publication lag is used for quarterly variables except for equity prices.

Table 3. Prediction probabilities and the usefulness values for fixed confidence level parameter.

Data covers 1980 to 2012 and the confidence level parameter α is fixed at 0.05.

		Usefulness				False ratios (%)			
		θ=0.5		θ=0.6		RADF		SADF	
Variable	F	RADF	SADF	RADF	SADF	FPR	FNR	FPR	FNR
Bank credit-to-GDP	Q	0.34	0.54	0.11	0.46	18.0	4.6	26.3	1.7
Bank credit	Q	0.34	0.39	0.17	0.34	29.8	3.4	46.8	0.9
Total credit-to-GDP	Q	0.44	0.53	0.25	0.45	16.3	3.9	28.5	1.5
Household credit-to-GDP	Q	0.04	0.34	-0.29	0.23	27.3	8.7	38.6	2.9
Total credit	Q	0.40	0.36	0.26	0.33	29.9	2.7	51.8	0.6
Total household credit (real)	Q	0.02	0.31	-0.26	0.28	35.6	7.6	54.4	0.8
Loans to income	Q	0.01	0.18	-0.27	0.03	32.9	13.7	38.2	7.6
Debt service ratio	Q	0.37	0.34	0.10	0.07	9.3	5.4	9.4	5.6
Household debt service ratio	Q	0.41	0.44	0.20	0.25	14.0	6.9	15.3	6.1
NFC debt service ratio	Q	0.25	0.16	-0.06	-0.21	10.2	9.2	9.3	10.8
Residential RE price-to-income	Q	0.07	0.20	-0.32	-0.11	13.4	9.1	17.8	7.0
Residential RE price-to-rent	Q	0.14	0.22	-0.20	-0.05	16.1	8.5	20.3	6.8
Residential RE price (OECD, real)	Q	0.15	0.30	-0.16	0.10	20.5	7.5	24.6	5.1
Stock price (nominal)	Q	0.11	0.23	-0.27	-0.05	12.7	7.9	19.8	5.8
Stock price	Q	0.10	0.19	-0.30	-0.15	9.5	8.4	11.2	7.2
Residential RE price (ECB, nominal)	Q	0.12	0.26	-0.15	0.15	29.4	6.5	45.2	2.7
Residential RE price (OECD, nominal)	Q	0.10	0.19	-0.18	0.06	29.3	6.8	48.3	3.1
Residential RE price (nominal)	M	0.22	0.14	0.15	-0.02	52.6	2.5	42.6	6.7
Residential RE price (real)	M	0.12	0.19	-0.07	-0.05	39.0	7.8	26.2	9.7
Stock price (nominal)	M	0.30	0.18	0.12	-0.17	29.6	4.3	12.4	8.3

F denotes frequency of the time series. Window length (RADF) and minimum window length (SADF) is 12 quarters (or 36 months). θ is the policymaker's preference parameter. FPR = False Positive Rate, FNR = False Negative Rate. A one-quarter publication lag is used for quarterly variables, except for equity prices.

Table 4. Out-of-sample performance comparison.

Threshold-optimized out-of-sample usefulness is based on evaluation period 2003–2012, where the confidence level parameter α is optimized based on training with 1980–1999 data. In case of fixed threshold, the confidence level parameter α is fixed at 0.05.

		Usefulness, opt threshold		Usefulness, fixed threshold		Optimal critical value		Usefulness, benchmark	Usefulness ratio	
Variable	F	RADF	SADF	RADF	SADF	RADF	SADF	signaling	RADF	SADF
Bank credit-to-GDP	Q	0.28	0.45	0.35	0.43	-0.98	1.24	0.25	1.12	1.79
Bank credit	Q	0.36	0.46	0.35	0.29	-0.31	3.23	0.38	0.94	1.21
Total credit-to-GDP	Q	0.45	0.31	0.51	0.38	-1.08	-0.12	0.27	1.68	1.15
Household credit-to-GDP	Q	0.10	0.24	-0.12	0.26	-1.68	0.36	0.28	0.36	0.86
Total credit	Q	0.57	0.43	0.57	0.30	-0.26	3	0.22	2.58	1.96
Total household credit (real)	Q	0.17	0.20	0.08	0.19	-2.94	1.07	0.27	0.62	0.74
Loans to income	Q	-0.02	0.15	0.01	0.18	0	0	-0.03	n/a	n/a
Debt service ratio	Q	0.50	0.46	0.52	0.47	-1.23	0.25	0.28	1.79	1.64
Household debt service ratio	Q	0.39	0.39	0.46	0.50	-1.55	-0.79	0.32	1.20	1.23
NFC debt service ratio	Q	0.35	0.38	0.29	0.18	-2.03	-1.32	0.09	3.85	4.23
Residential RE price-to-income	Q	-0.05	0.09	-0.07	0.06	0.14	0.05	0.18	n/a	0.49
Residential RE price-to-rent	Q	0.00	0.10	0.02	0.17	0.46	1.41	0.03	n/a	3.35
Residential RE price (OECD, real)	Q	0.02	0.29	0.07	0.33	0.6	1.38	0.07	0.23	4.16
Stock price (nominal)	Q	0.04	0.50	0.23	0.34	1.13	-0.12	0.37	0.10	1.36
Stock price	Q	0.03	0.20	0.17	0.20	0.94	0.93	0.29	0.12	0.69
Residential RE price (ECB, nominal)	Q	0.11	0.13	0.13	0.22	-1.22	2.82	0.11	1.04	1.18
Residential RE price (OECD, nominal)	Q	-0.01	0.03	0.15	0.28	1.77	3.32	0.43	n/a	0.07
Average		0.19	0.28	0.22	0.28	-0.48	0.98	0.22	1.20	1.63
Median		0.11	0.29	0.17	0.28	-0.31	0.93	0.27	1.04	1.22

F denotes frequency of the time series. Window length (RADF) and minimum window length (SADF) is 12 quarters. The usefulness ratio is the usefulness of the unit root method (RADF or SADF with optimized threshold) divided by the usefulness of the benchmark method. The benchmark signaling method is based on the one-sided HP-filtered trend gap (or relative trend gap for non-ratio variables) with smoothing parameter λ=400,000. The policymaker's preference parameter is θ=0.5. A one-quarter publication lag is used for quarterly variables, except for equity prices.

Table 5. Performance statistics for composite indicators.

Panel a) shows the variables included in each composite. Panel b) presents the evaluation results for the full-sample with fixed confidence level parameter α=0.05. Panel c) presents the 2003–2012 out-of-sample usefulness results both for confidence level parameter α that is optimized based on 1980–1999 data, and for the case that α is fixed to 0.05.

a) Variables included in each composite

Composite	Variable 1	Variable 2	Variable 3
SUM 2.1	Total credit-to-GDP	Debt service ratio	
SUM 2.2	Total credit-to-GDP	NFC debt service ratio	
SUM 2.3	Total credit-to-GDP	Household debt service ratio	
SUM 2.4	Total credit-to-GDP	Total household credit (real)	
SUM 3.1	Total credit-to-GDP	NFC debt service ratio	Residential RE price-to-rent
SUM 3.2	Total credit-to-GDP	Bank credit-to-GDP	NFC debt service ratio
SUM 3.3	Total credit-to-GDP	Total credit	Residential RE price-to-rent
SUM 3.4	Total credit-to-GDP	Total credit	Total household credit (real)
SUM 3.5	Total credit-to-GDP	NFC debt service ratio	Household debt service ratio
SUM 3.6	Total credit-to-GDP	Total household credit (real)	Debt service ratio
SUM 3.7	Total credit-to-GDP	Total household credit (real)	Stock price

The first number in the composite name denotes the number of variables in the composite. All composites are tested with all relevant alerting thresholds.

b) Full sample 1980-2012, fixed parameters

	At least one alerts						At least two alert						At least three alert					
	RADF			SADF			RADF			SADF			RADF			SADF		
Variable	Ur	FPR	FNR	Ur	FPR	FNR	Ur	FPR	FNR	Ur	FPR	FNR	Ur	FPR	FNR	Ur	FPR	FNR
Credit-to-GDP	0.44	16.3	3.9	0.53	28.5	1.5												
SUM 2.1	0.47	20.4	3.0	0.52	30.5	1.4	0.33	5.0	6.3	0.35	7.2	5.7						
SUM 2.2	0.46	18.2	3.5	0.53	28.9	1.5	0.24	6.8	10.0	0.16	8.4	10.9						
SUM 2.3	0.52	20.4	2.5	0.55	30.8	1.1	0.34	6.1	9.5	0.45	10.7	6.9						
SUM 2.4	0.39	34.4	2.3	0.40	48.4	0.6	0.15	10.2	9.8	0.54	26.3	2.1						
SUM 3.1	0.48	26.8	2.3	0.50	35.3	1.1	0.30	8.8	7.4	0.33	15.2	6.1	0.05	1.6	14.1	0.11	3.9	12.9
SUM 3.2	0.44	26.4	2.7	0.51	36.3	0.9	0.44	11.1	4.4	0.56	20.3	2.2	0.16	4.2	11.7	0.18	5.8	11.0
SUM 3.3	0.38	39.1	1.9	0.34	54.7	0.6	0.49	16.3	3.3	0.52	32.1	1.3	0.13	4.8	10.2	0.28	12.3	7.3
SUM 3.4	0.38	42.4	1.5	0.30	58.8	0.5	0.40	22.5	3.5	0.47	41.6	0.7	0.16	9.4	9.8	0.54	26.0	2.2
SUM 3.5	0.53	22.0	2.3	0.54	31.1	1.1	0.38	9.8	7.7	0.44	14.6	5.9	0.22	3.1	11.8	0.18	4.3	12.3
SUM 3.6	0.39	37.7	1.9	0.39	49.1	0.6	0.39	11.3	5.0	0.59	22.7	1.6	0.13	2.6	11.1	0.34	7.2	7.6
SUM 3.7	0.36	38.1	2.2	0.36	51.7	0.6	0.31	11.6	5.8	0.59	24.7	1.4	0.01	1.4	13.9	0.27	2.6	10.0

Window length (RADF) and minimum window length (SADF) is 12 quarters. Policymaker's preference parameter θ=0.5 is applied in the usefulness calculation. FP = False Positive, FN = False Negative. A one- quarter publication lag is used for quarterly data, except for stock market data.

c) Short sample 2003–2012, optimized parameters

	At least one alerts				At least two alert				At least three alert			
	α=0.05		α optimized		α=0.05		α optimized		α=0.05		α optimized	
Variable	RADF	SADF	RADF	SADF	RADF	SADF	RADF	SADF	RADF	SADF	RADF	SADF
Credit-to-GDP	0.51	0.38	0.45	0.31								
SUM 2.1	0.52	0.32	0.39	0.28	0.50	0.52	0.56	0.49				
SUM 2.2	0.49	0.36	0.41	0.25	0.31	0.20	0.41	0.45				
SUM 2.3	0.57	0.35	0.33	0.22	0.41	0.54	0.53	0.49				
SUM 2.4	0.41	0.21	0.18	0.18	0.20	0.38	0.48	0.35				
SUM 3.1	0.41	0.32	0.32	0.20	0.39	0.28	0.46	0.44	0.05	0.15	0.07	0.19
SUM 3.2	0.43	0.25	0.29	0.19	0.54	0.52	0.43	0.40	0.18	0.22	0.38	0.57
SUM 3.3	0.37	0.25	0.30	0.27	0.62	0.39	0.57	0.46	0.13	0.22	0.16	0.12
SUM 3.4	0.41	0.18	0.17	0.16	0.57	0.32	0.47	0.39	0.21	0.39	0.59	0.42
SUM 3.5	0.55	0.34	0.30	0.17	0.48	0.53	0.48	0.39	0.25	0.20	0.43	0.54
SUM 3.6	0.40	0.19	0.13	0.16	0.60	0.42	0.46	0.38	0.14	0.48	0.58	0.46
SUM 3.7	0.42	0.22	0.22	0.19	0.34	0.38	0.50	0.37	0.06	0.27	0.04	0.25

Window length (RADF) and minimum window length (SADF) are 12 are quarters. Policymaker's preference parameter θ=0.5 is applied in the usefulness calculation. A one-quarter publication lag is used for quarterly data, except for stock market data.

Table 6. Alerting leads of RADF and SADF test.

The alerting lead (i.e. how many quarters before a crisis we are most likely to get a signal from this variable) is calculated separately for the RADF and SADF tests using the methodology explained in Section 2.4.

	Alerting lead, quarters	
Variable	RADF	SADF
Bank credit-to-GDP	10	10
Bank credit	12	13
Total credit-to-GDP	8	5
Household credit-to-GDP	21	21
Total credit	8	5
Total household credit (real)	17	12
Loans to income	19	16
Debt service ratio	7	5
Household debt service ratio	6	9
NFC debt service ratio	5	4
Residential RE price-to-income	20	24
Residential RE price-to-rent	10	10
Residential RE price (OECD, real)	15	9
Stock price (nominal)	6	6
Stock price	6	10
Residential RE price (ECB, nominal)	16	16
Residential RE price (OECD, nominal)	10	24

Each data series has quarterly frequency. Data covers 1980 to 2012 and the confidence level parameter α is fixed at 0.05. Window length (RADF) and minimum window length (SADF) is 12 quarters (or 36 months). A one-quarter publication lag is used for quarterly variables, except for equity prices.

Table 7. Number of correctly predicted crises and false alarms for RADF and SADF test.

Data covers 1980 to 2012 and the confidence level parameter α is fixed to 0.05.

	Crises found		False alarms		
Variable	RADF	SADF	RADF	SADF	N of crises
Bank credit-to-GDP	13	17	20	22	17
Bank credit	17	17	31	31	17
Total credit-to-GDP	13	17	17	21	17
Household credit-to-GDP	10	14	16	16	16
Total credit	15	17	33	33	17
Total household credit (real)	13	16	25	25	16
Loans to income	5	5	3	2	6
Debt service ratio	10	9	7	9	16
Household debt service ratio	8	9	4	6	12
NFC debt service ratio	6	4	5	5	11
Residential RE price-to-income	7	10	6	10	15
Residential RE price-to-rent	10	13	13	15	16
Residential RE price (OECD, real)	13	13	16	16	17
Stock price (nominal)	5	8	14	25	17
Stock price	3	4	9	11	17
Residential RE price (ECB, nominal)	10	12	17	18	13
Residential RE price (OECD, nominal)	12	15	27	28	17
Residential RE price (nominal)	8	7	8	7	8
Residential RE price (real)	6	5	7	6	8
Stock price (nominal)	8	4	21	10	12

F denotes frequency of the time series. Criteria for predicted crisis = an alert of six or more consecutive quarters during a 5-year pre-crisis window. One break in alerts are allowed. Window length (RADF) and minimum window length (SADF) is 12 quarters (or 36 months). θ=0.5 is the policymaker's preference parameter for false positive rate vs false negative rate. A one-quarter publication lag is used for quarterly variables, except for equity prices.

Figure 1. RADF alerts for the debt service ratio variable

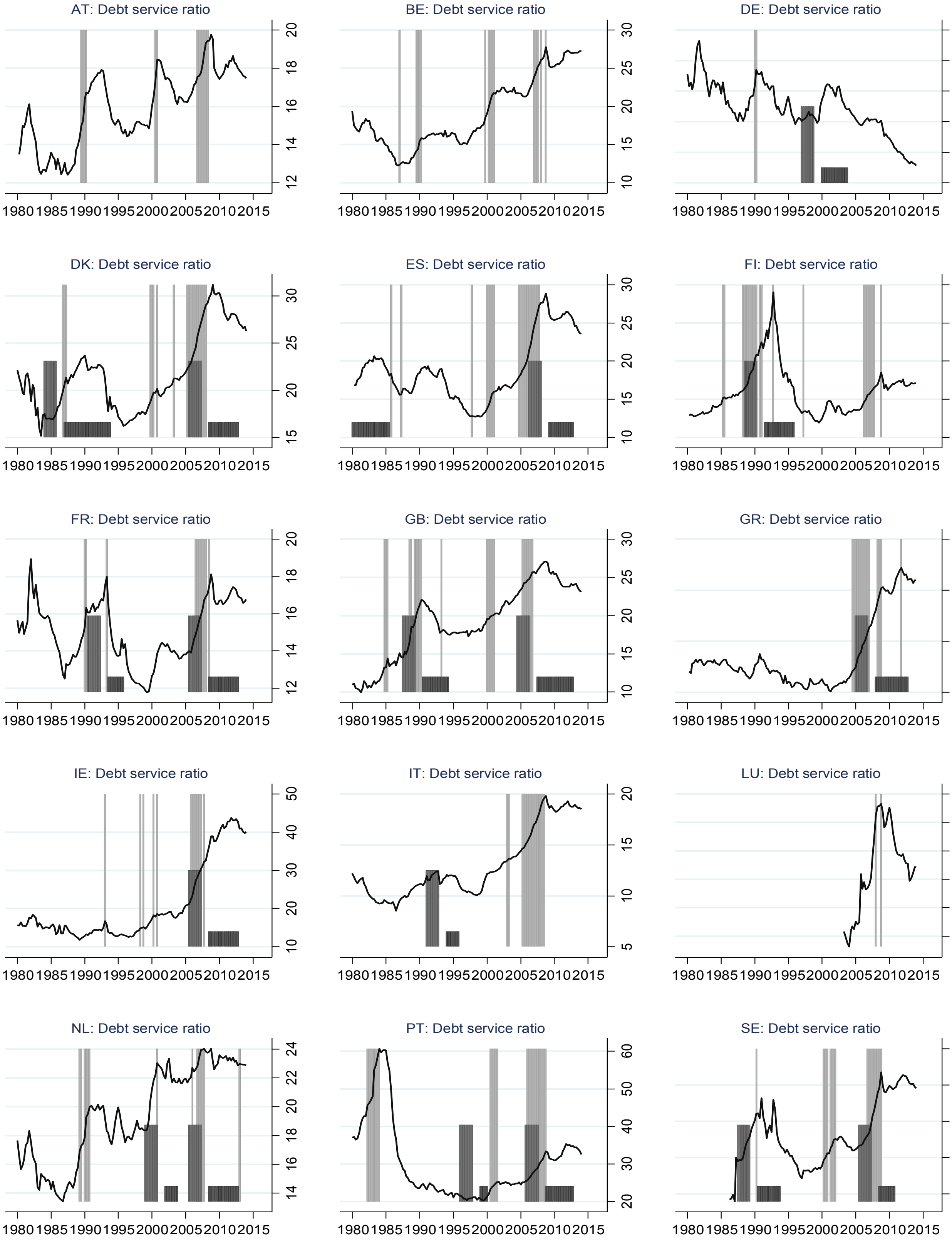


Figure 2. RADF alerts for the total credit-to-GDP ratio variable

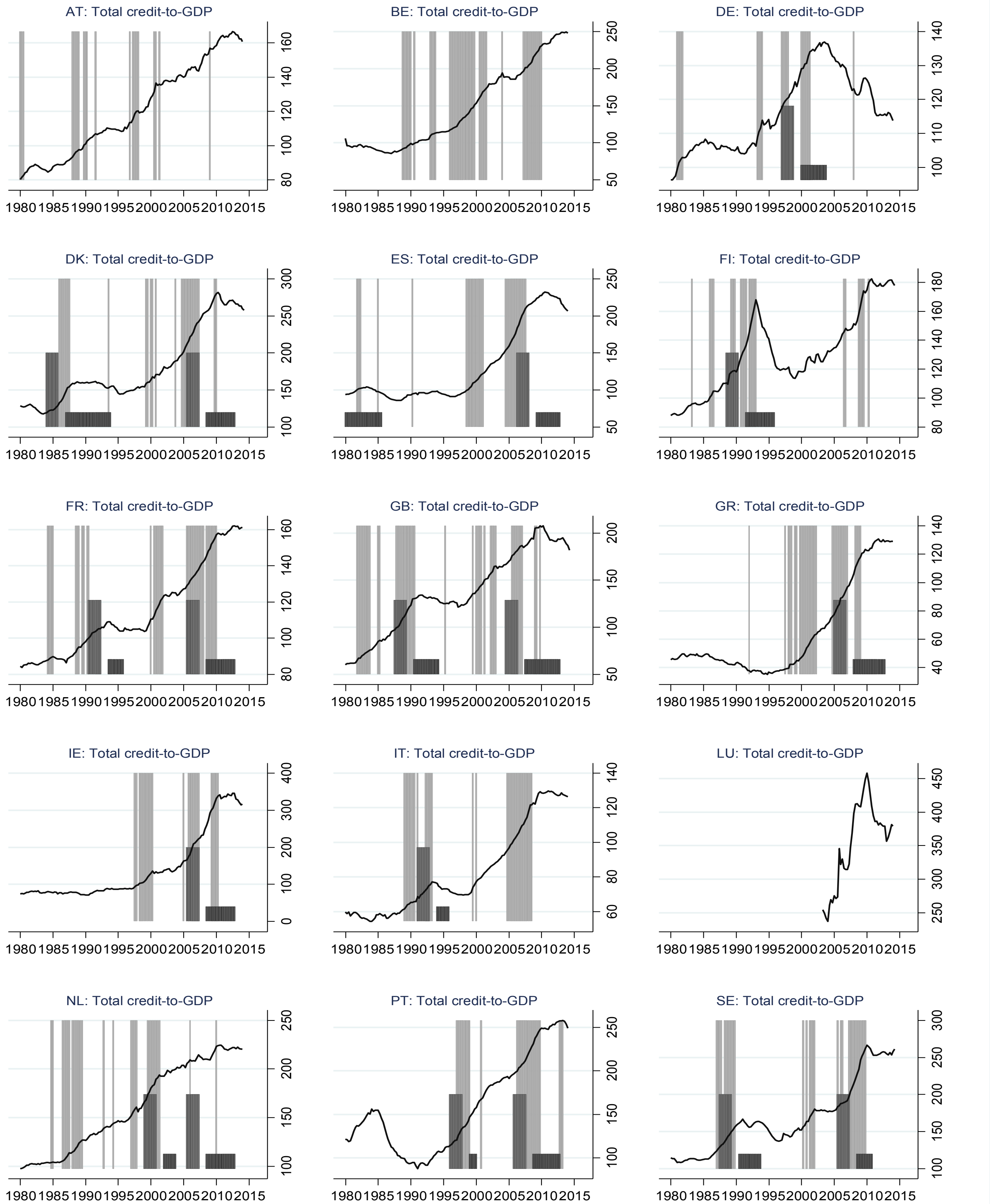


Figure 3. RADF alerts for the residential real estate price variable

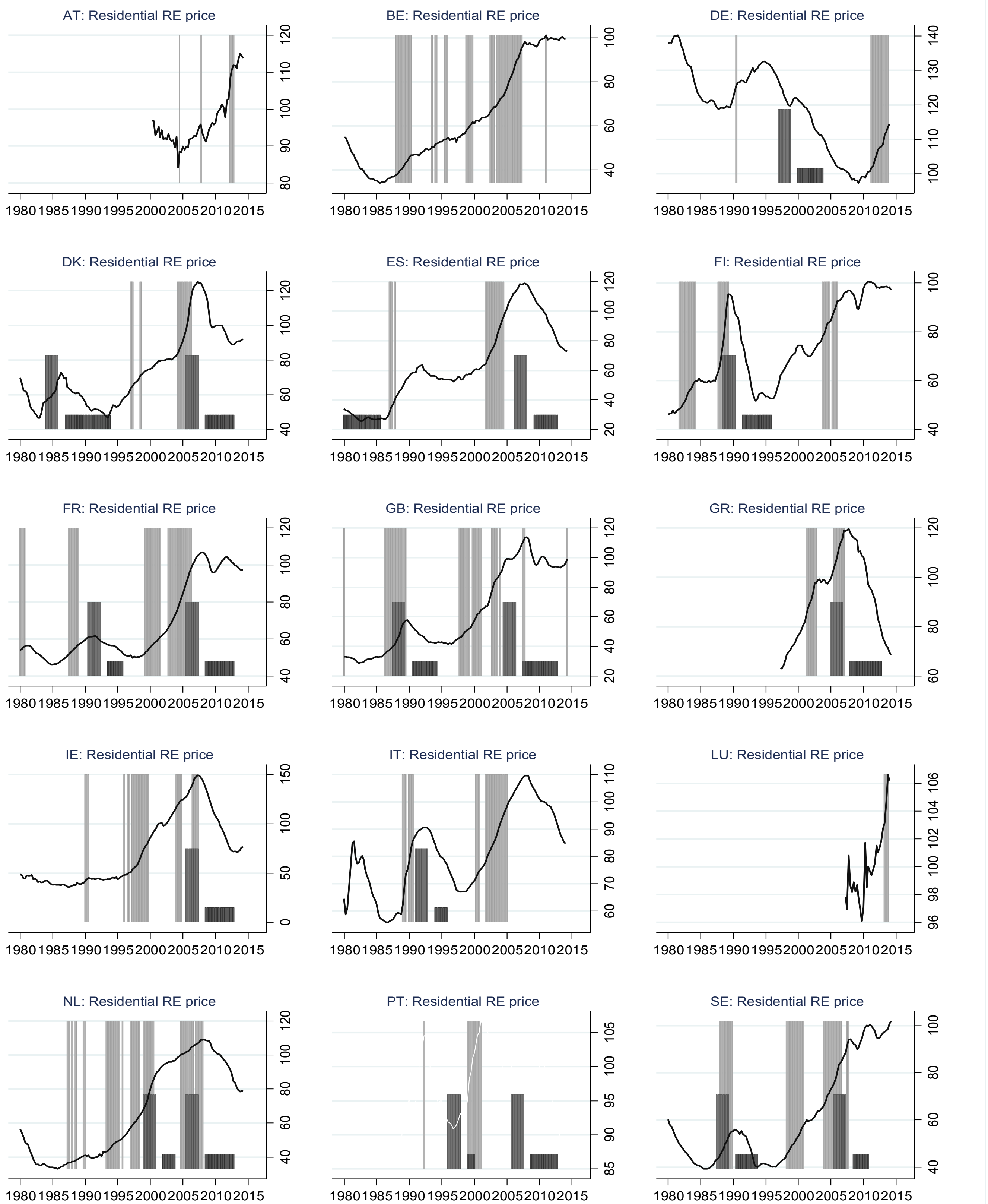


Figure 4. A sum of three RADF indicators (SUM 3.1)

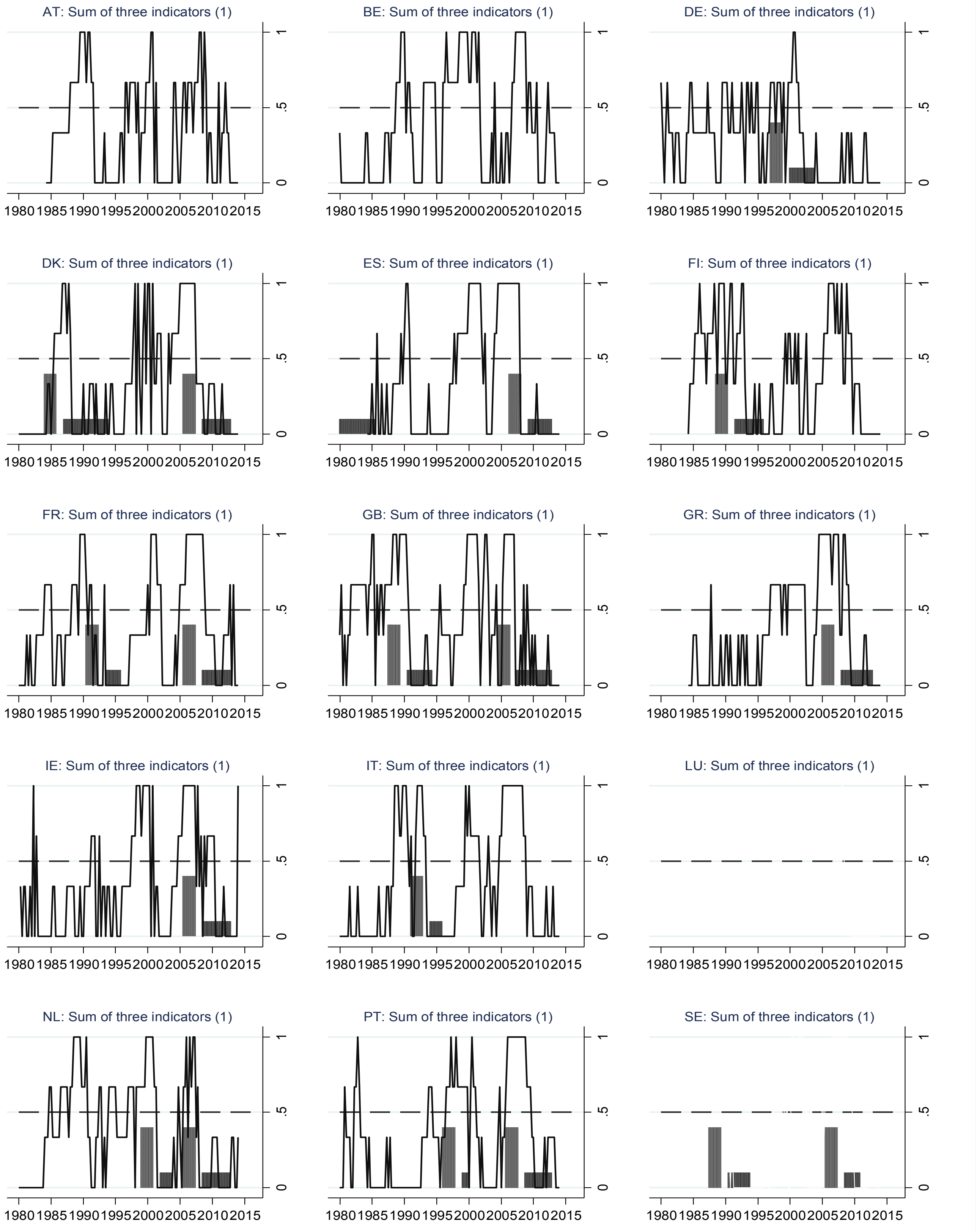
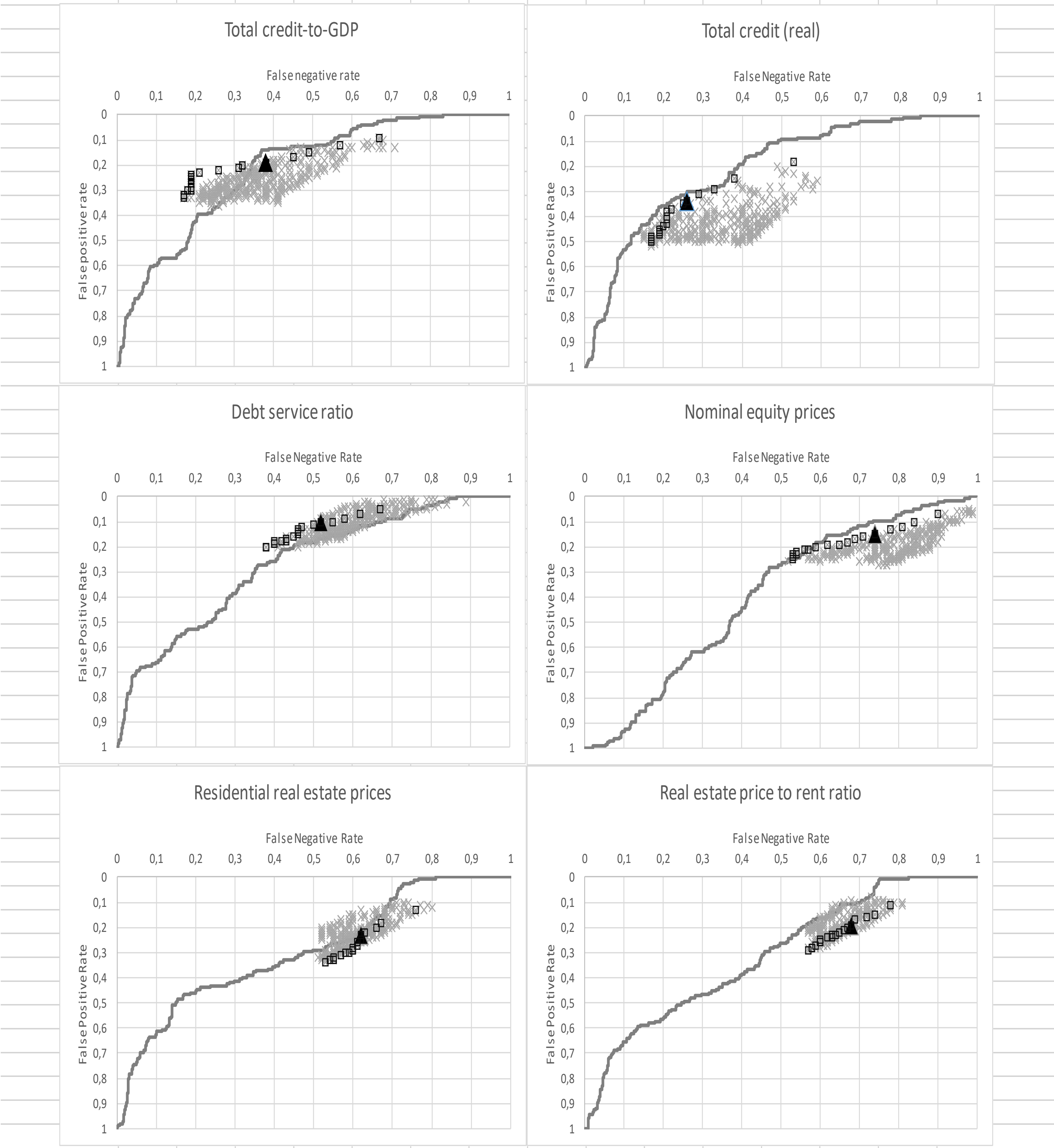
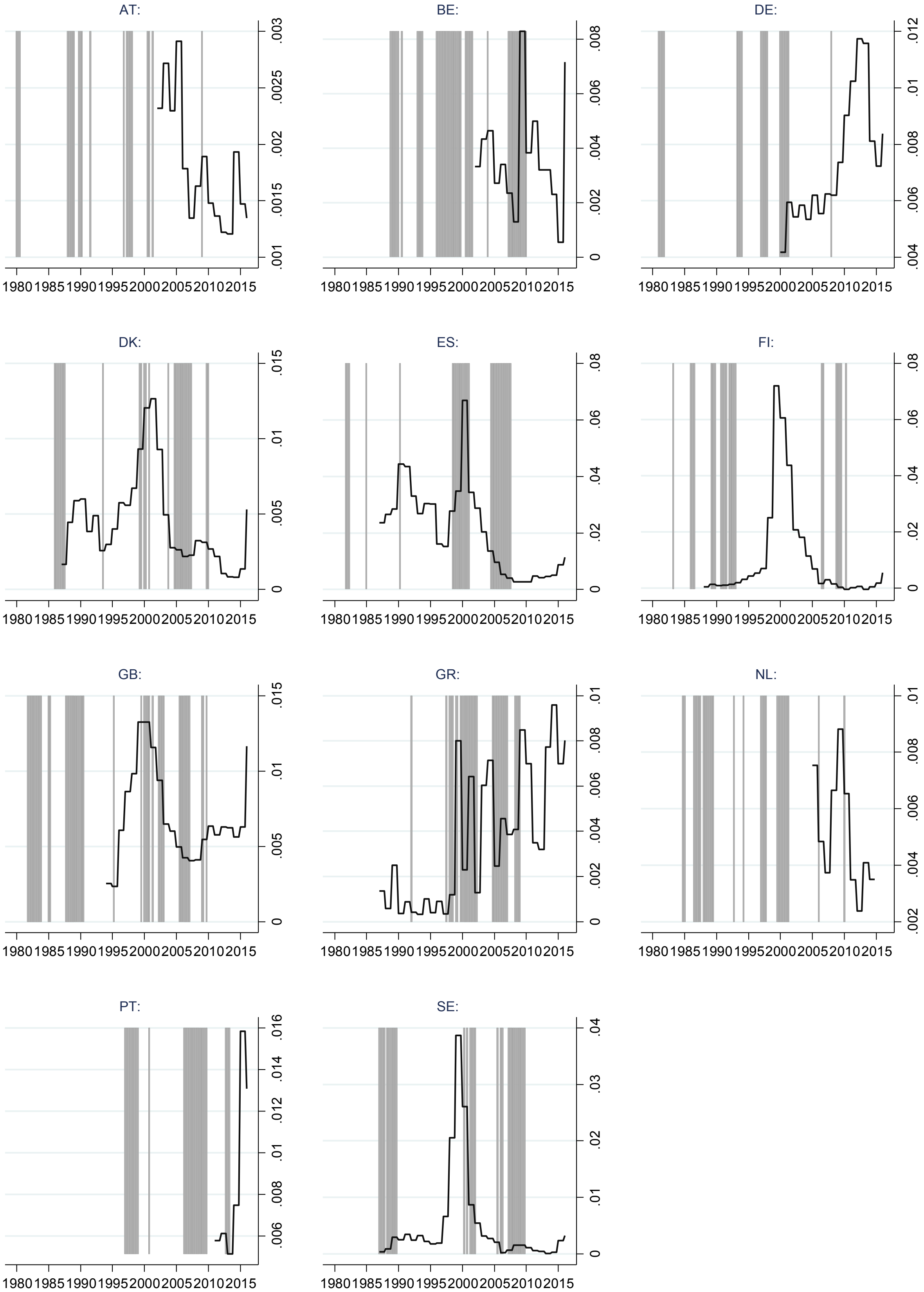


Figure 5. Parameter sensitivity for RADF test and ROC curve for conventional signaling method.



Solid line: The ROC curve of the relative trend gap. Gray boxes: FP and FN rates achieved with various window lengths and significance levels; Black boxes FP and FN rates achieved with a window length of 12 and various significance levels; black triangles: FP and FN rates achieved with a window length of 12 and the significance level of 0.95.

Figure 6. RADF alerts and banks' loan losses



The loan loss data was not available for France, Ireland, Italy and Luxembourg.

Imprint and acknowlegements

We gratefully acknowledge the comments of participants in the ECB's Macroprudential analysis group (MPAG) WS-B, the IIF's 2016 International Symposium on Forecasting (ISF) in Santander, and at the 2016 RiskLab/BOF/ESRB Conference on Systemic Risk Analytics in Helsinki.

Timo Virtanen
Turku School of Economics, University of Turku, Finland

Eero Tölö
Financial Stability and Statistics Department, Bank of Finland; e-mail: **eero.tolo@bof.fi**

Matti Virén
Turku School of Economics, University of Turku, Finland
Financial Stability and Statistics Department, Bank of Finland

Katja Taipalus
Financial Stability and Statistics Department, Bank of Finland

Postal address	60640 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Website	**www.esrb.europa.eu**

Note:
The views expressed in ESRB Working Papers are those of the authors and do not necessarily reflect the official stance of the ESRB, its member institutions, or the institutions to which the authors are affiliated.

ISSN	2467-0677 (online)
ISBN	978-92-95081-93-2 (online)
DOI	10.2849/654 (online)
EU catalogue No	DT-AD-17-012-EN-N (online)