The European Commission's Scoreboard of Macroeconomic Imbalances – The impact of preferences on an early warning system

Tobias Knedlik*

Abstract

The European Commission's Scoreboard of Macroeconomic Imbalances is a rare case of a publicly released early warning system. It allows the preferences of the politicians involved to be analysed with regard to the two potential errors of an early warning system – missing a crisis and issuing a false alarm. Such an analysis is done for the first time in this article for early warning systems in general by using a standard signals approach, including a preference-based optimisation approach, to set thresholds. It is shown that, in general, the thresholds of the Commission's Scoreboard are set low (resulting in more alarm signals), as compared to a neutral stand.

Keywords: Early warning system, Scoreboard, preferences, incentives, political economy

JEL: G01, F47, F53

Introduction

On 15 February 2012, the European Commission (EC) published its Scoreboard of Macroeconomic Imbalances (Scoreboard), which reports a set of macroeconomic indicators and provides thresholds that, if crossed, indicate imbalances (EC 2012a, 2012b). If imbalances are indicated, the EC undertakes an in-depth analysis to establish whether or not

* Halle Institute of Economic Research (IWH), Halle (Saale), Germany, tobias.knedlik@iwh-halle.de.

imbalances are present. If imbalances are indeed present, member states are asked to develop strategies to overcome these imbalances and may be fined if their strategies are found to be inappropriate or unsuccessful.[1] The exercise was undertaken in reaction to public debt crises in Europe and is included in the 'six-pack' of measures to reshape and tighten fiscal and macroeconomic supervision.[2]

The EC has published comments on why it has chosen particular thresholds for its indicators (EC 2012b), but it is unclear to what extent the choice of thresholds is driven by EC-specific preferences which do necessarily correspond to the preferences of the politicians in member states or other stakeholders. The EC's preferences in terms of the thresholds it has specified are explored in this article. The work is done for the first time for early warning systems in general. In addition, the forecasting results of the official thresholds of the Scoreboard are compared to thresholds derived from neutral preferences, and adjustments for thresholds are suggested. Although recent contributions suggest that, in general, the Scoreboard approach may be promising as an early warning system for public debt crises (Knedlik and von Schweinitz 2012), so far it has not been tested whether or not the Scoreboard in its current form is an appropriate early warning system. This task is undertaken in this article.

The article is structured as follows. In Section 1 the methodology is outlined. In Section 2 results regarding the EC's preferences are presented and the forecasting performance of the Scoreboard indicators with official thresholds is compared to a neutral stand. This is followed by a conclusion.

1. Method and data

Early warning systems feature prominently in both the academic literature and in practical policies to anticipate financial crises. However, recently, authors (re)discovered the need to

[1] Regulation (EU) No 1174/2011 of the European Parliament and of the Council of 16 November 2011.
[2] Regulations (EU) No 1173-1177/2011 and Council Directive 2011/85/EU.

model aspects of politicians' preferences into the construction of early warning systems.[3] They show that politicians' preferences regarding risk-averseness have a strong impact on the choice of crisis thresholds.

1.1 The signals approach and preferences

One of the simplest and most widely used methods to construct early warning systems is the signals approach developed by Kaminsky and Reinhart (1999)[4], which has been shown to produce statistically significant results even if it follows a non-parametric approach (El-Shagi et al. 2012). It assumes strong non-linearity in the relationship between indicator variables and financial crises. Indicator variables send a signal when a given threshold is crossed. The signal is interpreted as a sign of a looming crisis that can be expected to emerge within a predefined length of time. Thresholds are set to optimise the forecasting performance in a sample. In most prior studies, forecasting performance has been optimised by minimising a noise-to-signal ratio (e.g. Kaminsky and Reinhart, 1999). In Demirgüç-Kunt and Detragiache's (2000) study and more recent contributions,[5] thresholds are set to minimise the weighted sum of two potential forecasting errors. If thresholds are set too high, looming crises may be overlooked (Type I errors). Conversely, if thresholds are set too low, false alarms may be produced (Type II errors). Thus, an early warning system can have four potential results: first, a signal is issued and a crisis follows (State *A*); second, a signal is issued and no crisis follows (State *B*); third, no signal is issued and a crisis follows (State *C*); fourth, no signal is issued and no crisis follows (State *D*). States *A* and *D* are the desired results; State *C* constitutes Type I errors; State *B* constitutes Type II errors. Thus, $C/(A+C)$ is the share of

[3] The original contribution is Demirgüç-Kunt and Detragiache (2000).
[4] The signals approach has since been employed in several studies e.g. Alessi and Detken (2011), Edison (2003), Knedlik and von Schweinitz (2012).
[5] E.g. Alessi and Detken (2011), Bussière and Fratzer (2008), Knedlik and von Schweinitz (2012), Duca and Peltonen (2012).

Type I errors in pre-crisis periods, while $B/(B+D)$ is the share of Type II errors for tranquil periods.

All politicians may well be interested in accurate early warning systems, but in optimising an early warning system, there is a trade-off between the two potential types of error that could occur (e.g. Alessi and Detken, 2011). Minimisation of errors of one type leads to an increase in errors of the other type. Hence, the relative importance of both error types has to be defined. Early warning systems are instruments for decision-making (in other words, deciding whether or not pre-emptive action should be initiated), so the relative relevance of the two error types should be decided in line with politicians' preferences, which tend to depend on the costs of an error type to a politician. Type I errors imply costs because politicians could be blamed for not foreseeing and reacting to an emerging crisis with potentially high social costs. With Type II errors, politicians might be blamed for taking unnecessary and costly pre-emptive action.

The utility function for the politician is given by Alessi and Detken (2011):

$$U_I(\theta_I) = \min(\theta_I, 1-\theta_I) - \left(\theta_I \frac{C}{A+C} + (1-\theta_I)\frac{B}{B+D}\right) \quad (1)$$

The relative preference of a politician for avoiding the two types of error with regard to an indicator I is given by $\theta_I \in (0,1)$. Thus, $\theta_I = 1$ implies total ignorance of Type II errors, whilst $\theta_I = 0$ implies total ignorance of Type I errors. The weighted sum of both error types gives the loss to the policy-maker, which enters the utility function negatively. The first term of the equation (1) represents the secure loss to the policy-maker for two cases. In the case of $\theta_I > 0.5$, the policy-maker might set the threshold extremely low, resulting in no periods without signals (C=B=0) and a loss of $(1-\theta_I)$. In the case of $\theta_I < 0.5$, the policy-maker might set the threshold extremely high, so no signal is sent (A=D=0) and a loss of θ_I results. Thus, $\min(\theta, 1-\theta_I)$ can always be ensured, and employing an indicator only adds value if

the utility function is positive. Indicators with negative utility are best left out of consideration. For a given θ_I, one can calculate the utility for all potentially meaningful thresholds (which leads to different compositions with regard to the four states):

$$U_I(\theta_I) = \min(\theta_I, 1-\theta_I) - \left(\theta_I \frac{C}{A+C}[threshold_I] + (1-\theta_I)\frac{B}{B+D}[threshold_I]\right) \quad (2)$$

and would then choose the threshold that maximises utility as the optimal threshold ($threshold_I^*$):

$$U_I(\theta_I) \rightarrow \max: \frac{\delta U_I}{\delta threshold_I} = 0 \quad (3)$$

$$threshold_I^* = threshold_I^*(\theta_I) \quad (4)$$

Accordingly, the derived optimal thresholds can be used in early warning systems. The prior studies on early warning systems preferences go no further.

This study assumes that different incentive systems may lead to different preferences among politicians, as the costs associated with both error types to a politician vary according to institutional setting. The relevance of institutions in guiding political action is beyond doubt (e.g. Laffont, 2000), but their importance has not been stressed in the context of early warning systems that are employed by an international organisation and have a severe impact on policy in member countries. If, for example, a central bank is responsible both for anticipating a currency crisis and for taking pre-emptive action (e.g. increasing foreign exchange reserves, and tightening banking regulations), the preference for one or the other of the two error types may be more balanced than if there is an 'early warning committee' which might be held responsible for failing to forecast a crisis, but is not responsible for undertaking costly actions (and may therefore attach a lower relative weight to Type II errors). The EC largely plays the role of an early warning committee, because its responsibilities regarding taking costly precautionary action in case of a looming crisis are very limited (such action is basically left

to member states). Hence, it can be hypothesised that the EC would display a stronger preference for avoiding Type I errors than member state politicians would.

1.2 Exploring preferences in early warning systems

If the construction of an early warning system is known, and, in particular, if indicators and thresholds are known, it is possible to uncover politicians' preferences. Usually, early warning systems are kept secret, for good reason, including avoiding the risk of making self-fulfilling prophecies. The Scoreboard on macroeconomic imbalances of the EC is an exception to this tendency and allows decision-makers' preferences in an international organisation to be ascertained.

To unveil politicians' preferences we turn the modelling of preferences in an early warning system upside down. We run a standard approach for early warning systems, a signals approach, with the specific feature of a utility function to reflect relative error preferences (Kaminsky and Reinhart, 1999; Alessi and Detken, 2011). We then check at which specific relative preference the known threshold of the Scoreboard is also the optimal threshold, as derived from the model. We can then conclude that the respective relative preference is the relative preference of the politician(s) concerned. More formally, we build the inverse function of Equation (4):

$$\theta_I^* = threshold_I^*(\theta_I^*)^{-1} \quad (5)$$

Thus, in a case where the threshold is given, we can calculate the utilities for all θ_I – $threshold_I$ combinations and pick the θ_I for which the utility has a maximum at the given threshold. We call this θ_I the implicit θ_I^* or the implicit weight of Type I errors. Since there is an infinite number of potential θ_I^*, we restrict our calculations to two decimal places, in other words, 101 calculations per meaningful threshold of an indicator. If there is no hit (a given threshold is always either smaller or larger than the optimal threshold at any calculated θ_I),

we assume the implicit θ_I^* to be between the two θ_Is, where the optimal thresholds switch from above to below the given threshold. If the hit range comprises more than one θ_I (the given threshold is optimal for various θ_I), we assume that the implicit θ_I^* is the average θ_I of all θ_I within the hit range. With that information, we can assign one implicit θ_I^* to each given indicative threshold.

1.3 Assessing the usefulness of the Scoreboard for politicians

Once the preferences of the politicians involved in the construction of the Scoreboard are clear, we compare the official threshold with that derived (as described in Section 1.1) from a neutral stand of preferences, which we assume to be θ=0.5, i.e. equal weights for both error types.[6] This allows politicians with other preferences than the EC to judge the usefulness of the Scoreboard for their own purposes. It might also allow potential adjustments of the thresholds of the Scoreboard to gain wider acceptance among European Union members.

1.4 Data

We use the original EC data set with no transformations. The dataset covers 29 countries and 10 indicators[7] and is available annually. We employ the data from 1999 to 2010 for all indicators, except the house price panel, which only started in 2006. Crises the early warning system is meant to signal are defined as years in which a new IMF lending arrangement was instituted. We expect early warning signals in the year when the programme starts and in the two years preceding the lending arrangement. We do not consider the two years following a new arrangement. For countries where there is no current crisis, we also ignore the data for 2010, because we cannot know at the time of writing whether or not a crisis will follow. This results in crises windows as shown in Table 1.

[6] This assumption is in line with those of e.g. Alessi and Detken (2011), Knedlik and von Schweinitz (2012).
[7] The indicators are listed in Table 2 and described in Section 2.2.

Table 1: Definition of crisis windows

Crisis countries	Crisis windows
Bulgaria	2000-2004
Estonia	1999-2000
Greece	2008-2010
Hungary	2006-2008
Ireland	2008-2010
Latvia	1999-2001, 2006-2008
Lithuania	1999-2002
Portugal	2009-2010
Romania	1999-2004, 2007-2010

Source: Own definition based on IMF program data

We conduct our analysis for the whole set of countries, but also for two sub-groups: the euro area countries and the Central and Eastern European countries (CEECs). To avoid overlap of the groups, a CEEC that entered the euro area is counted as a euro area country from its date of entry. For the remaining group of non-CEECs, non-euro countries cannot be controlled for, since there was no incidence of a crisis in that sub-sample.

2. Results

2.1 General results concerning EC preferences

The results from the calculation of implicit preferences are shown in Table 2. The table shows that, on average, the weights on Type I and Type II errors are 0.56 and 0.44 respectively. Running a one-sided t-test shows that the weight on Type I errors is significantly larger (at a ten per cent level) than 0.5. Thus, the EC seems to put a greater weight on avoiding Type I errors than on Type II errors. Taking 0.5 as a reference weight, the EC chooses rather low thresholds, resulting in more crisis signals. The highest weights on Type I errors are found in the cases of export market share and the house price index, with θ_I^* of 0.79 and 0.82 respectively. There are, however, two exceptions from this general finding of relatively low thresholds. In the cases of the international investment position and the unemployment rate, the EC seems to take Type II errors more seriously, and chooses rather high thresholds. In almost all cases, there are differences regarding implicit preferences between the two sub-

samples. One may assume that the EC has specific country groups in mind in designing its thresholds. This is more evident when one looks at the different indicators.

Table 2: Implicit preferences for avoiding Type I errors (θ_I^*)

Indicators	**All countries**	**Euro-area**	**CEECs**
Three-year backward moving average of the **current account balance** in percentage of GDP	0.59	0.61	0.59
Net **international investment position** in percentage of GDP	0.35	1.00	0.45
Three-year percentage change of the **real effective exchange rate**	0.51	0.47	0.61
Five-year percentage change of **export market shares**	0.79	0.52	1.00
Three-year percentage change in nominal **unit labour costs**	0.53	0.47	0.62
Year-on-year changes in the **house price index** relative to a consumption deflator	0.82	0.76	1.00
Private sector credit flow in percentage of GDP	0.58	0.63	0.49
Private sector debt in percentage of GDP	0.53	0.69	0.35
Public sector debt in percentage of GDP	0.53	0.71	0.39
Three-year backward moving average of the **unemployment rate**	0.45	0.34	0.49
Average over all indicators	0.56	0.62	0.60

Source: Own calculations

2.2 Results for the different indicators

In Tables 3 to 5, we compare the thresholds set by the EC with the thresholds that would have to be chosen if θ was 0.5. Table 3 shows that, first, in half of the cases, the official thresholds lead to negative utility. This implies that for forecasting purposes, half of the set of indicators with those thresholds do not deliver useful insights. Second, utility is positive in all cases if equal weights of the error types are assumed. In only two cases, the unemployment rate and the current account balance, would the thresholds proposed by the EC have to be even lower. In the remaining eight cases, thresholds would have to be increased. Similar results are derived for the two sub-samples (see Tables 4 and 5). Looking at the different indicators in more detail reveals the following picture.

Table 3: Thresholds and utility for all countries

	Official threshold	**Utility at implicit θ_I^* of official threshold**	**Optimal threshold at θ=0.5**	**Utility at optimal threshold for θ=0.5**
Current account balance	Above +6% or below -4%	0.16	Above +6% or below -4%	0.25
International investment position	-35%	0.05	-20%	0.27
Real effective exchange rate	-/+5% for euro-area countries, -/+11% for non-euro-area countries	-0.01	-/+22% for euro-area countries, -/+28% for non-euro-area countries	0.00
Export market shares	-6%	-0.48	-19%	0.02
Unit labour costs	+9% for euro-area countries, +12% for non-euro-area countries	0.11	+18% for euro-area countries, +21% for non-euro-area countries	0.15
House price index	6%	-0.50	26%	0.08
Private sector credit	15%	-0.08	27%	0.05
Private sector debt	160%	0.00	240%	0.05
Public sector debt	60%	-0.01	110%	0.03
Unemployment rate	10%	0.06	6%	0.15

Source: EC, own calculations.

Table 4: Thresholds and utility for euro-area countries

	Official threshold for euro-area	**Utility at implicit θ_I^* of official threshold**	**Optimal threshold at θ=0.5**	**Utility at optimal threshold for θ=0.5**
Current account balance	+6% / -4%	0.16	+16.5 / -11	0.30
International investment position	-35%	0.00	-70%	0.45
Real effective exchange rate	-/+5%	0.00	-/+4%	0.03
Export market shares	-6%	0.19	-8%	0.21
Unit labour costs	+9%	0.13	+7%	0.17
House price index	6%	-0.53	19%	-0.00
Private sector credit	15%	-0.20	39%	0.05
Private sector debt	160%	0.03	240%	0.29
Public sector debt	60%	0.06	80%	0.27
Unemployment rate	10%	0.00	5%	0.19

Source: EC, own calculations.

Table 5: Thresholds and utility for CEECs

	Official threshold	**Utility at implicit θ_I^* of official threshold**	**Optimal threshold at θ=0.5**	**Utility at optimal threshold for θ=0.5**
Current account balance	+6% / -4%	0.10	+6% / -4%	0.15
International investment position	-35%	0.04	-20%	0.13
Real effective exchange rate	-/+11%	-0.20	-/+28%	0.00
Export market shares	-6%	-0.96	-12.5%	0.00
Unit labour costs	+12%	-0.06	+22%	0.09
House price index	6%	-0.50	26%	0.22
Private sector credit	15%	0.03	13%	0.04
Private sector debt	160%	0.00	100%	0.07
Public sector debt	60%	0.03	60%	0.07
Unemployment rate	10%	0.02	6%	0.04

Source: EC, own calculations.

2.2.1 Current account balance

The first Scoreboard indicator is the current account balance. The EC proposes the three-year backward moving average of the current account balance as a percentage of GDP, with a two-sided threshold of + six per cent and - four per cent. Given the data and the threshold, we can calculate the implicit θ_I^* = 0.59. Thus, the EC seems to worry more about Type I errors than about Type II errors. However, since the hit range of the given threshold with regard to implicit preferences is very broad (ranging from 0.38 to 0.79), other combinations of preferences (in this range) would lead to identical thresholds. The positive utility derived, if this indicator is employed, indicates its usefulness in an early warning system. The Scoreboard results allow three groups of countries to be identified. In the first group (Belgium, Denmark, France, Italy, Austria, and United Kingdom), no signal was issued at any time. The second group consists of countries with almost permanent signals (a maximum of two out of 12 periods with no crisis signal). This group consists of Bulgaria, Estonia, Greece, Latvia, Lithuania, Luxemburg, Hungary, Malta, Portugal, Romania, and Slovakia. The third group oscillated between signals and no signals (the Czech Republic, Germany, Ireland, Spain, Cyprus, the Netherlands, Poland, Slovenia, Finland, and Sweden). All the countries in

the ‘no signals’ group also belong to the group that never experienced a crisis in the sample period. Countries that did experience crises are almost all in the ‘permanent signals’group. The only exception is Ireland, where signals were only sent between 2007 and 2009, reflecting the Irish crisis window of 2008 to 2010 quite well.

Another case stands out in this group, namely Luxemburg, where signals were sent almost throughout the sample, but a crisis never occurred. Signals were sent in Luxemburg because the current account *surplus* exceeds the two-sided threshold, not the deficit. Similar examples can be found in the ‘oscillating’ group of countries. In Germany (2007-2009), the Netherlands (2005-2008), Finland (2000-2004), and Sweden (2004-2010) signals were sent for excessive surpluses. No crisis arose in any of these countries, which suggests that the current account balance indicator’s forecasting performance would increase considerably if it were used only as a one-sided indicator, focusing on deficits.

The recent signals in Spain and Cyprus suggest the risk of an upcoming debt crisis in these countries. Since we do not observe false alarms in CEECs due to surplus signals, the threshold of +six per cent/-four per cent seems to be the appropriate threshold for these countries. However, looking at the euro area countries, the threshold should be widened (if for any reason a two-sided threshold is preferred). A widening of the threshold to +16.5 per cent/-11 per cent would eliminate all wrong signals from current account surpluses, would simultaneously reduce the number of too early signals in the group of ‘permanent signals’ countries and would also increase the forecasting performance with regard to the current debt crises in the EMU. The caveat is that such a change would reduce the forecasting performance for the CEECs, because the earlier crises of these countries have already been signalled by lower current account deficit levels. For the current account balance indicator, three conclusions can be drawn. First, considering the current account balance as an early warning indicator for debt crises is very helpful. Second, instead of a two-sided threshold, a one-side

threshold for current account deficits should be used. Third, thresholds should be more specific with regard to the country groups.

2.2.2 International Investment Position

The second indicator is the net international investment position. It is expressed as percentage of GDP. The given one-sided threshold is -35 per cent. Based on the data, we can calculate the implicit preference for Type I errors: θ_I^* = 0.35. It seems that in the case of the net international investment position, the EC is more concerned about Type II errors and sets the threshold relatively high, compared to a reference situation with equal weights on both error types. Taking θ=0.5 as a reference, the optimal threshold would be -20 per cent. The higher threshold leads to a positive utility, making the net international investment position a valuable indicator. With the given threshold of -35 per cent, 11 out of 27 countries gave no signal during the period of observation. The group of 'permanent signals' countries comprises Estonia, Greece, Spain, Latvia, Hungary, and Portugal. The 'oscillating' group includes Bulgaria, the Czech Republic, Ireland, Cyprus, Lithuania, Poland, Romania, Slovenia, Slovakia, and Finland. Thus, the crisis countries are again mostly found in the group with almost permanent signals, except for Bulgaria, Ireland, Lithuania, and Romania. In Bulgaria and Lithuania, crisis signals were sent in all periods outside the crisis window and none within it. Thus, the international investment position is a perfectly wrong indicator for these countries. For Ireland, and the second crisis episode in Romania, the correct signals were sent. The only country in the 'permanent signals' group that did not experience a crisis (according to the definition used) was Spain. Among the 11 countries with no signal at all, no country experienced a crisis in the sample period.

Reducing the threshold to -20 per cent obviously leads to more signals (174 instead of 112). Better results are yielded in Romania, where the first crisis episode would have been forecast. In Lithuania and Bulgaria, crisis signals would also have been sent in the crisis window.

However, more false alarms would have been issued, for example in Austria and Sweden at the beginning of the sample period, and in Spain and Italy at the end of the sample period. The latter may again hint at unfolding crises in these countries. In sum, however, performance measures are again better for the threshold derived with an equal weight on both error types.

One interesting feature of the international investment position is that although the threshold should be reduced for all countries, that is not the case if the threshold is optimised for the euro area alone. As mentioned above, a reduction of the threshold would lead to better results in CEECs only, and would lead to more false alarms in euro area countries. Thus, for euro area countries the optimal threshold (assuming equal weights on both error types) would be -70 per cent – for the euro area, the official threshold would thus be too low. In sum, the international investment position is a good crisis indicator; however, compared to the neutral stand, the official EC threshold is set too high if all countries or CEECs are considered, but is set too low if only the euro area is considered.

2.2.3 Real effective exchange rates

The real effective exchange rate indicator is the percentage change over three years of the real (deflated by the consumer price index) effective (trade weighted) exchange rate. The EC proposes in the case of this indicator to use different two-sided thresholds for euro area countries (+/-five per cent) and non-euro area countries (+/-11 per cent). The use of these thresholds indicates that across all countries, the EC is somewhat less worried about Type II errors – hence, the threshold tends to be set relatively low, compared to a threshold with equal weights on both error types. Employing the indicator with the given thresholds leads to a utility below zero, meaning that the indicator is of little use in forecasting crises. The reason becomes clear in view of the results for different countries over time. We find that there are no countries in the group of countries with almost permanent signals. The group of countries with no signals at all contains only Denmark and Cyprus. All other countries belong to the

oscillating group and sometimes show some signals. This finding is surprising considering that only eight out of 27 countries experienced a crisis in the sample period.

Looking at the crisis countries, we find the following: in Bulgaria there is an almost equal number of correct signals and false alarms; in Ireland more false alarms than correct signals are sent; in Estonia, Greece, Hungary, Portugal only wrong signals are sent; in Lithuania more good signals than wrong signals are sent; and in Latvia and Romania only correct signals are sent.

Over all countries, signals since 2003 are only sent due to large real appreciations of the countries' exchange rates, except for three signals in Poland and the UK. Before 2003, in the euro area, signals are only sent because of the large real depreciation of exchange rates, while in the CEECs, signals are again only sent due to large real appreciations. Since the crises in euro area countries were recorded in the latter part of the sample, all signals due to real depreciations in the euro area are wrong. Regarding the CEECs, there is only one signal due to real depreciations (Poland in 2004), which is also wrong. Thus, the use of a two-sided indicator reduces the forecasting performance of the real effective exchange rate indicator dramatically. The Scoreboard should therefore use a one-sided indicator, focusing on real appreciations. However, even if a one-sided indicator is used, the real effective exchange rate may not be an appropriate indicator for the euro area, since for these countries the real effective exchange rate is, to a large extent, driven by developments of the nominal euro exchange rate, which limits the indicator's ability to differentiate between different euro area countries. If, for some reason, a two-sided indicator needs to be used, it should be used with much higher thresholds for CEECs (+/-28 per cent). For the euro area, the threshold should be reduced to +/-four per cent, which would have resulted in better forecasts for the crisis episodes in Ireland and Greece. That would at least increase the negative utility to around zero. To have a useful competitiveness indicator, one would need to consider a one-sided

indicator that takes intra-euro area imbalances into account more effectively. Alternatively, real effective exchange rates should not be used to judge the performance of countries with regard to crisis probabilities.

2.2.4 Export market shares

The export market share indicator is the five-year percentage change of the export market share. The EC threshold is set at -six per cent. At a first glance the indicator seems to be one of the indicators with the poorest performance on the Scoreboard, as its utility is very low. The θ_I^*=0.79 over all countries indicates that the threshold is set much lower than a neutral stand regarding the weights of the error types, whereas it would be set at -19 per cent. The picture looks very different if only euro area countries are considered – then, the indicator's forecasting performance is among the better ones, and the threshold derived based on equal weighting of the error types, -eight per cent, is also not far off the official EC threshold. It seems that the EC tailored this indicator specifically for the euro area countries, but still employs it for all countries. Regarding the euro area programme countries, we find some correct signals in all cases, while false alarms are limited. For the CEEC crisis countries, we find only one correct signal for an upcoming crisis, namely Bulgaria in 2000. The only countries with almost permanent (wrong) signals are France and Italy. No signals are sent over time in the Czech Republic, Estonia, Latvia, Lithuania, Luxemburg, Hungary, Poland, Romania, and Slovakia. The remaining 18 countries belong to the oscillating group, including the aforementioned euro area crisis countries (Greece, Ireland, Portugal) with rather correct signals. Assuming that the indicator works fairly well for the euro area, we find current (2010) alarm signals in all euro area countries, except Estonia, Malta, Slovenia, and Slovakia. Increasing the threshold to -12.5 per cent for the CEECs, as derived from a neutral stance regarding error types for these countries, would result in some correct crisis forecasts in Bulgaria and in limited false alarms, increasing utility to just zero. The indicator works so

poorly in the CEECs because these countries experienced a period of integration into the world markets where increases in the export market shares were reached in almost all countries in almost all periods, regardless of the incidence of crises.

2.2.5 Unit labour costs

The unit labour costs indicator is defined as the three-year percentage change of nominal unit labour costs. The one-sided thresholds are +nine per cent for euro area countries and +12 per cent for non-euro area countries. For unit labour costs, EC preferences seem to be similar to that for most of the indicators – it shows a higher preference for not overlooking a potential crisis than for avoiding false alarms. As in the case of real effective exchange rates, the findings vary over the sub-samples of the EU member countries. For euro area countries, the threshold of +nine per cent is rather high compared to a neutral stand, but the threshold for CEECs of +12 per cent is too low. As with export market shares, the phenomenon can be explained by the catching-up process of the CEECs. According to the Balassa-Samuelson effect, prices can be expected to grow faster in a catching-up process, thus labour costs may also increase at a higher rate without signalling risks for crises. For the EC, which only allows itself to differentiate between euro and non-euro area members, it is difficult to find a threshold for the latter group, which comprises CEECs as well as the more mature economies of Denmark, Sweden and the UK. A look at the signals reveals that only one country displays permanent signals, namely Romania. There are also only a few countries (Germany, France, and the UK, with the higher threshold of non-euro-area countries) that show no signal. The remaining countries show at least some signals. The crisis episodes in CEECs are signalled in about half of the periods (and all periods in Romania). In Ireland and Greece in two thirds of the pre-crisis periods signals are sent, whilst none are sent for Portugal. There are 85 false alarms in total, mainly in CEECs, but notably also in Spain and in non-crisis window periods in Ireland and Portugal. Decreasing the threshold to +seven per cent for the euro area (as

derived if equal weights are assumed for both error types) would lead to better forecasts in Greece and Portugal, with limited additional false alarms. Increasing the threshold for CEECs to +22 per cent (as derived from equal weights), many false alarms would be avoided, so the utility of the indicator could turn positive for this country group. In sum, unit labour costs is indeed one of the better indicators for crises. However, while the threshold values for euro area countries are almost adequate (albeit a little too high), they are far too low for CEECs at the current setting. This limits the overall performance of the indicator.

2.2.6 House prices

The house price indicator is defined as the annual growth rate of the real house prices index. The EC uses a one-sided threshold of +six per cent to indicate imbalances. The indicator performs exceptionally poorly in respect of forecasting crises in our sample. In particular, it performs very poorly in the euro area, even if adjusted thresholds are used. It seems to have some explanatory power for CEECs, if the threshold is increased considerably. The threshold used shows that for house prices, the EC seems almost to ignore the risk of false alarms in order to gain the advantage of missing as few crisis signals as possible, but this strategy is not very successful. All signals given by the indicator are false alarms, with the exception of Latvia in 2006 and 2007. There are two main problems relating to the house price indicator. First, the data sample starts only in 2006 and the panel is not balanced (e.g. there are no data for Hungary and Romania during their respective crisis windows). Thus, this very limited amount of data does not allow for a comprehensive judgement of the indicator. Second, it is very unclear why (except for the first problem) in the case of house prices the EC used the year-on-year percentage change instead of a change over a longer period, say of three years (as with the real effective exchange rate and the unit labour costs) or five years (as in the case of export market shares). That would be much more appropriate, as the case of the current crisis in Ireland exemplifies. The crisis that is indicated by the start of the IMF/EU

programme for Ireland emerged only a few years after the crash of the Irish property market. Thus, price increases of houses could only have been observed before that. Thus, while the bursting property price bubble in Ireland is probably one of the main causes of the crisis, an annual change of the house price is not a good indicator in the two-year early warning horizon. Because of both problems, the house price indicator should be dropped from the Scoreboard until appropriate data are available.

2.2.7 Private sector credit flow

The indicator measures private sector credit flows as a percentage of GDP. The threshold is set at 15 per cent. The analysis shows that in the case of private sector credit flows, the EC also aims to avoid Type I errors at the cost of Type II errors (θ_I^*=0.58 over all countries of the sample). Thus, the threshold is set relatively low, compared to a neutral stand. However, there are differences between country groups, as with other indicators. While a neutral threshold would be 27 per cent for all countries and even 39 per cent for the euro area, it would be just 13 per cent if the CEECs were considered in isolation. The overall poor performance of the indicator at the given threshold is expressed by negative values for the utility. That could, however, be turned into slightly positive values if the thresholds based on neutral stands were used – rendering the indicator a useful tool for the anticipation of risk, although it would still not be classified among the best performing indicators. Using a 15 per cent threshold, the indicator signals the more recent crisis episodes in Latvia, Hungary, and Romania quite well, but fails to signal the less recent crises in Latvia and Romania and also the crises episode in Estonia, Lithuania, and Portugal. Some indications are provided in the cases of Bulgaria, Ireland, and Greece. Almost permanently (wrong) signals are sent for the UK and Spain. However, there are also a few countries where (quite correctly) there are no signals at all: the Czech Republic, Germany, France, Italy, Poland, and Slovakia. Increasing the threshold considerably (to 27 per cent) would lead to much fewer false alarms, but would also reduce

correct signals (to a lower extent). As mentioned above, for CEECs the threshold should actually be lower. That would ensure even more correct crisis signals and a relatively low number of false alarms (these were mainly produced in euro area countries but also in the UK, Denmark and Sweden). In sum, the indicator would profit substantially from using regionally differentiated thresholds. However, its performance remains limited compared to that of other indicators on the Scoreboard.

2.2.8 Private sector debt

The eighth indicator of the Scoreboard is private sector debt. It is measured as the stock of private sector debt (the sum of non-consolidated loans and securities other than shares) as a percentage of GDP. The threshold is set at 160 per cent. The calculated θ_I^*=0.53 again indicates a preference for avoiding Type I errors. Although the calculated value is not far from equal weights, the use of θ=0.5 would result in a considerably increased threshold of 240 per cent over all countries, which would also be optimal if euro area countries were considered in isolation. For the CEEC subsample, however, the threshold would have to be decreased to 100 per cent. Accordingly, if the implicit weight of error type preferences is calculated for CEECs only, it seems the EC cares little about missing a crisis (θ_I^*=0.35 for CEECs only). Using a threshold of 160 per cent, the indicator would identify only the current crises in Portugal and Ireland. The indicator would produce false alarms almost constantly in Belgium, Denmark, Cyprus, Malta, the Netherlands, Sweden, and the UK. In nine countries, no signals are sent throughout the sample period, including crisis countries Greece and Romania. For CEECs, a reduction of the threshold to 100 per cent would mean that the recent crises in Hungary, Romania and Latvia would have been signalled with an acceptable level of additional false alarms, so that the negative utility of the 160 per cent threshold for CEECs would turn positive if 100 per cent were used as the threshold. However, if the 100 per cent threshold were to be employed for the euro area, permanent (wrong) signals would be

produced for almost all countries in almost all years. However, increasing the threshold to 240 per cent would still signal the crises of Ireland and Portugal (with indicator values as high as above 330 per cent in Ireland), while reducing the number of false alarms to zero. Thus, the level of private debt is a very good and accurate indicator for crises in some cases, but it does not play a central role in all crises. Again, the divergence of European countries needs to be considered in order to make the indicator useful. For the euro area, private sector debt would be the third best performing indicator (after the international investment position and the current account balance) – if an appropriate threshold is employed.

2.2.9 Public sector debt

The public sector debt indicator is defined as the general government debt as a GDP percentage. In line with the Maastricht criteria, the threshold is set at 60 per cent. The indicator threshold is again chosen by focusing more on Type I errors than on Type II errors (θ_I^*=0.53 over all countries). The 60 per cent threshold leads to almost permanent crisis signals in Austria, Belgium, Germany, Greece, and Italy, which have – except for Greece – not experienced a crisis. On the other hand, no signals are sent in eight countries, including the crisis episodes in Estonia and Romania. The only correctly signalled crises are those in Greece, Hungary and Portugal, and to some extent also those in Bulgaria and Ireland. Hence, the indicator (with the given threshold) performs poorly in forecasting crises and produces negative utility if the whole sample of countries is considered. That would change if adjusted thresholds were used. Over all countries, the neutral threshold would be 110 per cent. With this indicator threshold, only the crisis in Greece would have been predicted, but false alarms would be reduced to two. Thus, utility would turn positive, but the indicator would still be amongst the indicators that performed poorest. These results are counterintuitive, but can again be explained by the very different applicability of the threshold to the different country groups.

The 110 per cent threshold would be chosen to minimise false alarms also for the non-euro area non-CEECs. For euro area countries, the loss of correct crisis signals would not be outweighed by more correct predictions of non-crisis periods. Accordingly, the threshold should be set lower if only euro area countries are considered, to 80 per cent. That would restore crisis signals in the cases of Portugal, Greece and, to some extent, Ireland, still producing fewer false alarms than the 60 per cent threshold (e.g. the false alarms for Germany would disappear). Then the indicator would be the fourth best performing indicator for the euro area at the 80 per cent threshold. Regarding the CEECs, the 60 per cent threshold corresponds with a wide hit-range regarding error preferences. For preferences for Type I errors of between θ_I=0.23 and θ_I=0.54, a threshold of 60 per cent would be optimal. Thus, for the neutral stand of θ=0.5, the given threshold is optimal for these countries.

2.2.10 Unemployment rate

The final indicator on the Scoreboard is the unemployment rate, measured as the three-year backward-moving average of Eurostat's unemployment rate. The threshold is set at ten per cent. The unemployment rate is one of the few indicators with positive utility, even using the official threshold. It is one of two indicators (besides the international investment position) where the EC seems to prefer to avoid Type II errors (false alarms) – indicated by θ_I^*=0.45. Thus, as compared to a neutral stand, the threshold is set too high for the all-country case, and if the two subsamples of the euro area (θ_I^*=0.37) and the CEECs (θ_I^*=0.49) are considered separately. With a ten per cent threshold, only the earlier crises in Latvia and Lithuania would have been correctly forecast, with some indication for the looming crises in Estonia, Ireland, and Portugal. The crisis periods in Romania, Hungary, Greece and the later episode in Latvia would have been missed. With this high threshold, false alarms are limited, occurring mainly in Slovakia, Poland and Spain. By reducing the threshold to six per cent, which would be optimal based on a neutral stand for the samples of all countries and the CEEC subsample,

basically all crisis periods (with just two exceptions) would be correctly called. This leads to an increased number of false alarms, but overall, utility would be largely positive. Reducing the threshold even further, to five per cent, as would be optimal for euro area countries, would eliminate the remaining two periods with no signal before a crisis, by producing only a few additional false alarms in the euro area. Still the number of false alarms is high, including almost permanent (wrong) signals in Belgium, Germany, Spain, France, Italy, and Finland, if such a low threshold is used. The indicator is characterised by limited variations over time, but quite excessive variations over countries. However, if unemployment is rising, it is a good indicator that a crisis is looming.

2.3 Discussion of results

2.3.1 Discussion of the EC's preferences

Thus far, academia has had to speculate about politicians' preferences regarding Type I and Type II errors in the construction of early warning systems. With regard to costly asset price booms, Alessi and Detken (2011) state that they 'believe a θ smaller than 0.5 is a realistic description of central bankers' loss functions', while Bussière and Fratzscher (2002) assume that in the case of currency crises 'Type 2 errors may be less worrisome from a policy-maker's perspective'. With regard to the EC's Scoreboard, we can now state that, on average, Type II errors do indeed seem to be less worrisome to policy-makers. This provides evidence for the hypothesis that the EC, acting more as an early warning committee than as a full-fledged crisis preventer (that would also be responsible for undertaking costly action), is more concerned about missing a crisis than about raising a false alarm if there is indeed only limited evidence of an upcoming crisis. However, preferences also seem to vary with regard to the indicators employed. While in most cases, thresholds are set lower than if both error types were equally important, there are also cases in which the thresholds seem to be set rather high. In the case of the international investment position, the relatively high threshold

might be due to a compromise between the CEEC group and the euro area country group. Whilst in the CEECs the indicator works very well with a low threshold, in the euro areas a higher threshold would be optimal. In the case of the unemployment rate, it may be relevant that false alarms due to a lower threshold would mainly be found in the largest euro area economies (e.g. in Germany, France, Italy, and Spain). That may indicate that the EC uses some weighting scheme of countries when deriving thresholds. Additionally, there are also differences with regard to the different sub-samples.

2.3.2 Discussion of forecasting performance and threshold adjustment

The general picture that most indicators (using official thresholds) perform poorly could be changed in most cases if thresholds are adjusted. Already the use of an optimal threshold based on equal weighting of the error types would result in positive utility for all indicators, and would thus increase their relevance for early warning systems fundamentally. Moreover, the consideration of country group specifics would increase the forecasting performance of EC Scoreboard indicators further. While the differentiation between euro and non-euro countries, as used by the EC in some cases, can be considered a first step in that direction, it does not go far enough. In particular, the grouping does not allow differentiation between different levels of economic development and it ignores catching-up processes in CEECs. Also the characteristics of the non-euro non-CEECs are ignored. The characteristics of these countries are in many ways closer to the euro-area countries than to the CEECs (with which the EC groups them).

There is just one indicator, unemployment, for which the derived optimal thresholds are at least almost the same for the two subsamples of the analysis. A threshold of about six per cent would do a proper job for the whole region, although the indicator performs still relatively poorly in CEECs. With all other indicators, different thresholds for CEECs and the euro area would increase performance substantially. For the two indicators for which the EC proposed a

higher threshold for non-euro area countries, this proposal is supported by the results of the present study. However, the differentiation should be even stronger: Instead of -/+five per cent for the euro area and -/+11 per cent for non-euro-area countries for the real exchange rate indicator, -/+ four per cent should apply for the euro area and -/+28 per cent for CEECs. In the case of unit labour costs, the neutral stand would lead to an even lower threshold for the euro area and a much higher threshold for the CEECs. Moreover, with regard to export market shares and house prices, the CEECs need higher thresholds than the euro area. These indicators have in common the fact that they reflect price developments that might be more dynamic in emerging market economies without causing crisis risks (the exception is export shares, which is a poor indicator for CEECs in any case, as stated above). For all other indicators, the euro area should have higher thresholds than CEECs. This concerns the foreign trade indicators (current account balance and international investment position) and public debt, where larger imbalances can be sustained by the more mature economies, and private sector debt and credit. The latter cases might reflect the general development of financial markets, which results in both higher flows and stock of loans. In both cases, much higher figures could be allowed for euro area countries.

Conclusions

The current construction of the Scoreboard allows the error preferences of the policy-makers involved to be uncovered. In general, the EC shows a higher relative preference for avoiding Type I errors than for avoiding Type II errors. This can be explained by the EC's specific design as an early warning mechanism without a direct responsibility to undertake costly action if an emerging crisis is signalled. Assuming that politicians who also have to account for the costs of potentially unnecessary pre-emptive action may have different preferences and

would attribute a higher weight on avoiding Type II errors, we use equal weights for both error types as a reference system. Consequently, we find that the current Scoreboard is, in most cases, too alarmist, and that the threshold for two indicators (unit labour costs and the unemployment rate) are set rather high. Thus, preferences have an important influence on thresholds in early warning systems. To avoid setting a threshold in accordance with the preferences of only some of the stakeholders and thereby limiting its acceptance for others, it is recommended that thresholds be discussed openly.

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*, Vol. 27, No. 3, pp. 520-33.

Bussière, M. and Fratzscher, M. (2002). 'Towards a New Early Warning System of Financial Crises'. ECB Working Paper, 145.

Bussière, M. and Fratzscher, M. (2008) 'Low Probability, High Impact: Policy Making and Extreme Events'. *Journal of Policy Modeling*, Vol. 30, No. 1, pp. 111-21.

Demirgüç-Kunt, A. and Detragiache, E. (2000) 'Monitoring Banking Sector Fragility: A Multivariate Logit Approach'. *World Bank Economic Review*, Vol. 14, No. 1, pp. 287-307.

Duca, M.L. and Peltonen, T.A. (2012) 'Assessing Systemic Risks and Predicting Systemic Events. *Journal of Banking & Finance*, in print.

Edison, H.J. (2003) 'Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System'. *International Journal of Finance and Economics*, Vol. 8, No. 1, pp. 11-53.

El-Shagi, M., Knedlik, T. and von Schweinitz, G. (2012) 'The (statistical) significance of the signals approach'. IWH-Discussion Papers, 3/2012.

European Commission. (2012a) *Alert Mechanism Report*. Available at: <<http://ec.europa.eu/economy_finance/economic_governance/documents/alert_mechanism_report_2012_en.pdf [06.07.2012]>>.

European Commission. (2012b) 'Scoreboard for the Surveillance of Macroeconomic Imbalances'. Occasional Papers, No. 92, February 2012, (Brussels: European Commission).

Kaminsky, G.L. and Reinhart, C.M. (1999) 'The Twin Crises: The Causes of Banking and Balance-of-payments Problems'. *American Economic Review*, Vol. 89, No. 3, pp. 473-500.

Knedlik, T. and von Schweinitz, G. (2012) 'Macro-economic Imbalances as Indicators for Debt Crises in Europe'. *Journal of Common Market Studies*, Vol. 50, No. 5, pp. 726-45.

Laffont, J.-J. (2000) *Incentives and Political Economy*, (New York: Oxford University Press).