

## Assessing the risk of banking crises – revisited<sup>1</sup>

*Historically, unusually strong increases in credit and asset prices have tended to precede banking crises. Could the current crisis have been anticipated by exploiting this relationship? We explore this question by assessing the out-of-sample performance of leading indicators of banking system distress developed in previous work, also extended to incorporate explicitly property prices. We find that they are fairly successful in providing a signal for several banking systems currently in distress, including that of the United States. We also consider the complications that arise in calibrating the indicators as a result of cross-border exposures, so prominent in the current episode.*

*JEL classification: E37, E44, F34, G21.*

The current banking crisis is already widely regarded as among the most severe since the Great Depression. It has given renewed impetus to work on developing frameworks to address financial stability threats more effectively. Quantitative tools to inform assessments of the build-up of risk in the financial system are a natural element of any such framework. But the construction of reliable ones has proved elusive (eg Borio and Drehmann (2008)).

In previous work, Borio and Lowe (2002a,b) argue that focusing on the behaviour of credit and asset prices is a promising line of enquiry to develop simple and transparent leading indicators of banking system distress. Across a variety of policy regimes, these variables have tended to grow at unusually rapid rates for long periods prior to crises. However, a serious concern has been that, while performing fairly well with the benefit of hindsight, leading indicators based on those variables might not produce reliable signals for *future* crises. That is, they might work well in sample, but not out of sample.

In this special feature, we investigate this question by assessing the out-of-sample performance of those indicators over the period 2004 to 2008, in the light of the current financial crisis. We carry out two variants of out-of-sample exercises. In the first, we use the indicators as specified in the original studies, based exclusively on credit and equity prices. In the second, we also incorporate the information from property prices. While recognised as important

---

<sup>1</sup> We would like to thank Stephen Cecchetti, Bob McCauley, Pat McGuire, Frank Packer, Kostas Tsatsaronis and Karsten von Kleist for helpful comments and Marjorie Santos for excellent research assistance. The views expressed are the authors' and not necessarily those of the BIS.

in those studies, this was not possible there because of data limitations. We find that the indicator based exclusively on equity prices fails to issue warnings of the current financial strains, while the one that incorporates property prices does so for several countries, including the United States. At the same time, one significant limitation of the indicators is that they do not take into account cross-border exposures of banking systems. As a result, they fail to pick up crises associated with losses on foreign portfolios when the domestic economy does not show signs of credit and asset price booms. Drawing on the BIS international banking statistics, we show how these limitations can be addressed, although not fully resolved.

The first section recalls briefly the structure of the indicators, the basic philosophy underlying them and the findings of the previous studies. The second evaluates the in-sample and out-of-sample performance of the indicators. The third considers the limitations associated with the failure to incorporate explicitly cross-border exposures. The conclusions discuss possible future extensions.

## The indicators: structure, rationale and previous findings

The indicators are based on the view that banking crises often result from the growing fragility of private sector balance sheets during benign economic conditions – henceforth referred to as “financial imbalances”. These financial imbalances, associated with aggressive risk-taking, are driven by, but also feed, an unsustainable economic expansion. At some point, however, they unwind, potentially causing widespread financial strains. The precise timing of the unwinding is impossible to predict, but the longer the imbalance persists, the higher the likelihood of the reversal. This view is rooted in a long intellectual tradition that sees occasional financial crises as inherent in the dynamics of the economy and as the result of mutually reinforcing processes between the financial and real sides of the economy: the boom sows the seeds of the subsequent bust.<sup>2</sup>

The obvious difficulty, however, is how to identify in a reliable way the build-up of the imbalances as they develop, ie to distinguish what is sustainable from what is not in real time. After all, expansions of this kind are typically associated with developments supporting the belief that the trend growth of the economy has increased (eg structural reforms, real and financial innovations). Under these conditions, there is a very fine line between what is “far” and “too far”. And the relevance of historical relationships is unclear. Moreover, to be useful for policy, any indicator has to identify the risk of future financial strains with a lead sufficient to allow the authorities to take remedial action.

Banking crises are often preceded by a build-up of financial imbalances ...

---

<sup>2</sup> In the postwar period, prominent exponents of this view include Kindleberger (2000) and Minsky (1982). The full formalisation of such endogenous financial cycles has proved more elusive, but elements can be found in models that stress the interaction of credit and asset price “bubbles” (eg Allen and Gale (2000)). See Borio and Drehmann (2008) for a further discussion of these issues.

... that manifest themselves in unusually rapid credit expansion and buoyant asset prices

Despite these difficulties, previous work suggests that even some simple exercises can help us make progress towards an answer. Borio and Lowe (2002a,b) argue that it is possible to construct indicators that provide a fairly good sense of the build-up of imbalances as they develop (see the box for details).<sup>3</sup> The basic idea is that the imbalances manifest themselves in the *coexistence* of unusually rapid cumulative growth in private sector credit and asset prices. The indicators are intended to capture the coexistence of asset price misalignments with a limited capacity of the system to withstand the asset price reversal. Both of these are measured based on deviations of variables from their trends (“gaps”). The gaps are calculated so as to incorporate only information that is available at the time the assessments are made (ie are based on one-sided trends). Asset price misalignments are captured by asset price gaps, in inflation-adjusted terms, while the shock absorption capacity of the system is proxied by credit gaps, in terms of the ratio of private sector debt to GDP – a coarse measure of leverage for the economy as a whole. Signals of future crises are issued when these gaps exceed certain thresholds. As the precise timing of the unwinding of the financial imbalances is impossible to predict, the authors use a flexible horizon.

Simple leading indicators of banking crises can exploit this relationship

That body of work finds that, in sample, the performance of these indicators is quite good. They identify episodes of banking distress with a lead that, depending on the calibration, can vary between one and four years (Borio and Lowe (2004)). They also exhibit comparatively low noise-to-signal ratios<sup>4</sup> despite their parsimony, alleviating the false positives problem.<sup>5</sup>

One drawback stressed in those studies is that, owing to data limitations, the only asset price that could reliably be used in the construction of the indicator was equity prices (stock price indices). Property prices, which have played such a prominent role in banking crises, were not available for many emerging market countries. Moreover, for many industrial countries the length of the series was not regarded as sufficient to allow estimation of the initial trend values with an acceptable degree of confidence.<sup>6</sup> With the benefit of several more years of observations, in this exercise we also consider versions of the previous indicator that incorporate property prices.

---

<sup>3</sup> There is a small but growing literature on estimating early warning indicators for banking crises. For recent surveys, see Demirgüç-Kunt and Detragiache (2005) and Davis and Karim (2008a). Davis and Karim (2008b) examine whether different early warning indicators developed by them could have predicted the current crises but find them not to be successful. Alessi and Detken (2008) propose real-time indicators for costly asset price booms and find that some specifications would have issued persistent warning signals prior to the current crisis.

<sup>4</sup> The noise-to-signal ratio is the ratio of the fraction of type 2 errors (ie the number of (false) positive signals issued relative to non-crisis periods) over 1 minus the fraction of type 1 errors (ie the number of instances in which no signal was issued relative to the number of crises observed).

<sup>5</sup> In addition, Tarashev (2008) finds that these indicators improve the performance of widely used indicators of credit risk, such as KMV EDFs (probabilities of borrowers' default).

<sup>6</sup> The first decade of data is used simply to calculate the trend, before any forecast is actually made.

## Determining the optimal indicator

The indicators are based on a signal extraction method, which is one of the most common approaches to estimating early warning indicators (Kaminsky and Reinhart (1999)). For each period,  $t$ , a signal,  $S$ , is calculated. The signal takes the value of 1 (is “on”) if indicator variables ( $V_{1,2,3}$ ) exceed critical thresholds ( $\theta_{1,2,3}$ ); it is 0 (is “off”) otherwise. In the special feature, we analyse combinations of two- and three-indicator variables. For a signal to be issued, both critical thresholds have to be breached in the case of two-indicator variables. In the case of three-indicator variables, a signal is issued if the first indicator variable,  $V_1$ , exceeds its threshold,  $\theta_1$ , and at the same time at least one of the remaining two variables breaches its own (see panel below).  $V_1$  always refers to a credit variable and  $V_2$  and  $V_3$  to an asset price (see the main text for a definition of the series).

Two-indicator variables	Three-indicator variables
$S_t = \begin{cases} 1 & \text{if } (V_1^t > \theta_1 \ \& \ V_2^t > \theta_2) \\ 0 & \text{else} \end{cases}$	$S_t = \begin{cases} 1 & \text{if } (V_1^t > \theta_1 \ \& \ (V_2^t > \theta_2 \ \text{or} \ V_3^t > \theta_3)) \\ 0 & \text{else} \end{cases}$

The individual indicator series  $V_i$  are all measured as deviations from one-sided Hodrick-Prescott trends (“gaps”), calculated recursively up to time  $t$ , which is the point at which the signals are issued. The value of the smoothing parameter (lambda) for the estimation of the trend is quite high for the annual frequency of the data, 1600. The high degree of smoothing is intended to better capture the gradual and cumulative build-up of imbalances, which could be missed if the trend followed the actual data too closely.

We use multiple horizons to analyse the performance of the signals. A signal of 1 (0) is judged to be correct if a crisis (no crisis) occurs any time within the chosen horizon, ie any time within one, two or three years ahead, respectively.

Ideally, the vector of thresholds  $\theta$  would be chosen so that the indicator variables would always exceed the critical thresholds ahead of crises and never during non-crisis periods. Empirically, however, type 1 errors (no signal is issued and a crisis occurs) and type 2 errors (a signal is issued but no crisis occurs) are observed. In general, lower thresholds for  $\theta$  predict a higher percentage of crises (as more positive signals are issued), reducing the fraction of type 1 error ( $T_1$ ), but at the cost of predicting more crises that do not occur, raising the fraction of type 2 errors ( $T_2$ ). The optimal indicator has to find the right trade-off. Ultimately, this will depend on the relative costs of type 1 errors versus type 2 errors.

In this box, we explore three different approaches, which minimise different loss functions,  $L$ , with respect to the vector of thresholds,  $\theta$ :

$$\min_{\theta} [L_1] = \min_{\theta} [\alpha T_1 + (1 - \alpha) T_2] \quad (1)$$

$$\min_{\theta} [L_2] = \min_{\theta} [\text{noise - to - signal ratio}] = \min_{\theta} \left[ \frac{T_2}{1 - T_1} \right] \quad (2)$$

$$\min_{\theta} [L_3] = \min_{\theta} \left[ \frac{T_2}{1 - T_1} \mid (1 - T_1) \geq X \right] \quad (3)$$

In the first approach, we minimise the weighted sum of type 1 and type 2 errors, given different weights  $\alpha$  for type 1 and  $(1-\alpha)$  for type 2 errors. This approach would be ideal if policymakers could express their preferences based on views about their relative costs (eg Demirgüç-Kunt and Detragiache (1998)). It requires the costs to be sufficiently measurable and the preferences over them identifiable, which is hard in practice. In the second, we minimise the noise-to-signal ratio (eg Kaminsky and Reinhart (1999)), a very popular method. This in fact amounts to trading off type 1 and type 2 errors in proportion to the noise-to-signal ratio itself. The third, mixed, approach is to minimise the noise-to-signal ratio subject to predicting a minimum percentage of crises,  $X$ . For example, the thresholds chosen by Borio and Lowe (2002a,b) and Borio and Drehmann (2008) are broadly consistent with, although not formally derived from, this method, with minimum thresholds for crises predicted varying between around 60% and two thirds. Of course, if the minimum  $X$  is set to 0, this approach is equivalent to just minimising the noise-to-signal ratio.

The table below illustrates how these different approaches perform over the period used for the in-sample exercise.<sup>1</sup> To save space, we only show the results for the (cumulative) three-year horizon, ie assessing the validity of the signal depending on what happens any time within the three years following the one in which it is issued. We evaluate the indicators based on different weights for type 1 and type 2 error and different thresholds of minimum percentages of crises predicted. Some points are worth highlighting.

### Selecting the optimal indicator<sup>1</sup>

	Weight on type 1 error ( $\alpha$ )					Min N/S	At least x% of crises predicted	
	5	10	25	50	75–95 <sup>2</sup>		0%	66% <sup>3</sup>
Credit and equity gaps								
Credit ( $\theta$ )	8	6	2	2	2	8	2	2
Equity ( $\theta$ )	60	60	60	40	40	60	60	60
Predicted (%)	46	62	77	92	92	46	77	77
Type 2 error (%)	2	3	4	11	11	2	4	4
Noise/Signal	0.04	0.04	0.06	0.12	0.12	0.04	0.06	0.06
Credit and aggregate asset price gaps								
Credit ( $\theta$ )	18	18	6	6	6	18	6	6
AAP ( $\theta$ )	10–20	10–20	10	10	5	10–20	10	10
Predicted (%)	15	15	77	77	85	15	77	77
Type 2 error (%)	0.3	0.3	11	11	27	0.3	11	11
Noise/Signal	0.02	0.02	0.14	0.14	0.32	0.02	0.14	0.14
Credit and either property or equity price gaps								
Credit ( $\theta$ )	8	6	2	2	2	22–24	6	2
Property ( $\theta$ )	40–50	40–50	25	30–50	30–50	10–25	25	40–50
Equity ( $\theta$ )	60	60	60	40	40	20–150	60	60
Predicted (%)	46	62	85	92	92	8	69	77
Type 2 error (%)	2	3	6	11	11	0	4	4
Noise/Signal	0.04	0.04	0.07	0.12	0.12	0.00	0.06	0.06

<sup>1</sup> The estimation period is 1980–2003. The figures refer to the cumulative three-year horizon. N/S = noise-to-signal ratio; AAP = aggregate asset price index. The thresholds  $\theta$  shown are optimal with respect to the criteria listed in the rows of the table. The first set weighs type 1 errors (no signal issued but crises occurred) as indicated, with the corresponding weight on type 2 errors (signal issued but no crises occurred) equal to 1 minus the weight on type 1 error. The second minimises the noise-to-signal ratio. The third minimises the noise-to-signal ratio conditional on at least x% of the crises being predicted. <sup>2</sup> The results are the same for this range of weights. <sup>3</sup> Relative to the minimum of 75% of crises predicted, the 66% minimum is binding only in the case of the indicator that disaggregates property and equity prices.

Sources: National data; authors' calculations.

First, as expected, the more concerned a policymaker is about missing crises (type 1 error), the lower are the critical thresholds to be crossed before signalling crises and the noisier the indicators become. The noise can be quite high. For example, if the policymaker puts at least 75% weight on type 1 error, the corresponding indicators pick between 85 and over 90% of the crises, but with a noise-to-signal ratio in the range of 12 to 32%. This means that even more than one in four signals can be incorrect. In fact, the table above shows that the noise-to-signal ratio can be cut by half while still predicting an acceptable number of crises (see below).

Second, at the other end of the spectrum, minimising the noise-to-signal ratio generally results in an unacceptably low percentage of crises predicted. The percentage of crises predicted is as low as 8 and 15% for two of the indicators, with a noise-to-signal ratio never exceeding 0.04 and being effectively 0 in the case of the indicator that includes both the property and equity price gaps.

On balance, minimising the noise-to-signal ratio subject to at least two thirds of the crises being correctly predicted appears to provide a good compromise and is our preferred criterion.

Depending on the indicator, the noise-to-signal ratio is reduced by at least half compared with assigning a 75% weight to type 1 error. Raising the bar further by setting a floor of at least three quarters of crises predicted has very little effect on the performance of the indicators. The noise-to-signal ratio increases only for the indicator which disaggregates property and equity prices (and beyond the level of accuracy shown in the table). In this case, however, we feel there may be a risk of “overfitting”, given the exceptional performance of the indicators despite the very ambitious floor. If so, better in-sample performance could be gained at the expense of out-of-sample predictive power.

At the same time, the strict statistical approach used in the table can provide a spurious degree of precision. We observe only 13 crises in our sample of 18 countries. This implies that capturing one more crisis increases the percentage of crises predicted by as much as 7.7 percentage points. As non-crisis periods far outnumber crises, percentage changes in type 2 errors are far smaller per observation. Generally, type 2 errors are minimised by higher thresholds. Therefore, a mechanical optimisation procedure implies that any “optimal” indicator will be just at the tipping point of indicating one more crisis: this ensures a given number of predicted crises with the lowest percentage of type 2 errors. Policymakers should keep this in mind and not focus on specific thresholds but look at broad ranges, especially given the concern with out-of-sample performance. This is what we do in the analysis in the main text.

---

<sup>Ⓞ</sup> The optimisation procedure was run using a grid search with a relatively coarse grid. Incremental changes for credit are set at 2, for asset and property at 5 and for equity at 10, so as to avoid misleadingly precise numbers. A different grid will lead to different thresholds. However, as shown in Table 2 in the main text, the performance of the indicators across a range of thresholds is very robust.

## The indicators: recent performance

We now explore formally the performance of various versions of the leading indicator. We first calibrate them in sample, from 1980 to 2003, and then perform an out-of-sample exercise for the years 2004 to 2008.

We consider three versions of the indicator (see the box for a technical description). All of them include a credit gap, but differ in terms of the asset prices included. The first version includes only equity prices as originally specified by Borio and Lowe (2002a,b). The second aggregates equity, commercial and residential property prices based on some rough estimates of their shares in private sector wealth – an aggregate asset price index (Borio et al (1994)). The third splits equities out, but aggregates the two types of property prices. In this case, a signal is issued if the credit gap exceeds the critical threshold together with either the equity or the property price gap. Following previous work, when equities are included separately, the corresponding gap is lagged two periods, in order to take into account the fact that they peak well ahead of a crisis.<sup>7</sup>

Because of limitations in the availability of property prices, the sample covers only 18 industrial countries.<sup>8</sup> A gap is only calculated if at least 10 years of data are available before any prediction is made. This is why the period used for the in-sample calibration of the thresholds is only from 1980

Original indicators extended to incorporate property prices with benefit of more data

---

<sup>7</sup> Equities are not lagged in the aggregate asset price index because the index is seen as a simple measure of aggregate private wealth.

<sup>8</sup> Australia, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the United Kingdom and the United States. Property price indices for residential and commercial property are available for most countries only from 1970 onwards.

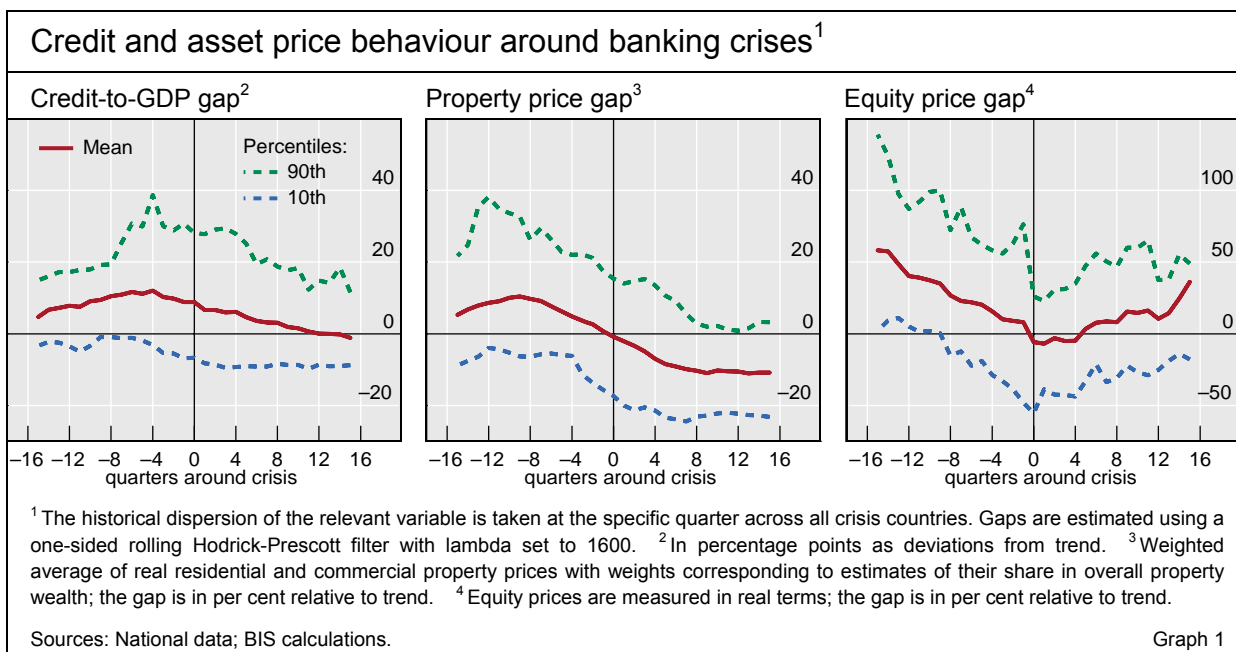
until 2003. In this period, 13 crises occur. The identification and timing of banking crises is based on Borio and Lowe (2002a), who draw on Bordo et al (2001),<sup>9</sup> and extended based on Laeven and Valencia (2008). The data are annual.

Technical features of the indicators

We pay particular attention to the criterion for the choice of “optimal” indicator (see box). Rather than minimising the noise-to-signal ratio per se, we explore different criteria for optimality. The reason is that policymakers may assign more weight to the risk of missing crises (type 1 error) than calling those which do not occur (type 2 error) as the costs of the two differ. Below we initially present our preferred choice: minimising the noise-to-signal ratio subject to predicting at least three quarters of the crises. In our view, given how the indicator behaves, this provides a good balance between identifying costly crises and missing them, without being too ambitious. But we then explore the robustness of the resulting thresholds by checking their sensitivity to specific choices and focus more on ranges.

The use of a flexible horizon means that we consider the performance of the indicator over multiple ones. Specifically, a signal that points to a crisis is judged to be correct if a crisis occurs any time within three possible horizons, namely within one, two and three years ahead, respectively. We therefore expect the performance to improve as the (cumulative) horizon is lengthened.<sup>10</sup>

Before the performance of the indicators is discussed, Graph 1 illustrates



<sup>9</sup> Following Borio and Lowe (2002b), and in contrast to Borio and Lowe (2002a) and to the crises identified by Bordo et al (2001), we added two serious financial stress episodes in the United Kingdom and United States in the early 1990s. These are intended to capture severe financial strains experienced in these economies at the time.

<sup>10</sup> We consider only the first year of any given crisis: correctly predicting a crisis in its second year would be too late; moreover, signals are not designed to predict the length of crises. Technically, we do not use any signals that are issued during the first year of the crises and for the following two years.

the empirical basis for their possible predictive ability. The graph plots the behaviour of credit, equity and property price gaps around the crisis episodes in the sample. It shows that, on average, credit, property and equity price gaps tend to be large and positive in the run-up to crises. In addition, the property and equity price gaps peak well before the crisis, with those of equity prices peaking before property prices and being much larger. By contrast, the credit gap exhibits more inertia. At the same time, there is considerable dispersion around this central tendency.

### *In-sample performance*

The core in-sample results are shown in Tables 1 and 2. Table 1 indicates the performance of the three indicators based on our preferred optimisation criterion. Table 2 explores the sensitivity of the performance to a range of thresholds.

The general performance of the indicators is quite good, confirming previous work. At the three-year (cumulative) horizon, between 69 and 77% of the crises are predicted with a noise-to-signal ratio ranging from 6 to some 14% (Table 1). This means that, for every 20 signals issued, between one and three incorrectly point to a crisis. By construction, the performance tends to improve as the valid horizon over which a crisis may occur is lengthened, as the noise-to-signal ratio necessarily falls.<sup>11</sup>

In-sample performance of the optimal indicators, 1980–2003 <sup>1</sup>									
Horizon (years)	Credit >2 & Equity >60 <sup>2</sup>			Credit >6 & AAP >10 <sup>2</sup>			Credit >6 & (Property >25 or Equity >60) <sup>2</sup>		
	Pred (%) <sup>3</sup>	Type 2 error (%)	Noise/Signal	Pred (%) <sup>3</sup>	Type 2 error (%)	Noise/Signal	Pred (%) <sup>3</sup>	Type 2 error (%)	Noise/Signal
1	46	7	0.16	54	14	0.27	46	6	0.13
1, 2	62	5	0.09	69	12	0.18	62	5	0.08
1, 2, 3	77	4	0.06	77	11	0.14	69	4	0.06

<sup>1</sup> Optimal indicators are chosen based on minimisation of the noise-to-signal ratio conditional on capturing at least two thirds of the crises over a cumulative three-year horizon (see box). A signal is correct if a crisis occurs in any of the years included in the horizon ahead. The noise is measured by the wrong predictions within the same horizon. <sup>2</sup> All variables are measured as gaps, ie as percentage point (credit-to-GDP ratio) or as percentage deviation (asset price indices) from an ex ante (one-sided), recursively calculated Hodrick-Prescott trend with lambda set to 1600. Numbers that follow the sign ">" indicate the critical threshold. Credit is the ratio of private sector credit to GDP. Equity is the (real) equity price (stock market) index, lagged by two periods. AAP is the (real) aggregate asset price index, which combines equity prices and residential property and commercial property prices based on rough estimates of their shares in private sector wealth. Property is the price index that combines residential and commercial property prices, based on the weights used in the AAP. <sup>3</sup> Percentage of crises predicted (1 minus type 1 error).

Sources: National data; authors' calculations. Table 1

<sup>11</sup> Closer examination reveals that a number of these signals are "wrong" only in the sense that the indicators *start* going on too early, ie they signal a crisis which will materialise in four to five years. Similarly, the percentages of crises predicted over a one-year horizon tend to be rather low compared with those over longer ones. The reason is precisely that the indicators are designed to identify risks of distress during boom conditions, and before crises emerge asset prices gaps narrow as asset prices soften, possibly switching signals off. These observations indicate that the indicators' lead is quite long. Moreover, they also suggest that, if so desired, calibration could be quite successful also starting the valid interval of prediction not one, but as far as three years ahead, as done in Borio and Lowe (2004), where the relevant interval is three to five years ahead.



Sensitivity of the indicators to different thresholds									
Horizon (years)	Predicted <sup>1</sup>			Type 2 error			Noise-to-signal ratio		
	Mean <sup>2</sup>	Min <sup>2</sup>	Max <sup>2</sup>	Mean <sup>2</sup>	Min <sup>2</sup>	Max <sup>2</sup>	Mean <sup>2</sup>	Min <sup>2</sup>	Max <sup>2</sup>
Credit (4–6) & Equity (40–60) <sup>1</sup>									
1	47	38	54	7	4	10	0.15	0.10	0.19
1, 2	62	54	69	6	3	9	0.10	0.06	0.13
1, 2, 3	69	62	77	6	3	8	0.08	0.04	0.11
Credit (4–6) & AAP (5–10) <sup>1</sup>									
1	63	54	77	18	14	23	0.28	0.26	0.30
1, 2	75	69	77	16	13	21	0.22	0.18	0.28
1, 2, 3	77	77	77	15	11	20	0.19	0.14	0.26
Credit (4–6) & (Property (15–25) or Equity (40–60)) <sup>1</sup>									
1	56	46	62	10	6	15	0.18	0.12	0.25
1, 2	71	62	77	9	5	14	0.12	0.08	0.18
1, 2, 3	72	69	77	8	4	13	0.11	0.06	0.18
<sup>1</sup> Percentage of crises predicted (1 minus type 1 error). <sup>2</sup> Mean, minimum and maximum of the percentage of crises predicted, type 2 error (in per cent) and the corresponding noise-to-signal ratio if the thresholds for the indicator vary between the numbers in brackets.									
Sources: National data; authors' calculations.								Table 2	

In-sample, indicators show strong performance ...

Across the different types of indicator, the picture varies somewhat. In sample, the credit-cum-equity indicator performs remarkably well. At a three-year horizon, it captures the highest percentage of crises (77%) with the lowest noise-to-signal ratio (6%). Separating out property prices, however, improves the performance slightly at the one- and two-year horizons. The aggregate asset price index is not as good as the other two indicators. It performs better at shorter horizons only if a high value is attached to predicting crises correctly at the expense of issuing wrong positive signals. This probably reflects the loss in predictive content that results from not lagging equity prices once they are aggregated with property.

... and are robust to different specifications

The performance of the indicators is quite robust to the specific choice of the threshold (Table 2). For example, for the disaggregated indicator, ranges that vary between as far as 4 and 6 (credit), 15 and 25 (property) and 40 and 60 (equities) yield a range of crises predicted between 69 and 77% over a three-year horizon, with an average noise-to-signal ratio of 0.11, varying from 0.06 to 0.18.<sup>12</sup> This is encouraging for policy purposes, in the sense that the success of the indicator does not hinge on very specific combinations of thresholds. In assessing the out-of-sample performance, therefore, we will consider these ranges rather than taking particular point estimates of the thresholds too literally (see also the box).

<sup>12</sup> Similarly, calibrating the thresholds so as to optimise the performance of the indicator over a cumulative two-year horizon, instead of the three-year one as shown in Table 1, changes the specific thresholds but has little impact on the performance (not shown).

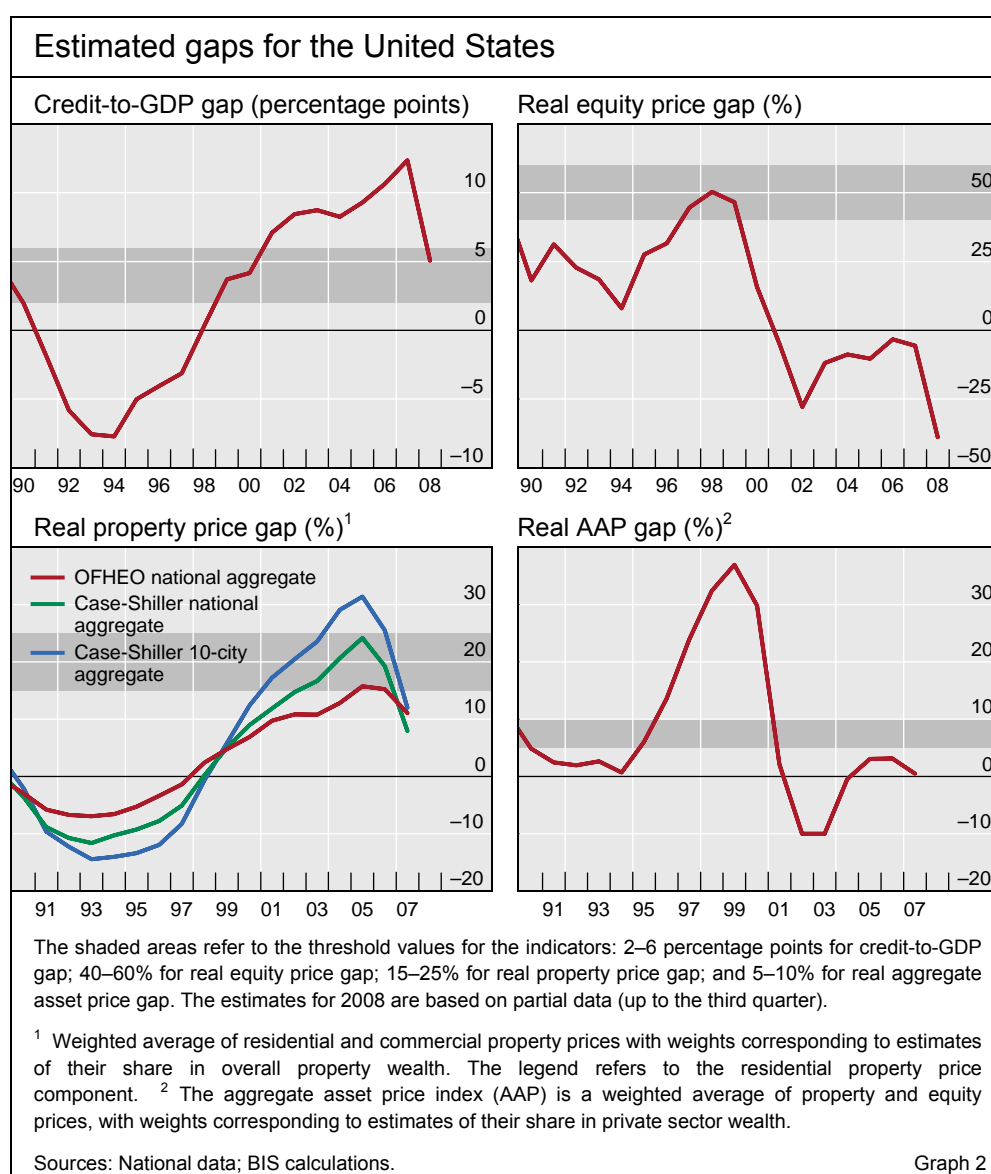
### Out-of-sample performance

How do the indicators perform out of sample, from 2004 to 2008? We explore this question in two steps. First, we consider the United States, the epicentre of the current crisis and one where financial distress has been particularly acute. It would be problematic if the indicators failed to issue warnings for that country. We then assess their performance across countries.

#### (i) The case of the United States

Graph 2 plots the behaviour of the various gaps for the United States, together with the discussed ranges for the thresholds. While the credit gap indicates a potential build-up of vulnerabilities at least as early as 2001, when it crosses the relatively strict threshold of 6%, the performance of the overall indicator varies depending on how asset prices are treated.

The graph indicates that those indicators that rely on equity prices on their own would have failed to signal the build-up of risks. Admittedly, if calibrated on pre-2000 data, the indicator would have given some warnings of the



The indicators that disaggregate property prices signal rising risks before the current crisis in the United States ...

impending strains and recession associated with the dotcom bust.<sup>13</sup> But the bust, despite the subsequent recovery, undermines the information content of the gap in the more recent period. Moreover, the weight of equity prices in the aggregate asset price index is such that the same shortcomings are transferred to the corresponding indicator.

By contrast, the graph suggests that the indicator that treats equity and property prices separately would have picked up the vulnerabilities. How early depends on the specific thresholds and property price series used (shaded area in the graph).<sup>14</sup> Signs of financial imbalances began to emerge as far back as the beginning of the century, as both the credit gap and the property price gap started to exceed indicative thresholds jointly. If the residential component of the property price index is measured by the Case-Shiller 10-city index, the strictest criterion, which has the property price gap exceeding 25%, is met as early as 2004. On the other hand, if the much less variable OFHEO index is used, the property price gap peaks at nearly 16% in 2005.

(ii) *The cross-country experience*

Extending the out-of-sample exercise to all the industrial countries in our sample is harder to perform at this early stage, in the midst of the crisis. At least two problems arise. First, given that the flexible horizon extends up to three years, we can only fully assess the predictive content of the signals issued in 2005; for subsequent ones, the full horizon has not yet materialised. This is an issue whenever banking distress has not yet emerged. Second, and more importantly, defining which country is in distress can be ambiguous. The datasets that identify the crises used in sample have not as yet been extended to cover the recent episode.

To address the ambiguity in the identification of the crisis, we adopt two definitions, going from the more to the less restrictive:

Definition 1: Countries where the government had to inject capital in more than one large bank and/or more than one large bank failed.

Definition 2: Countries that undertook at least two of the following policy operations: issue wholesale guarantees; buy assets; inject capital into at least one large bank or announce a large-scale recapitalisation programme.

Which definition of distress is more appropriate? Definition 1 may be too narrow, and definition 2 too broad as it may include cases where measures are only announced as a precaution or in response to policies adopted in other countries. The extension of guarantees to prevent a drain of funding in the

---

<sup>13</sup> This is based on the thresholds highlighted in the original Borio and Lowe (2002a) study, on a different sample of countries and different period, ending before 2000. Technically, if taken literally, the indicator was very close, but did not quite issue a full signal. While the equity gap, at 46.6, did indeed breach the threshold of 40, the credit gap reached 3.7, slightly below 4.

<sup>14</sup> The in-sample results indicated in the table use the Case-Shiller national home price index extended backwards with the OFHEO national aggregate house price index using the first common period link method. We examined the sensitivity of these in-sample results to the choice of these two indices as well as to the Case-Shiller 10-city index and found that they were quite robust, resulting only in very small changes in the noise-to-signal ratio.

domestic market is an obvious such example. For instance, it might be argued that, so far, despite the measures taken, the actual strains faced in Australia, Canada and Italy have been quite mild. Together, however, the two definitions encompass a reasonable range.

By the end of January 2009, based on definition 1, seven countries had faced a crisis: the United States, the United Kingdom, Belgium, France, Germany, Ireland and the Netherlands. Based on definition 2, 14 out of the 18 countries had faced distress: the ones just mentioned plus Australia, Canada, Denmark, Italy, Spain, Sweden and Switzerland. In all countries, the criteria for a crisis are fulfilled only in 2008.<sup>15</sup>

As in the case of the United States, the indicators based exclusively on credit and equity prices fail to issue warning signals (not shown in the table). The likely reason is the longer lag between peaks in equity and property prices compared with the experience in sample. In sample, equity prices typically peaked two years before property prices (BIS (1993), Borio and McGuire (2004)). Using equity prices, with the appropriate lead, could thus also partially proxy for them.<sup>16</sup> In the current episode, however, equity prices peaked in the early 2000s and property prices peaked around 2006 or later. In the late 1980s, the time of the previous property boom in industrial countries, monetary policy had to be tightened in several countries in order to fight emerging inflation pressures, thereby triggering the reversal in property prices. This was not the case in more recent years, as inflation pressures remained subdued until at least 2006.

... as well as in a number of other countries

Not surprisingly, the performance of the indicator that includes also the property price gap is encouraging, although far from perfect (Table 3).<sup>17</sup> The variant of the indicator based on the lowest threshold for property prices (15%) performs best; that based on the top of the range (25%) appears too strict. The lower bound predicts over 50% of the crises, regardless of the definition; the higher bound only the one in the United States, and based on the Case-Shiller 10-city index. Inevitably, not least given the small sample, the noise-to-signal ratios increase substantially compared with the in-sample estimates.<sup>18</sup>

A look behind the aggregate numbers is instructive. Using definition 2, three false positive signals are issued: for Finland, Norway and New Zealand.

---

<sup>15</sup> The exception is Denmark, where some measures were taken only in January 2009. For present purposes, we treat them as if they had been taken in 2008.

<sup>16</sup> Moreover, as found in Borio and Lowe (2002b) for emerging market economies, a measure of (real) exchange rate appreciations could also help. The reason is that appreciations tend to go hand in hand with the capital flow surges that typically fuel property price booms.

<sup>17</sup> The performance of the indicator with the aggregate asset price index falls somewhere in between the indicators discussed.

<sup>18</sup> The increase arises for three reasons. First, for some countries signals are issued even though no crises according to the definition used have as yet materialised. Second, most type 2 errors occur for countries for which signals are issued quite early, eg in 2004, which is too early for our three-year horizon, even if the crisis eventually does materialise. Finally, given that the out-of-sample exercise only covers the period from 2004 to 2008, the number of "non-crisis" periods is very low, which can lead to large swings in the noise-to-signal ratio in response to small changes in the absolute number of type 2 errors.

Out of sample performance, 2004–08						
	Crisis definition 1 <sup>1</sup>			Crisis definition 2 <sup>1</sup>		
Horizon (years)	Predicted <sup>2</sup>	Type 2 error (%)	Noise/Signal <sup>3</sup>	Predicted <sup>2</sup>	Type 2 error (%)	Noise/Signal <sup>3</sup>
Credit >4 & (Property >15 or Equity >40)						
1	29	38	1.35	29	40	1.39
1, 2	57	36	0.63	50	36	0.73
1, 2, 3	57	35	0.62	50	33	0.67
Credit >6 & (Property >20 or Equity >60)						
1	0	18	–	7	19	2.66
1, 2	29	17	0.60	29	16	0.56
1, 2, 3	29	18	0.62	29	17	0.47
Credit >6 & (Property >25 or Equity >60)						
1	0	6	–	0	7	–
1, 2	14	5	0.36	7	7	0.95
1, 2, 3	14	4	0.27	7	7	0.93
<sup>1</sup> Crisis definition 1: Countries where the government had to inject capital in more than one large bank and/or more than one large bank failed (seven crises). Crisis definition 2: Countries that undertook at least two of the following policy operations: issue wholesale guarantees; buy assets; inject capital into at least one large bank or announce a large-scale recapitalisation programme (14 crises). Signals are assessed over a three-year horizon. <sup>2</sup> Percentage of crises predicted (1 minus type 1 error). <sup>3</sup> If no crisis is predicted, the noise-to-signal ratio cannot be calculated.						
Sources: National data; authors' calculations.						Table 3

The latter two countries have already taken extra policy measures to enhance financial stability, but without meeting the criteria of our definitions.<sup>19</sup>

However, they do not signal potential vulnerabilities emerging from cross-border exposures ...

By contrast, the countries that are missed, based on the lower bound of the range, are Germany and the Netherlands (definitions 1 and 2) as well as Switzerland and Canada (definition 2). The indicator does not capture these cases as banks have run into trouble as a result of losses on their international exposures in the absence of clear signs of financial imbalances in the domestic economy. This is no surprise, since by construction the indicator assumes that banks in any given country are exposed only to the financial cycle in that country.

On balance, these findings suggest that the recent credit crisis confirms the usefulness of this type of indicator. At the same time, they point to some of its limitations and the potential scope for improvement. A key limitation is its failure to consider cross-border exposures, to which we turn next.

### The indicators: cross-border exposures

One possible way of incorporating the risks arising from the cross-border exposures of a banking system whilst maintaining the underlying logic of the

<sup>19</sup> New Zealand introduced retail as well as wholesale guarantees, and Norway introduced a programme allowing banks to swap collateralised debt obligations for government securities.

indicator is the following.<sup>20</sup> First, establish the geographical distribution of the foreign exposures of the institutions in the system. Second, calculate the preferred indicator(s) for the countries to which those institutions are exposed. Check if a signal is issued (the value of the indicator is 1), or not (the value is 0). Third, calculate a weighted average of the indicator (0 or 1) for the overall balance sheet exposure, including that to the domestic market. The resulting number, which varies between 0 and 1, is an index of the riskiness of the overall portfolio. A value of 1 would indicate that all the exposures of a banking system are to countries for which the indicator signals future banking distress; a value of 0, that none is. Finally, if the data went back sufficiently in time and covered a sufficient number of crises, one could go one step further. It would in turn be possible to calculate critical thresholds for this derived index, and seek to predict crises along similar lines as before.<sup>21</sup>

Ideally, this exercise would be performed based on individual bank data and a full picture of the exposures. In practice, this is not possible except for national supervisors, as this type of information is not publicly available. The BIS international banking statistics, however, can provide useful information at the national banking system level. The data are drawn from the consolidated banking statistics, which capture the exposures of reporting banks to counterparties, regardless of the location of the office from which the funds are provided. These data include a counterparty breakdown (interbank, public sector, non-bank private sector) and therefore allow different possible aggregations.

... which can partly be captured by the BIS international banking statistics

There are two main drawbacks of any such exercise, combining as it does the BIS statistics with other data sources. First, the BIS statistics include information on reporting countries as counterparties only since 1999. This implies that the consolidated exposures to industrial countries are available only as from that date. More importantly, because of limitations on the availability of property prices, the indicator of domestic financial imbalances cannot be constructed for most of the countries outside our sample, which includes emerging market economies. This means excluding those foreign exposures from the analysis.

As a result, at this stage we can only perform an indicative exercise. We can calculate the weighted average of the riskiness of foreign exposures of a given banking system in the years just prior to the crisis, but are unable to estimate the critical values of the index. Moreover, that weighted average is not complete, as we are unable to construct a leading indicator of banking distress for a varying, at times sizeable, portion of the foreign exposures.

---

<sup>20</sup> Another dimension of the cross-border exposures is direct cross-border lending *into* a given country. The figures that we have used in this feature are based on national statistics. As such, they only include lending by institutions *located* in a given country. The BIS statistics could also be used to remedy this deficiency. We leave this potential improvement to future work.

<sup>21</sup> An alternative to this two-step procedure would be to estimate the thresholds specific to a given banking system in one go, based on the information of the geographical distribution of its exposures.

Table 4 summarises the results of the exercise. The left-hand side of the table provides a weighted average of the riskiness of the foreign exposures, based on two representative thresholds of the disaggregated indicator that incorporates property prices. It also indicates the percentage of foreign exposures covered. The right-hand side includes an estimate of the riskiness of the domestic portfolio, with its size approximated by the private sector domestic credit aggregate used in the previous analysis.<sup>22</sup> It also shows the weight of the foreign exposures in the overall portfolio.

Two points stand out. First, the riskiness of the cross-border exposures of the banking systems for which the indicator failed to predict crises in the previous analysis is considerably higher than that of their domestic ones. This partly helps to explain the financial strains incurred in those systems. For example, the ranges for the index of foreign exposures for Germany and

Indicators weighted by domestic and international exposures <sup>1</sup>						
	Foreign <sup>2</sup>			Foreign plus domestic <sup>3</sup>		
	Credit >4 & (Property >15 or Equity >40)	Credit >6 & (Property >20 or Equity >60)	% of foreign portfolio captured	Credit >4 & (Property >15 or Equity >40)	Credit >6 & (Property >20 or Equity >60)	Foreign as % of total portfolio captured
Belgium	0.62	0.36	81	0.32	0.18	48
Canada	0.89	0.69	84	0.14	0.11	14
Germany	0.79	0.46	78	0.30	0.18	34
France	0.59	0.38	78	0.87	0.82	27
Ireland	0.77	0.27	88	0.94	0.11	43
Italy	0.35	0.21	62	0.05	0.03	14
Japan	0.79	0.65	73	0.14	0.11	15
Netherlands	0.67	0.44	84	0.31	0.21	43
Norway	0.73	0.47	67	0.98	0.03	6
Spain	0.78	0.28	69	0.97	0.87	15
Sweden	0.60	0.31	80	0.87	0.77	27
Switzerland	0.80	0.61	77	0.48	0.40	47
United Kingdom	0.68	0.60	73	0.92	0.26	32
United States	0.54	0.14	63	0.99	0.98	3

<sup>1</sup> Sum of indicators corresponding to the country to which the banks headquartered in the country shown in the table are exposed, weighted by the share of the exposure in the portfolios indicated (foreign and foreign plus domestic). Indicator is 1 if thresholds were exceeded in any of the years 2005 to 2007 in a particular country. <sup>2</sup> Foreign claims are cross-border claims plus locally booked claims in all currencies on residents of a given country. Only claims on banks and the non-bank private sector considered. <sup>3</sup> Domestic and foreign claims on non-bank private sector.

Sources: BIS; authors' calculations. Table 4

<sup>22</sup> In technical terms, the "foreign" index of riskiness  $FIR_i$  for country  $i$  is  $FIR_i = \sum_{j \neq i} (E_j * S_j / \sum_j E_j)$ , where  $S_j$  is the signal in country  $j$  and  $E_j$  are all cross-border claims vis-à-vis country  $j$  plus locally booked claims in all currencies on residents of country  $j$ . All foreign claims are on an ultimate risk basis, and only claims on banks and the non-bank private sector are considered. The combined "foreign and domestic" index  $FDIR_i$  is constructed as  $FDIR_i = [\sum_{j \neq i} (E_j * S_j / (\sum_j E_j + D_i)) + D_i * S_i / (\sum_j E_j + D_i)]$ , where  $S_j$  and  $E_j$  are defined as in the case of  $FIR$  except that  $E_j$  only takes account of foreign claims on the non-bank private sector, in order to increase the comparability of the figures as no information on domestic interbank exposures is available.  $S_i$  is the signal in the home country and  $D_i$  is domestic credit to the non-bank private sector. For a more detailed description of the international banking statistics, see McGuire and Wooldridge (2005).

Switzerland are 0.46–0.79 and 0.61–0.80, respectively, depending on the thresholds chosen. For Switzerland, in particular, this raises the index for overall exposures from 0 to nearly 0.5; the impact on Germany is lower, owing to the smaller relative weight of cross-border assets. Second, for most banking systems in the sample the riskiness of foreign exposures is quite high. This suggests that ignoring them could miss a significant source of vulnerabilities. One exception is Italy (0.21–0.35). For that country, however, we only capture a comparatively low percentage of cross-border exposures (slightly above 60%), most of which is to Germany.

## Conclusion

This special feature suggests that it is possible to build relatively simple indicators that can help inform assessments of the build-up of risks of future banking distress in an economy. These indicators are based on the coexistence of unusually strong and protracted increases in credit and asset prices. We find that they perform reasonably well also out of sample, as indicated by their ability to point to potential banking distress ahead of the current crisis.

At the same time, a number of caveats should be borne in mind. First, the analysis confirms the critical role of judgment. And for some, this role may be uncomfortably large. The out-of-sample performance is not an unqualified success. The indicators would have failed in recent years had they been based exclusively on equity prices, which perform so well in sample. The extension to property prices is essential for the current episode. Similarly, we caution against deciding on “optimal” performance in sample purely based on strict statistical criteria, without acknowledging the “fuzzy” nature of the exercise. This, too, could have failed to identify the risks correctly. For policy purposes, we support the use of ranges rather than point thresholds. Second, a full assessment of the indicators’ performance will require more time, as the current financial strains are still unfolding.

The indicators could be improved in several dimensions. First, one could seek to incorporate cross-border exposures more systematically. While the BIS international banking statistics can be helpful, they do not provide a complete picture. This would require specific data collection efforts at the national level. Similarly, considering the information content of more global measures of credit and asset price increases, rather than country-by-country, could help to better capture the international dimension of the problems. Second, one could seek to make improvements to the individual series included. It is worth exploring how to overcome the current heterogeneity of the property price series across countries. Efforts by national authorities to improve the underlying data, in terms of both quality and historical availability, could be extremely useful. Third, the performance of further asset price series could be examined. Beyond exchange rates, as in Borio and Lowe (2002b), credit risk spreads merit particular attention: prolonged periods of unusually *low* credit risk spreads during expansion phases would signal potential stress further down the road. Finally, one could improve on the measures of “leverage” included. For



example, the indicators do not consider leverage *within* the financial system itself, which appears to have been so prominent in the current episode. We would conjecture, however, that the basic architecture of the indicators would survive. This would involve the *coexistence* of a measure of asset price misalignments with one that captures the limited shock absorption capacity of the economy and hence its “leverage”.

## References

Alessi, L and C Detken (2008): “Real time early warning indicators for costly asset price boom/bust cycles: a role for global liquidity”, paper presented at the Euro Area Business Cycle Network and CREI conference on *Business cycle development, financial fragility, housing and commodity prices*, Barcelona, 21–23 November.

Allen, F and D Gale (2000): “Bubbles and crises”, *The Economic Journal*, no 110, pp 236–56.

Bank for International Settlements (1993): *63rd Annual Report*, Basel, June.

Bordo, M, B Eichengreen, D Klingebiel and M S Martinez-Peria (2001): “Financial crises: lessons from the last 120 years”, *Economic Policy*, April.

Borio, C and M Drehmann (2008): “Towards an operational framework for financial stability: ‘fuzzy’ measurement and its consequences”, 12th Annual Conference of the Central Bank of Chile, *Financial stability, monetary policy and central banking*, Santiago, 6–7 November.

<http://www.bcentral.cl/eng/conferences-seminars/annual-conferences/2008/program.htm>

Borio, C, N Kennedy and S Prowse (1994): “Exploring aggregate asset price fluctuations across countries: measurement, determinants and monetary policy implications”, *BIS Economic Papers*, no 40, April.

Borio, C and P Lowe (2002a): “Asset prices, financial and monetary stability: exploring the nexus”, *BIS Working Papers*, no 114, July.

——— (2002b): “Assessing the risk of banking crises”, *BIS Quarterly Review*, December, pp 43–54.

——— (2004): “Securing sustainable price stability: should credit come back from the wilderness?”, *BIS Working Papers*, no 157, July.

Borio, C and P McGuire (2004): “Twin peaks in equity and housing prices?”, *BIS Quarterly Review*, March, pp 79–93.

Davis, E P and D Karim (2008a): “Comparing early warning systems for banking crises”, *Journal of Financial Stability*, vol 4, June, pp 89–120.

——— (2008b): “Could early warning systems have helped to predict the sub-prime crisis?”, *National Institute Economic Review*, vol 206(1), pp 25–37.

Demirgüç-Kunt, A and E Detragiache (1998): “The determinants of banking crises: evidence from developing and developed countries”, *IMF Staff Papers*, vol 45, pp 81–109.

——— (2005): “Cross-country empirical studies of systemic bank distress: a survey”, *IMF Working Paper*, no 05/96, May.

Kaminsky, G L and C M Reinhart (1999): “The twin crises: the causes of banking and balance-of-payments problems”, *American Economic Review*, vol 89(3), pp 473–500.

Kindleberger, C (2000): *Manias, panics and crashes*, Cambridge University Press, Cambridge, 4th edition.

Laeven, L and F Valencia (2008): “Systemic banking crises: a new database”, *IMF Working Paper*, WP/08/224.

McGuire, P and P Wooldridge (2005): “The BIS consolidated banking statistics: structure, uses and recent enhancements”, *BIS Quarterly Review*, September, pp 73–86.

Minsky, H P (1982): “Can ‘it’ happen again?”, *Essays on Instability and Finance*, Armonk: M E Sharpe.

Tarashev, N (2008): “An empirical evaluation of structural credit risk models”, *International Journal of Central Banking*, vol 4(1), March, pp 1–53. Also available as *BIS Working Papers*, no 179, July 2005.