Two new tools to assess risks and vulnerabilities in the spanish housing market

Pana Alves
Banco de España

The housing market plays an important economic and social function[1]. The activity linked to this market has an important weight on the economy, as a source of job creation, investment and economic growth (ESRB, 2019a). A significant part of housing purchases is financed with credit and housing loan portfolios also often make up a large part of banks' balance sheets, therefore the housing market has relevant implications for financial stability. Given all of the above, having an appropriate set of indicators that enable the correct monitoring of vulnerabilities and risks in this sector is crucial. In addition, such risk identification metrics serve to guide decision-making on activating, deactivating and calibrating the macroprudential tools[2].

This paper presents a methodology developed by the Banco de España[3] for monitoring and evaluating risks and vulnerabilities in the residential real market in Spain. This methodology consists of two

1 The Spanish residential real estate accounts for 70 % of households' asset holdings and housing loans are the most common form of their debt (around 73 %).

2 See ESRB (2013).
3 See Alves et al (2023).

TABLE 1. HOUSING MARKET INDICATORS BY CATEGORY

Category	Indicators
Real activity and demographic factors	Housing approvals House purchases Building construction workers registered with Social Security Difference between housing starts and change in households
Credit developments	Stock of loans for construction and real estate activities Stock of housing loans New housing loans Non-performing loans for house purchase
Credit standards for new loans	Loan-to-value (LTV) ratio Loan-to-price (LTP) ratio Overvaluation ratio (LTP/LTV) Maturity period Proportion of mortgages with certain maturity and LTP characteristics Interest rates
Household financial position	Financial and total wealth Indebtedness Debt burden Saving rate
House valuation	Real transaction prices Valuation models: i) deviation of the housing prices from its long-term equilibrium level, ii) deviation of the housing prices from the Hodrick-Prescott trend, iii) deviation of the price-to-gross disposable income (GDI) ratio from its historical average and iv) the residual of the regression of housing prices on the GDI. Affordability Gross housing rental yield Housing rental prices

Source: Banco de España.

tools: a heat map and a synthetic index. The heat map is a data visualization technique that color-codes the magnitude of the risks. The variations in color, as well as possible changes in the shade, allow the indicators to be read over time. Thus, drawing on a selection of key housing market indicators, a heat map is established in which the observed values of each indicator are evaluated against certain critical thresholds. Meanwhile, the synthetic risk indicator is constructed by aggregating all the information obtained from the individual indicators, and detects whether and to what extent the degree of vulnerability in the housing market increases or decreases over time.

1. Heat map

To construct the heat map 26 indicators were initially selected. Those variables contain the information considered most relevant on various aspects characterizing the Spanish residential real estate market. Therefore, these indicators cover features ranging from real activity (such as the number of house being purchased) and financing (the amount of mortgages granted and the conditions of those loans) to price developments (how much transaction and rental prices change) and household financial position (indebtedness and saving rate). The selection of indicators was done taking into account all the sources available that provided information on the housing market in Spain (either publicly available or not). The Banco de España created an interdisciplinary panel of experts to select the most relevant indicators with three objectives in mind: give a holistic view of the housing market, consider the advantages and disadvantages of using different indicators and sources of information and the relevance of the variables. Table 1 summarizes the 26 indicators used and classified them into five categories.

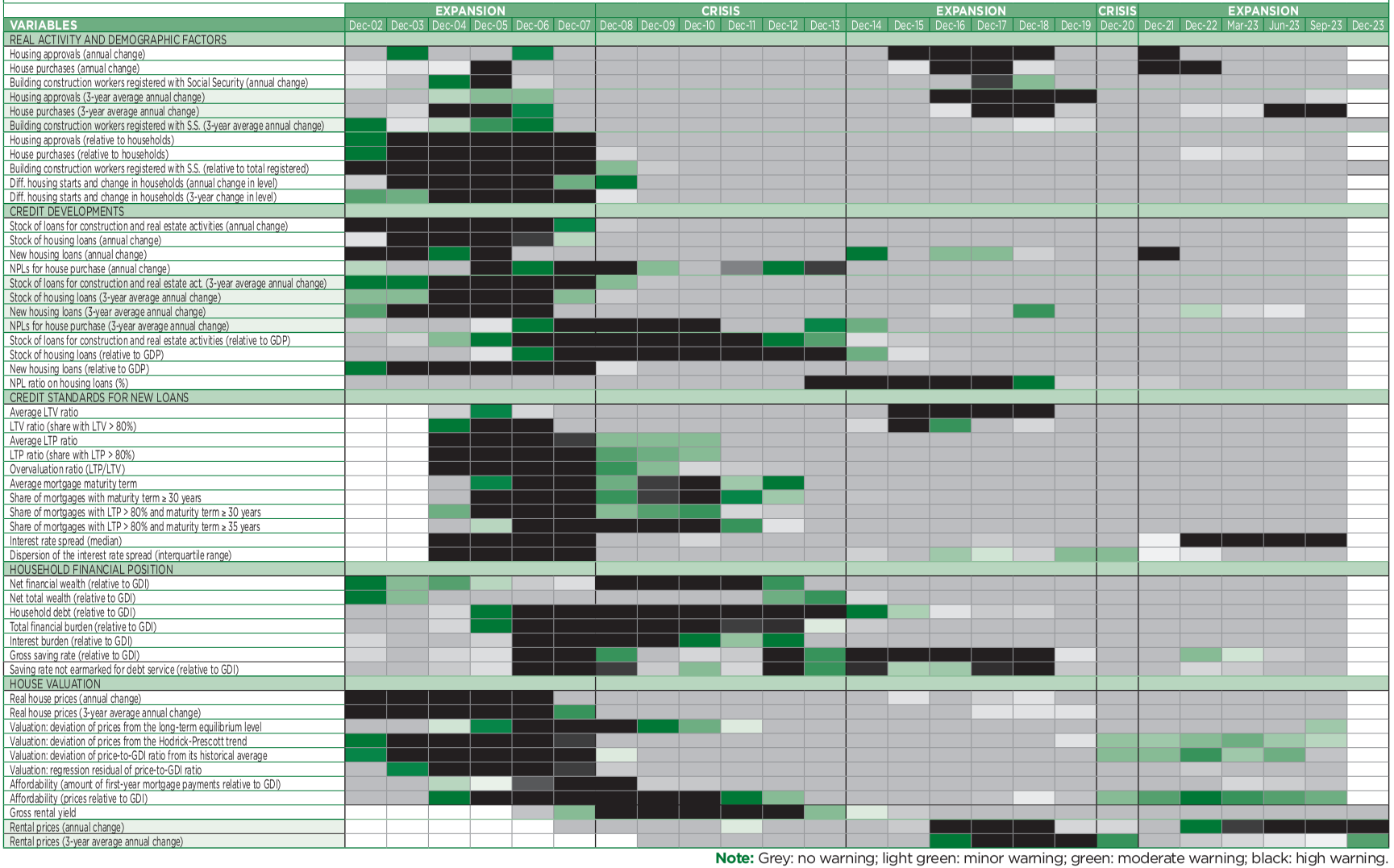
The indicators considered only provide information at the aggregate level (national), therefore do not take into account more granular developments that could potentially be relevant for identifying risks. The heat map is constructed by drawing on changes in the original indicators selected and calculating some transformations (such as ratios between different variables, rates of change, and differences between the levels of two variables). Some of the variables used, such as some house valuation indicators, are derived from the results of econometric modelling. Thus, the original 26 datasets yield a final set of 52 indicators, following the various

TABLE 2. HEAT MAP COLOR CODES

Right-tail indicators					
	Increasing values → Greater vulnerability				
Warning level	NO WARNING		MINOR	MODERATE	HIGH
Percentile	Below 50		Between 50 and 69	Between 70 and 84	85 or above
Left-tail indicators					
	Increasing values → Lesser vulnerability				
Warning level	HIGH	MODERATE	MINOR	NO WARNING	
Percentile	Below 15	Between 15 and 29	Between 30 and 49	50 or above	
Indicators for house valuation models					
	Increasing values → Greater overvaluation				
Warning level	NO WARNING		MINOR	MODERATE	HIGH
Threshold and percentile	0 or negative		Below 40	Between 40 and 69	70 or above
			Percentiles calculated using only positive values		

Source: Banco de España.

TABLE 3. HEAT MAP OF THE SPANISH HOUSING MARKET SINCE 2002



Note: Grey: no warning; light green: minor warning; green: moderate warning; black: high warning.
Source: Banco de España.

transformations and manipulations of the original data[4].

The heat map is created by associating a color to each indicator value according to the warning level. Four colors are used: grey (no warning), light green (minor warning), green (moderate warning) and black (high warning). The warning level thresholds are calculated drawing on the historical percentiles of each indicator's distribution[5]. As shown in Table 2, in the case of right-tail indicators (the majority of the variables), for which an increase in value denotes a higher risk, grey is assigned when the value is below the 50th percentile of the historical statistical distribution, light green when it is between the 50th and 69th percentiles, green when it is between the 70th and 84th percentiles, and black when it is above the 84th percentile. In the case of the left-tail indicators, for which a lower value represents a higher risk[6], the colors are assigned symmetrically with the indicators mentioned above. Finally, as regards the four indicators for the house valuation models[7], which are also right-tail, a different procedure is applied. Grey is allocated when the value is zero or negative, as in such cases there are no signs of overvaluation. A positive value indicates there are signs of overvaluation and thus triggers a warning, whose color is assigned based on the historical percentiles of the distribution, calculated just with positive values. Specifically, light green is assigned when the value is below the 40th percentile of the distribution, green when it is between the 40th and 69th percentiles, and black when it is above the 70th percentile. This procedure was established to generate the same relative size of the observations located in the warning zone (color other than grey) as in the two foregoing cases (right-tail and left-tail indicators).

Table 3 presents the results of heat map at the year-end since 2002, and with quarterly frequency for 2023.

2. Synthetic risk indicator

The synthetic index condenses into a single indicator the level of housing market vulnerability. This aggregate indicator is constructed from a subset of the individual heat map indicators (20 in total[8]), and is partly inspired by similar exercises done by other authorities (ESRB, 2019b). The methodology used consists of three stages (see Figure 1). In the first stage, given their very different statistical distributions, indicators need to be transformed, to standardize them. This yields new variables with the same scale and distribution *(level 1)*. In the second stage, all the individual variables standardized are aggregated into four categories (real activity, house valuation, credit conditions and household financial position) to produce four intermediate indices. To do so, linear combinations of these indicators with optimal weights for each of the four categories are calculated by conducting a principal component analysis *(level 2)*. These weights are optimal because they summarize the redundant information in the individual indicators by drawing on a statistical procedure. Finally, in stage three, the synthetic index used a simple average of the intermediate indices *(level 3)*.

4 For a complete description of the original 26 variables used, together with their definitions, the transformations carried out for each time series and how they were constructed, their source of information, frequency and the date from when they are available, see Table A.1 in Alves et al. (2023).

5 In the years leading up to the 2008 global financial crisis, risk-taking was clearly excessive and inclusion of this period could influence the tolerance level of the indicators. Therefore, an alternative exercise has been carried out excluding the period 2003-2007. The results show that, broadly speaking, there would be no significant changes on the warning signals.

6 Specifically, those indicators are the ones related to: interest rates, households' wealth, the saving rate and gross rental yield.

7 Specifically, this criterion applies to the following indicators: deviation of the housing prices from its long-term equilibrium level and deviation of the housing prices from the Hodrick-Prescott trend, deviation of the price-to-gross disposable income (GDI) ratio from its historical average and the residual of the regression of housing prices on the GDI.

8 Out of the 26 original indicators included in the heat map, six are excluded to construct the synthetic index: building construction workers registered with Social Security, non-performing loans for house purchase, financial wealth, total wealth, gross rental yield and rental prices.

2.1. Transformation of the indicators

In order to transform the indicators, their empirical cumulative distribution function (ECDF) is used, in line with the literature on financial stress indicators. The calculation of the ECDF is relatively immediate. First, the observed values of the indicators ×t with sample size T are ordered so that for each original time series of indicator $x_t = (x_1, x_2, ..., x_T)$, a new time series is obtained with its values ordered $x_{[t]} = (x_{[1]}, x_{[2]}, ..., x_{[T]})$. In this transformed series, $x_{[1]}$ represents the indicator's lowest value and $x_{[T]}$ its highest value. To obtain the transformed indicator z_t, the numerical ranking is then assigned to each x_t (r) value and this result is divided by sample size T:

$$z_t = r/T \text{ for } x_{[r]} \leq x_t < x_{[r+1]}, \; r = 1, 2, ..., T \quad (1)$$

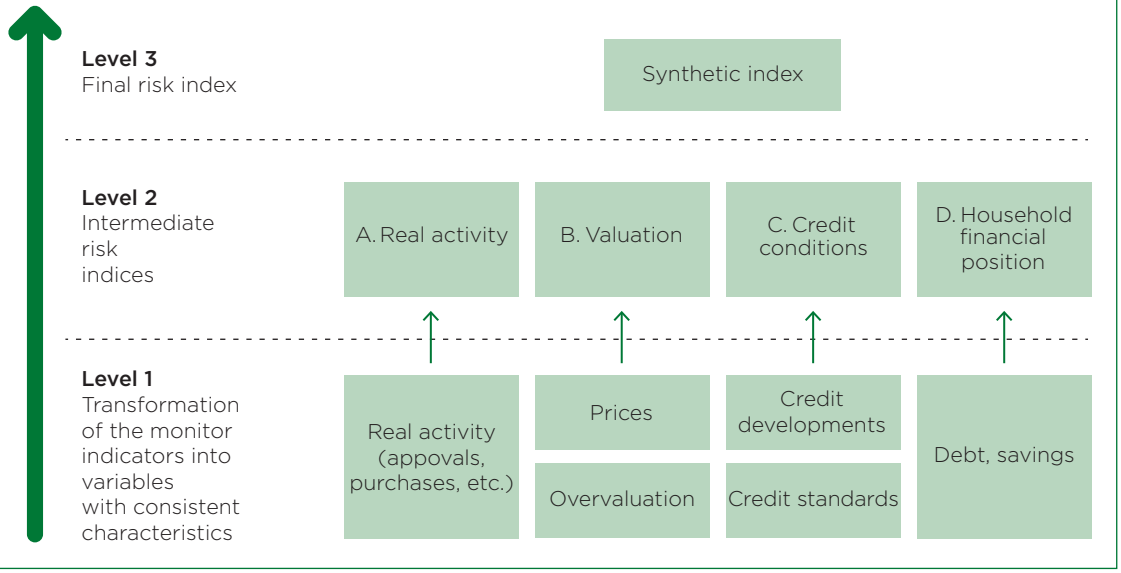
where r indicates the position assigned to each value of the variable[9]. The new variables range from 1/T to 1, which represent the minimum and maximum values, respectively, of the original indicator's distribution. In other words, a value near zero would indicate that this figure is close to the variable's minimum value, while a value close to one indicates that it is close to the maximum value. To construct synthetic index and intermediate indices, the transformed individual variables, z_t, were calculated so that higher values indicate higher risk and vice versa[10]. In addition, to achieve a more stable synthetic index, from 2013 Q4 onwards all transformations were computed recursively, and the transformed indicators are recalculated whenever new observations are included, whereas before that date, the same sample is always considered for calculating the transformed indicators (z_t).

By construction, the distance between two consecutive observations in the transformed indicators is always the same (1/T), so statistics such as the mean or the variance are comparable between these indicators. In addition, with this methodology it is easier to

9 Repeated values are assigned to the average of their position in the variable's order.

10 Certain indicators that signal risks when very low values are reached (left-tail indicators), therefore a prior operation is necessary. Such as indicators are modified as follows, Y = 1 - X, where X is the value of the variable. This transformation means that greater value of Y signal greater risk.

FIGURE 1. METHODOLOGY TO CONSTRUCT THE HOUSING MARKET SYNTHETIC INDEX



Source: Banco de España.

distinguish inflection points in the cycle. This is because changes in the transformed variables are more pronounced at the central points of the distribution (start of upswings/slowdowns) than at the tails (peaks/troughs of the cycle). However, the distance between consecutive observations of z_t being constant may result in some loss of information in the analysis of outliers. Therefore, z_t indicators should not be used to compare the severity of imbalances between the peaks in two upturns, or the extent of the downturns.

2.2. Construction of the intermediate indices for each risk category

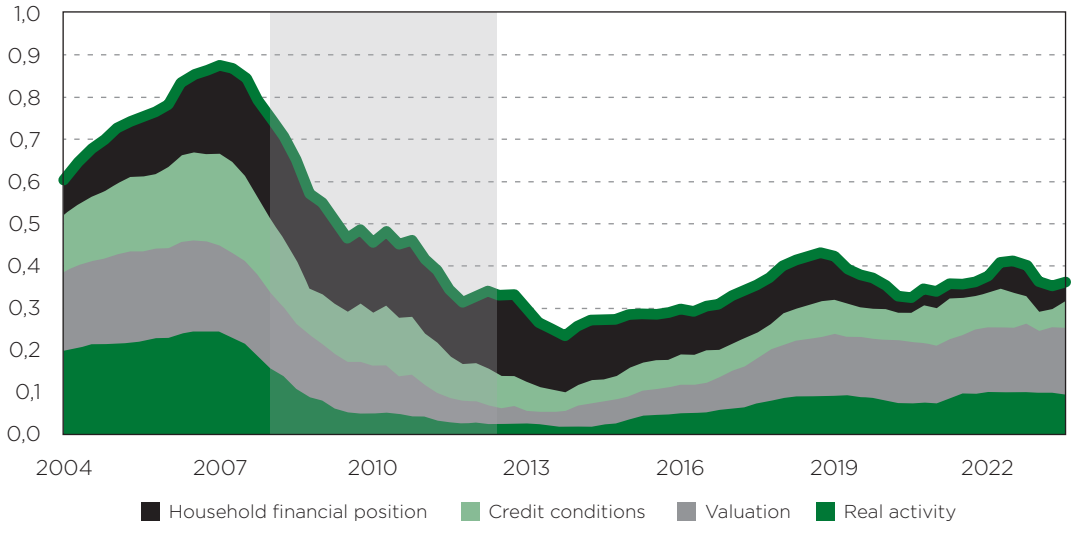
Once the individual indicators have been transformed, they are aggregated into intermediate indices for each of the four defined categories mentioned above. These intermediate indices have the same statistical properties as the z_t indicators defined previously. The method used to aggregate the transformed individual indicators is based on a principal component (PC) analysis. This technique makes it possible to find linear combinations of the z_t indices with which to obtain one or a few PCs that explain most of the variability in these indicators, and whose information does not overlap.

This PC-based statistical approach to aggregating individual indicators is particularly useful in those cases where no objective economic criteria are available to determine the contribution of the individual indicators to risk severity. Another advantage of the PC-based method is that it strips out redundant information from highly correlated variables.

This methodology is applied to all cases, except for credit standards, which fall under the credit conditions category, for which there is a model that approximates the probability of default based on credit standards[11]. The probability of default is subsequently transformed into a z_t indicator using its ECDF. The intermediate index of the credit conditions category is calculated as the simple average between this indicator and the indicator

11 A logit model is used. For a description of this model, see Galán and Lamas (2023).

FIGURE 2. SYNTHETIC INDEX OF SPANISH HOUSING MARKET



Note: The grey shaded area shows the period of the real estate crisis (2009 Q1-2013 Q4).
Source: Banco de España. Latest observation: 2023 Q3.

summarizing the credit in the housing market, obtained via the PC method.

The PC-based methodology used to aggregate the individual indicators is described below. First, for each risk category the PCs are obtained as follows:

$$PC_1 = \Sigma\,(a_{[t,\,1]} \times z_t),$$
$$PC_2 = \Sigma\,(a_{[t,\,2]} \times z_t),$$
$$\ldots$$
$$PC_N = \Sigma\,(a_{[t,\,N]} \times z_t),$$

where a indicates the weights corresponding to the transformed indicators z_t in each PC, and N the number of indicators z_t in each category[12].

Given the strong correlation between the indicators within each category, one or two PCs are enough to explain most of the variability in the individual indicators (more than 80 % of the variance). When in one category a single PC contains enough information on the indicators (i.e. when this PC explains more than 80 % of the variance), the weights are obtained directly from that PC (by construction, the sum of the squares of *a* is equal to one, so the weights are calculated immediately). By contrast, when more than one PC needs to be considered to ensure that the explained variance exceeds 80 %, the intermediate index is obtained by weighting each PC by its contribution to the variance of the original indicators in each category[13].

Figure 2 presents the results of synthetic index since 2003 up to 2023 Q3. ●

12 The weights satisfy several conditions: 1) they are orthogonal; 2) the PCs are ordered so that the first PC explains the highest proportion of the Z t indicator variance, while the last indicator explains the lowest percentage of the variance; and 3) a t1 2 + a t2 2 + ... + a tN 2 = 1, where t = 1, 2, ..., N.

13 To summarise information from several PCs, the same methodology as the one proposed in OECD (2008) is used.

References

- Alves, Pana, Carmen Broto, María Gil and Matías Lamas. (2023). "Risk and vulnerability indicators for the Spanish housing market". *Occasional Papers*, 2314, Banco de España. https://doi.org/10.53479/33154
- European Systemic Risk Board (ESRB). (2013). "Recommendation of the European Systemic Risk Board of 4 April 2013 on intermediate objectives and instruments of macro-prudential policy". *Official Journal of the European Union*, 15 June. https://www.esrb.europa.eu/pub/pdf/recommendations/2013/ESRB_2013_1.en.pdf
- European Systemic Risk Board (ESRB). (2019a). *Vulnerabilities in the residential real estate sectors of the EEA countries*. https://www.esrb.europa.eu/pub/pdf/reports/esrb.report190923_vulnerabilities_eea_countries~a4864b42bf.en.pdf
- European Systemic Risk Board (ESRB). (2019b). *Methodologies for the assessment of real estate vulnerabilities and macroprudential policies: residential real estate*. https://www.esrb.europa.eu/pub/pdf/reports/esrb.report190923_methodologies_assessment_vulnerabilities_macroprudential_policies~7826295681.en.pdf
- Galán, Jorge E., and Matías Lamas. (2023). "Beyond the LTV ratio: lending standards, regulatory arbitrage and mortgage default". *Journal of Money, Credit and Banking*, February. https://doi.org/10.1111/jmcb.13041
- Organisation for Economic Co-operation and Development (OECD). (2008). *Handbook on Constructing Composite Indicators*. https://www.oecd.org/sdd/42495745.pdf