Quantitative Finance and Economics

http://www.aimspress.com/journal/QFE

QFE, 2(4): 904–930.
DOI:10.3934/QFE.2018.4.904
Received: 22 April 2018
Accepted: 27 July 2018
Published: 26 November 2018

Research article

Out-of-sample forecasting of housing bubble tipping points

Diego Ardila[1],***Dorsa Sanadgol**[1] **and Didier Sornette**[1,2]

[1] ETH Zurich, Department of Management, Technology, and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland

[2] Swiss Finance Institute, Switzerland

* **Correspondence:** Email: dardila@ethz.ch; Tel: +0446328917.

Abstract: This paper analyzes the information content of statistical tests for bubble detection in the context of international real estate markets. We derive binary indicators from the causal application of five statistical tests to log house prices, and via logit regressions we assess the indicators' out-of-sample performance in the forecasting of tipping points of housing bubbles and systemic financial crises. In our assessment, three of the indicators - two based on the identification of super-exponential trends and one based on the scaled ratio of the sum of squared forecast errors - exhibit significant out-of-sample results. Combining the indicators via simple threshold-rules yields the most robust and best results.

Keywords: housing bubbles; out-of-sample forecasting; house price forecasting; monitoring forecasting; OECD countries
JEL codes: C40, C52, C53, E31, E37, G01

1. Introduction

With the goal of developing early warning signals of future crises, many economic indicators have been analyzed in the aftermath of the 2007–2009 financial crisis. This literature has its beginnings in the prediction of stock price corrections, and by now includes measures such as credit-to-GDP growth, the debt service ratio, and gaps of property prices (Gençay and Gradojevic, 2010; Drehmann and Juselius, 2014). Attempts to estimate large scale deviations from equilibrium relations in housing markets often rely on ad-hoc rules that follow the application of the Hodrick-Prescott filter (Jordà et al., 2015). In contrast, the use of formal statistical tests, based on precise definitions of what a large deviation (i.e. a bubble) is, has been very limited. In addition, the studies that actually apply these techniques tend to presuppose rather than to examine the validity of the tests (Pavlidis et al., 2013). Thus, the question of whether statistical tests for bubble detection can be used to practically identify bubbles in housing markets remains largely unexplored*.

The assessment of these tests is relevant not only within the context of early warning systems, but also for any attempt to empirically study the causes and consequences of bubble regimes (Sarno and Taylor, 2003). An inappropriate definition of bubbles or a flawed bubble identification strategy might rule out, even in hindsight, valid and alternative interpretations of the data. If crashes are for example used to identify bubbles, bubble regimes are likely to be associated with poor economic performance, as only episodes with crashes, and potentially very evident ones, will enter the sample of the study (a typical case of selection bias). On the contrary, bubble identification strategies that do not assume a crash or economic decline as consequences of a bubble, have already yielded empirical evidence that asset price bubbles can be a welfare enhancing phenomenon (Narayan et al., 2016; Shi, 2017). Let us also mention the "social bubble hypothesis" (Gisler and Sornette, 2009, 2010; Gisler et al., 2011), which describes the existence of beneficial bubbles fostering innovations, during which strong social interactions between enthusiastic supporters weave a network of reinforcing feedbacks that lead to widespread endorsement and extraordinary commitment by those involved, beyond what would be rationalized by a standard cost-benefit analysis.

The present paper analyzes the information content of statistical tests for bubble detection in the context of international real estate markets. We evaluate the ex-ante performance of five different indicators derived from three types of statistical tests for bubble detection: 1) two tests based on the identification of super-exponential trends; 2) one test based on the detection of mildly explosive regimes; 3) two tests based on the scaled ratio of forecasting errors. We mainly compare the indicators individually, though towards the end of the paper we also analyze simple rules to combine them.

Following the literature on early warning signals (Drehmann and Juselius, 2014; Edison, 2003), we study the performance of the bubble indicators in the forecasting of tipping points of real estate bubbles and systemic financial crises. We run logit regressions having the indicators as dependent variables, and evaluate the performance of the regressions via the statistical significance of the area of the receiver operating characteristic curve (AUROC statistic). Our dependent variable consists of a binary indicator for the dates of the bubbles' tipping points and systemic banking crises. The dates of the bubbles' tipping points were obtained from a literature review on drivers of house prices. In total, we reviewed 19 single-country, 3 bi-country, and 3 multi-country studies to identify the periods of extreme overvaluation. We argue that our dating approach is more objective than any other technical rule previously employed in the literature, as such rules do not take into account the particularities of each housing market and depend on strong (and often symmetric) corrections to identify the end of a bubble period. The dates of systemic banking crises are those reported by Laeven and Valencia (2008), Valencia and Laeven (2012) and Drehmann et al. (2012) for the countries in our sample.

We emphasize that, although we employ the methodologies of the literature on early warning signals, our aim, more generally, is to contribute to the understanding of the tools available to identify and analyze bubble regimes. Thus, we are not working on an early warning signal system because this would require endogenizing the cost of false positives as well as false negatives, which necessarily depends on the applications and circumstances.

Overall, our results suggest that the statistical tests for bubble detection contain significant in-sample and out-of-sample forecasting power. All tests present significant in-sample performance,

*Although the Federal Reserve Bank of Dallas regularly applies the Phillips test to international markets, there are no formal assessments regarding the performance of this procedure in forecasting the tipping point of past bubbles (Mack et al., 2011).

whereas three out of the five indicators present significant out-of-sample forecasting power. We contribute to the literature on early warning indicators by documenting and comparing the forecasting skill of these statistical methods for bubble detection. Our results also emphasize the importance of identifying and understanding the consequences of periods of explosive price development, as they directly link statistical exuberance to historic episodes of strong deviations from fundamental factors. In addition, the fact that the supLPPLS and supPL contain significant value relative to the other approaches highlights the need to further employ diagnosis that are sensitive to the super-exponential growth pattern, which is increasingly supported as being a key stylized fact of financial and housing bubbles (Hüsler et al., 2013; Leiss et al., 2015; Gjerstad and Smith, 2014). Finally, our results give insights on the use of price-to-rent and price-to-income ratios, which have been broadly employed to diagnose deviations from fundamental values (Bourassa et al., 2016). We find that, although the ratios indeed contain valuable information related to the bubble, the ex-ante power of the statistical tests, when applied to price-to-income and price-to-rent ratios, does not seem sufficient to ex-ante identify the bubble regime.

This paper is organized as follows: section 2 describes the sample set and the chronology of bubbles. Section 3 presents the bubble detection tests from which our indicators are derived. Section 4 explains the methodology to construct and assess the indicators. Section 5 discusses the results. Section 6 presents the conclusions.

2. Database and the chronology of bubbles

Our sample comprises 18 countries, with quarterly data between 1975Q1 and 2013Q3[†]. These economies were selected based on relevance and availability. We mostly rely on the house indices in (Mack et al., 2011), though we also use the BIS database (Scatigna et al., 2015) to obtain the real house prices from two countries from the Asia-pacific region: Thailand and South Korea. These two countries are important as Asian countries suffered a systemic financial crises between 1997 and 1998, which stopped the real estate booms taking place in these economies. Thus, their real estate cycles remained disconnected from the last housing boom of the US and Europe, and significantly enrich the information in our sample. All series are in real terms (deflated by CPI), and normalized by their respective value in 2005Q1 to ensure comparability of the units.

To compare our results against existing benchmarks used in the literature of early warning signals, we gathered the countries' (log) price-to-income and (log) price-to-rent time series from the OECD database[‡].

2.1. *Bubble tipping points*

In order to identify the bubble tipping points, we surveyed the literature on drivers of house prices. Empirical detection of bubbles poses several challenges as there is no general theoretical or empirical definition of a bubble (Case and Shiller, 2003). Furthermore, hindsight might reveal or obscure the presence of bubble episodes. A crash might be regarded as evidence in favor of a bubble. On the

[†]Excluding Thailand, for which the time series starts in 1990Q4.
[‡]Prices: Analytical house price indicators `http://www.oecd-ilibrary.org/economics/data/prices/prices-analytical-house-price-indicators-edition-2016-1_6f6a769e-en`. When using these ratios, we had to exclude Thailand, as this country is not part of the OECD.

contrary, absence of a strong correction following a protracted boom might lead scholars to incorrectly conclude that a period of exuberance was indeed justified by unrecognized fundamental factors.

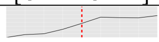
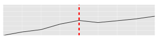
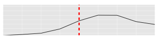
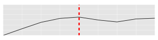
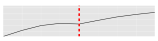
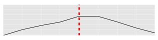
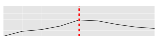
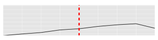
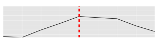
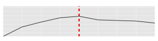
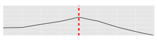
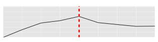
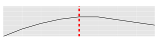
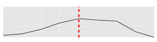
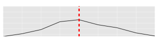
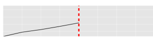
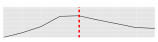
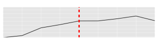
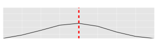
Hence, to conduct this study and contrary to the vast majority of indicators used in the literature of macroeconomic crises (see e.g. Jordà et al., 2015), we opt to identify bubbles by compiling the results from 25 different studies based on a fundamental analysis of prices (19 single-country, 3 bi-country, and 3 multi-country studies). These studies are typically based on inverted demand equations or life cycle consumptions models (Anundsen, 2015), and include variables such as interest rate, property tax, housing stock, disposable income, among others. We argue that our identification strategy provides us with a more objective benchmark against which to evaluate the performance of the indicators as fundamental-based studies take into consideration the internal macroeconomic conditions of the countries, especially those studies associated specifically to a single country.

Admittedly, our survey is far from perfect as bubbles remain a contentious concept. The identification strategy is based only on articles that interpret deviations from fundamentals as signs of bubbles. Their primary assumption - unless expectations are explicitly modeled - is that the lagged appreciation unexplained by the fundamental variables represents speculative pressures. However, it is fair to acknowledge that many studies identify such deviations without explicitly referring to speculative or exuberant periods. An alternative but not necessarily incompatible explanation is that real house prices have a tendency to overshoot following a shock to the market. As a result, observed house prices are likely to deviate from the long-run equilibrium level. Another issue typically raised is that the identification strategy might be subject to misspecification, as the econometrician has no way to know the "true" model (Flood and Hodrick, 1990).

The studies seldom report explicitly the peak of the bubble. Therefore, we chose the peak of the overvaluation period as the variable of analysis. In total, we identified 19 housing bubbles over the two long real estate cycles that the world has experienced plus the housing bubbles in Asia. Table 1 presents our complete list of tipping points matched to their original sources. The table includes the real log returns four years before and after the tipping point (Δ_{-4} and Δ_4 respectively), as well as the ratio of these two quantities ($\frac{\Delta_4}{\Delta_{-4}}$). The vertical lines in the subplots in the table represent the peaks of the bubbles.

A salient feature is the asymmetric behavior of booms and bursts, in which the former tend to be more pronounced than the latter. Only five countries, Ireland (2007), Japan, Sweden, Switzerland, and the US, exhibit symmetric boom-bust behaviors. Four bubbles, those in Canada, France, the Netherlands, and UK (2005), present almost constant price development four years after the peak when computed in real terms. House prices in Ireland (2002) quickly recovered after the bubble-burst episode. In the Netherlands and Denmark, prices continued rising after the peak of the overvaluation so that the peak does not coincide with a local maximum in the price level, illustrating the fact that adjustments in fundamental factors, such as a drop in the interest rate, might also align prices with fundamentals and prevent a visible correction from happening. The existence of such patterns challenges identification strategies based on local maxima. Bubbles do not always suddenly burst, the end phase of the bubble can follow very different dynamics, and as Brunnermeier and Oehmke (2012) have remarked, we do not know enough yet about how and why prices evolve the way they do once the expansion phase of the real estate cycle has passed.

Table 1. Summary of 19 bubble peaks in 15 countries according to studies surveyed. The table shows the years of the peaks, the real log returns four years before and after the peaks (Δ_{-4} and Δ_4 respectively), the real house price development around these years, and the ratio $\frac{\Delta_4}{\Delta_{-4}}$. The vertical lines in the small figures denote the bubble peaks. No peaks are reported for Germany, Italy, and Thailand as Nobili and Zollino (2012) and Glindro et al. (2011) concluded that prices in these countries have evolved (mostly) according to the fundamentals.

Country	Peak t_p	Δ_{-4}	Δ_4	$\frac{\Delta_4}{\Delta_{-4}}$	$\ln p_t$ $[t_p - 4, t_p + 4]$	Reference
Australia	2003	0.44	0.14	0.31		Glindro et al. (2011) Berry and Dalton (2004)
Canada	2008	0.31	0.1	0.29		Walks (2014)
Denmark	2006	0.33	-0.08	-0.25		Sørensen (2013) Dam et al. (2011)
France	2008	0.35	-0.02	-0.06		Antipa and Lecat (2010)
Ireland	2002	0.42	0.39	-0.93		Connor et al. (2012) Stevenson (2008)
	2007	0.39	-0.33	-0.87		Hott and Monnin (2008)
Japan	1991	0.3	-0.3	-0.97		Hott and Monnin (2008) Barsky (2009)
Netherlands	2006	0.08	0.01	0.075		Hott and Monnin (2008) Francke et al. (2009)
Norway	1989	0.43	-0.32	-0.75		Anundsen and Jansen (2011) Jacobsen and Naug (2005)
S. Africa	2008	0.47	-0.16	-0.34		Das et al. (2011)
S. Korea	1991	0.23	-0.41	-1.74		Glindro et al. (2011) Kim and Min (2011)
Spain	1992	0.42	-0.19	-0.46		Ayuso and Restoy (2006) Neal and García-Iglesias (2013)
	2007	0.42	-0.18	-0.42		Antipa and Lecat (2010)
Sweden	1990	0.29	-0.33	-1.13		Hort (1998) Andreas Claussen (2013) Sørensen (2013)
Switzerland	1990	0.30	-0.29	0.97		Hott and Monnin (2008)
	2013	0.19	-	-		Hott (2012)
UK	1990	0.5	-0.29	0.58		Muellbauer and Murphy (1997) Cameron et al. (2006)
	2005	0.43	0.01	0.14		Black et al. (2006) Hott and Monnin (2008)
US	2006	0.22	-0.22	-1		Hott and Monnin (2008) Mikhed and Zemčík (2009) Anundsen (2015)

2.2. *Crises*

The dates of systemic crises that we take are the years with a beginning of a systemic banking crises, reported by Laeven and Valencia (2008), Valencia and Laeven (2012) and Drehmann et al. (2012) for the countries in our sample. In case of discrepancies, we choose the earliest reported year among these two sources. Table 2 presents the actual values. Canada and South Africa do not appear in the table as there are no crises reported for these two countries.

Table 2. Systemic banking crises. The table reports the years with a beginning of a systemic banking crises. For the countries in our sample, we consolidated the dates reported by Laeven and Valencia (2008), Valencia and Laeven (2012) and Drehmann et al. (2012). In case of discrepancies, we chose the earliest reported year. Canada and South Africa do not appear with crisis events.

Country	Year crisis	Country	Year crisis
Australia	1989	Norway	1990
Denmark	2008	S. Korea	1997
Germany	2007	Spain	2008
France	2008	Sweden	1991, 2008
Ireland	2008	Switzerland	2008
Italy	2008	Thailand	1983, 1997
Japan	1997	UK	1990, 2007
Netherlands	2008	USA	1988, 2007

3. Bubble detection tests

In this section, we present several bubble detection tests proposed in the literature of asset bubbles. We are interested in identifying a bubble in a log prices time series $\ln p_t$ with $t = 1, ..., [\tau T], [\tau T] + 1, ..., T$, $\tau \in (0, 1)$ and $[\tau T]$ denoting the greatest integer smaller than or equal to τT. $[\tau T]$ corresponds to the starting period in which the bubble is detected.

The tests can be motivated within the theory of rational expectations bubbles (REB). The departure point of REB is the familiar no arbitrage condition for the price of an asset,

$$P_t = \frac{1}{1+R} E_t(P_{t+1} + D_{t+1}) \tag{1}$$

where P_t is the asset price at time t, D_t is the dividend received from the asset for ownership between $t-1$ and t (i.e. the inputed rent in a real estate context), $R > 0$ is the discount rate, and $E_t(\bullet)$ denotes the expectation conditional on the information at time t. A log linear approximation of equation 1 and forward iteration yields (Campbell and Shiller, 1988; Phillips et al., 2011),

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

where

$$p_t^f = \frac{\kappa - \gamma}{1-\rho} + (1-\rho) \sum_{t=1}^{\infty} \rho^i E_t(d_{t+1+i}) \tag{3}$$

$$b_t = \lim_{i\to\infty} \rho^i E_t(p_{t+i}) \tag{4}$$

$$E_t(b_{t+1}) = \frac{1}{\rho} b_t = (1 + \exp(\overline{d-p}))b_t \tag{5}$$

with $p_t = \ln P_t$, $d_t = \ln D_t$, $\gamma = \ln(1+R)$, $\rho = 1/(1+\exp(\overline{d-p}))$, $\overline{d-p}$ equals the average log dividend price ratio, and $\kappa = -\ln\rho - (1-\rho)\ln(\frac{1}{\rho} - 1)$. p_t^f is called the fundamental component of the price, and b_t the rational bubble component. In the absence of a bubble (i.e. $b_t = 0$), $p_t = p_t^f$ and p_t is solely determined by the dividends. Conversely, if $b_t \neq 0$, equation 5 implies the following process,

$$b_t = \frac{1}{\rho} b_{t-1} + \epsilon_{b,t} = (1+g)b_{t-1} + \epsilon_{b,t}\ , \qquad E_{t-1}(\epsilon_{b,t}) = 0 \tag{6}$$

where $g = \frac{1}{\rho} - 1 = exp(\overline{d-p})$ is the growth rate of the natural logarithm of the bubble and $\epsilon_{b,t}$ is a martingale difference. Since $g > 0$, equation 5 implies that b_t is a sub-martingale, and equation 6 states that bubbles, if they are present, should manifest explosive characteristic in log prices.

The rational expectations bubble theory has several drawbacks. Models based on REB provide conditions for an economic equilibrium with bubbles, but they do not explain the dynamics that lead to the boom. In other words, bubbles in these models do not emerge, but they exist or appear randomly and it is unclear when/under what circumstances a market moves from a stable to an unstable regime. Rational expectation bubbles, if they exist, must be positive, infinitely lived, and hence, require assets with infinite maturity (Blanchard and Watson, 1982). These properties are at odds with empirical evidence such as, for example, the bubble on Chinese call warrants discussed by Palan (2013). Giglio et al. (2014) exclude infinitely lived bubbles on the housing markets of Singapore and the U.K [§]. Scheinkman and Xiong (2004) argue that models of rational bubbles are incapable of explaining the increase in trading volume that is typically observed in the historic bubble periods. Lux and Sornette (2002) demonstrate that exogenous rational bubbles are not compatible with some of the stylized facts of financial data at a very elementary level.

While these limitations are serious, there exists versions of REB that attempt to circumvent them, such as those developed in (Johansen et al., 2000) and (Lin and Sornette, 2013) for instance. Sornette (2002) proposed a resolution of the objections raised by Lux and Sornette (2002). We thus use below the REB theory as a common framework to illustrate the main differences among the bubble tests. At this point, it is worth stressing that the analysis of bubbles remains incomplete, and their research requires pragmatic choices as well as a constant dialogue between any proposed theory and the supporting empirical studies. In this vein, let us mention the framework recently advanced in (Schatz and Sornette, 2018) for explosive semi-martingales, which generalises the REB theory and is based on the antagonistic combination of (i) an excessive, unstable pre-crash process and (ii) a drawdown starting at some random time.

3.1. Super-exponential based procedure

The first type of bubble detection test is based on the identification of a transient super-exponential trend in the dynamics of the log prices. This kind of dynamics was first proposed on the basis of empirical observations in (Sornette et al., 1996; Feigenbaum and Freund, 1996). It was later justified by

[§]According to the authors, these markets give the best chances of detecting a bubble in the data.

Johansen et al. (2000) and Johansen and Sornette (1999) within the framework of rational expectation model of bubbles. From a theoretical view point, the authors argue that the no arbitrage condition, together with a hierarchical self-reinforcing organization of the market, and the need of investors to be compensated for the risk of the crash, generates a power law finite-time singular price dynamics as the bubble approaches its end. As a result of the positive feedbacks, such super-exponential dynamics is unsustainable as it ends in a finite time singularity, which signals a change of regime (the end of the bubble).

Statistically, a bubble can thus be identified via a non-nested hypothesis test of model selection (Davidson and MacKinnon, 1981). If there is enough evidence to reject the null hypothesis of a non-explosive process in favor of the alternative super-exponential trend, the bubble hypothesis can be supported. Specifically, we compare a stationary $AR(1)$ process in the log returns against $\Delta \ln \hat{p}_{t\,sexp}$, the log returns predicted by a fitted super-exponential trend in the subsample between $[\tau T]$ and T:

$$\Delta \ln p_t = \rho \Delta \ln p_{t-1} + \alpha_{sexp} \Delta \ln \hat{p}_{t\,sexp} + \epsilon_t, \text{for } t = [\tau T], [\tau T] + 1, ..., T \tag{7}$$

where ϵ_t is a white noise process. The null hypothesis of no bubble after $[\tau T]$ period is rejected if the t-statistic $\hat{t}_{\alpha_{sexp}}$ for the estimate of α_{sexp} exceeds the corresponding critical value[¶]. When the starting date of the super-exponential trend is not known, the statistic takes the following form:

$$\text{supSEXP}(\tau_0) = \sup_{\tau \in [0, \tau_0]} \hat{t}_{\alpha_{sexp}} \tag{8}$$

where τ_0 determines the interval in which the super exponential trend is tested.

We employ two different super-exponential specifications to obtain $\Delta \ln \hat{p}_{t\,sexp}$. First, the fitted values given by the log periodic power law singularity (LPPLS) model:

$$\ln p_{t\,sexp} = A + (t_c - t)^m [B + C \cos(\omega \ln (t_c - t) - \phi)] \tag{9}$$

where $0 < m < 1$, $B < 0$, $3 < \omega < 15$, and $|C| (\omega^2 + m^2)^{1/2} \leq |B| m$ (Filimonov and Sornette, 2013). t_c corresponds to the non-random time of the termination of the bubble. As calibration of equation 9 on quarterly data can be difficult due to the low frequency of the volatility of house prices and the relatively large number of parameters (7 in total, 3 nonlinear, 4 linear after the reformulation in (Filimonov and Sornette, 2013)), we also explore a simplification of the LPPLS model that excludes the log periodic oscillations (hereafter PL model):

$$\ln p_{t\,sexp} = A + B(t_c - t)^m \tag{10}$$

where as in equation 9, $0 < m < 1$ and $B < 0$. These last two conditions ensure that the instantaneous expected return diverges at t_c. In practice, it does not of course, but the hypothesis is that the average price trajectory can be approximated over a time interval until close to its turning point by such a process with increasing returns.

[¶]As a remark, $\rho \geq 1$ would also suggest super-exponential behavior, but we chose not to test this alternative to focus on the super-exponential trend.

3.2. The Kim-Busetti-Taylor statistic

In the case where d_t is stationary (i.e. $I(0)$), as it is observed for the inputed rent of a real estate asset, absence of bubbles implies by equation 2 that Δp_t should also be stationary. Hence, rejection of the stationarity of Δp_t in favor of the non-stationarity hypothesis can be interpreted as evidence of bubbles. In this context, Homm and Breitung (2012) employed the tests of Kim (2000) and Busetti and Taylor (2004) for a bubble detection setup.

Kim (2000) and Busetti and Taylor (2004) proposed independently a statistic for testing the null hypothesis that a stationary time series switches to a process with higher persistence (e.g. from $I(0)$ to $I(1)$). Their focus is on the Gaussian unobserved components model,

$$y_t = \beta_t + \mu_t + \epsilon_t \quad \text{for } t = 1, 2, ..., T \tag{11}$$

$$\mu_t = \mu_{t-1} + 1(t > [\tau T]\,\eta_t) \quad \text{for } \tau \in (0, 1) \tag{12}$$

where $1(\bullet)$ is the indicator function, ϵ_t and η_t are mutually independent mean zero IID Gaussian processes with variance σ^2 and σ_n^2 respectively, and β_t is a deterministic component that we take as constant. Equations 11 and 12 describe a process that is stationary up to $[\tau T]$, but is $I(1)$ after the break if and only if $\sigma_n^2 > 0$. Consequently, a test for stationarity against the non-stationary hypothesis can be framed as testing the null hypothesis

$$H_0 : \sigma_n^2 = 0 \tag{13}$$

against the alternative,

$$H_1 : \sigma_n^2 > 0 \tag{14}$$

A test statistic that rejects (13) for large values is,

$$KBT_\tau = \frac{[(1-\tau)T]^{-2} \sum_{t=[\tau T]+1}^{T} (\sum_{i=[\tau T]+1}^{t} \hat{\epsilon}_{1,i})^2}{[\tau T]^{-2} \sum_{t=1}^{[\tau T]} (\sum_{i=1}^{t} \hat{\epsilon}_{0,i})^2} \tag{15}$$

where $\hat{\epsilon}_{0,i}$ are the OLS residuals from the regression of y_t on an intercept in $t = 1, ..., [\tau T]$, and $\hat{\epsilon}_{1,i}$ are the OLS residuals from the regression of y_t on an intercept in $t = [\tau T] + 1, ..., T$. Equation 15 is a Chow-type test and can be interpreted as the scaled ratio of the sum of squared forecast errors.

Although Homm and Breitung (2012) adapted the KBT_τ statistic to be applicable directly on log prices, we opt to use the statistic given by equation 15 to test stationarity of the returns in order not to rely on the forecast associated with the random walk assumption. Thus, under the alternative hypothesis, the return process switches at $[\tau T]$ from a stationary process to a process with persistent returns. If the returns are persistent, the corresponding time series exhibits explosive behavior. When the breakpoint $[\tau T]$ is unknown, the statistic can be expressed as

$$\text{supKBT}(\tau_0) = \sup_{\tau \in [\tau_0, 1-\tau_0]} KBT_\tau \tag{16}$$

and rejects the null hypothesis of no bubble for large values.

3.3. The Busetti-Taylor statistic

A second test also adapted by Homm and Breitung (2012) for a bubble setup is the Busetti-Taylor statistic. The Busetti-Taylor statistic takes again the two cases 13 and 14 as the null and the alternative hypotheses respectively. When τ is known, the statistic corresponds to the locally best invariant test against changes in the order of integration under the assumption of Gaussianity. Namely, it has maximum local power close to the null hypothesis under invariant transformations.

We deviate again from Homm and Breitung (2012) and proceed as with the KBT statistic by implementing directly the test as originally proposed in (Busetti and Taylor, 2004),

$$BT_\tau = \hat{\sigma}^{-2}(T - [\tau T])^{-2} \sum_{t=[\tau T]+1}^{T} (\sum_{j=t}^{T} \hat{\epsilon}_j)^2 \tag{17}$$

where $\hat{\epsilon}_j$ are the OLS residuals from the regression of $\Delta \ln p_t$ on an intercept, and $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_t^2$. When the breakpoint is not known, the Busetti-Taylor statistic takes the following form

$$\text{supBT}(\tau_0) = \sup_{\tau \in [0,1-\tau_0]} BT_\tau \tag{18}$$

and, as in the previous statistics, rejects the null hypothesis of no bubble for large values.

3.4. The Phillips statistic

Phillips and Yu (2011) pointed out that, since $g > 0$ in equation 6, if $b_t \neq 0$, equation 2 implies that p_t will also be explosive irrespective of whether d_t is an integrated process or a stationary process. As a direct way to test for bubbles, they thus proposed to examine evidence of explosive behavior in p_t via a time-varying AR model and an econometric procedure based on recursive Dickey-Fuller (DF) t-statistics. The time-varying AR model reads

$$p_t = \rho p_{t-1} + \epsilon_t \ , \tag{19}$$

where ϵ_t is a white noise process with $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, and $p_0 = c < \infty$. Under the null hypothesis, the process starts as a random walk (i.e. $\rho = 1$) but under the alternative, ρ becomes larger than 1 and the process changes to an explosive process at an unknown time $[\tau T]$. Thus, the procedure consists of testing directly for explosive behavior via right-tailed unit root test for certain sub-periods of the data. Let $\hat{\rho}_t$ denote the OLS estimator of ρ and $\sigma_{\rho,\tau}$ the usual estimator for the standard deviation of $\hat{\rho}_\tau$ using the subsample $\{p_1, ..., p_{[\tau T]}\}$. The forward recursive DF tests is given by,

$$\text{supDF}(\tau_0) = \sup_{\tau \in [\tau_0,1]} DF_\tau \tag{20}$$

with $DF_\tau = \frac{\hat{\rho}_\tau - 1}{\hat{\sigma}_{\rho,\tau}}$. The test rejects the null hypothesis of no bubble for large values of $\text{supDF}(\tau_0)$[‖].

[‖]Phillips and Yu (2011) also developed an asymptotic distribution theory for mildly explosive processes in order to date the emergence or collapse of the bubbles. As we focus on bubble detection, we refrain from presenting that part of their contributions.

4. Ex-ante identification and evaluation of bubbles

4.1. Ex-ante identification strategy

We apply the procedures described in section 3 to test for bubbles in the log of real house prices for every country in our sample. The supBT, supKBT, and supDF tests are applied on rolling windows with length of 60 quarters (15 years), correcting for serial correlation in the residuals, and setting $\tau_0 = 0.2$. These values for τ are not far from the 0.1 value typically used in other studies, and ensure that all of our estimations are based on at least 12 quarters (3 years). The super-exponential-trend tests employ a rolling window with a maximum length of 60 quarters, setting $\tau_0 = 0.3$. Both types of nonlinear trends (the LPPLS and the PL) are estimated by minimizing the sum of squared residuals using the procedure described in (Filimonov and Sornette, 2013). Given the relatively small sample size of our estimations, we use Monte Carlo simulations with $15,000$ replications to calculate the critical values of all tests. Results of these simulations are shown in table 3.

Table 3. Critical values for the different bubble tests. The reported values correspond to Monte Carlo simulations for samples of size 60 over 15′000 repetitions. The * indicates that more than 90% of the estimations did not fulfill the LPPLS constraints, so critical values do not apply for this level.

Level	90%	95%	97.5%	99%	99.9%
supLPPLS	*	2.19	3.2	4.27	6.67
supPL	1.66	2.18	2.59	3.078	4.47
supKBT	11.54	15.58	20.39	26.75	51.5
supBT	1.17	1.46	1.76	2.11	3.00
supDF	1.49	1.91	2.32	2.89	5.52

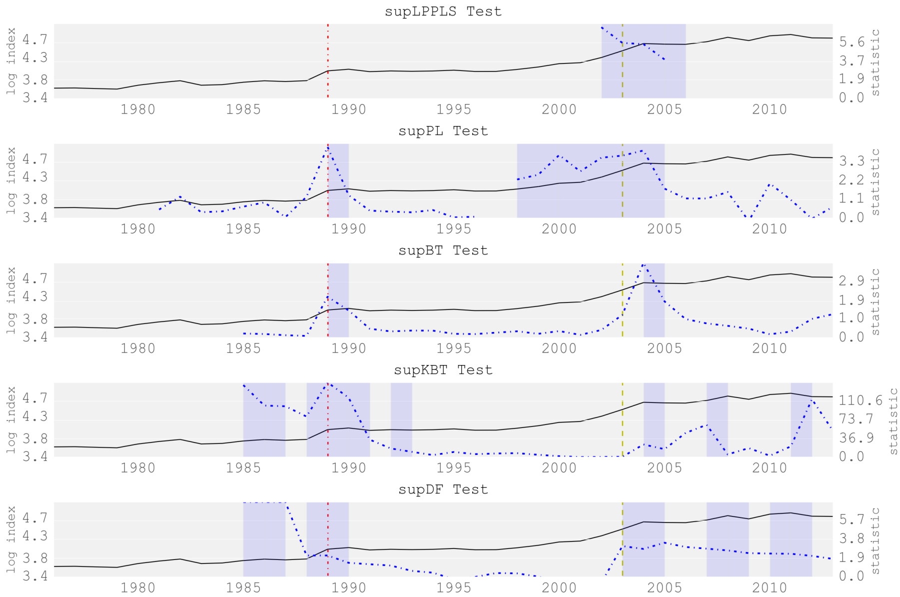
A bubble signal by test j is triggered at time t when the statistic is significant at a 5% α-level:

$$BI_t^j = \begin{cases} 1 & \text{p-value}_t^j < \alpha \text{ and } \Delta^y \ln p_t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

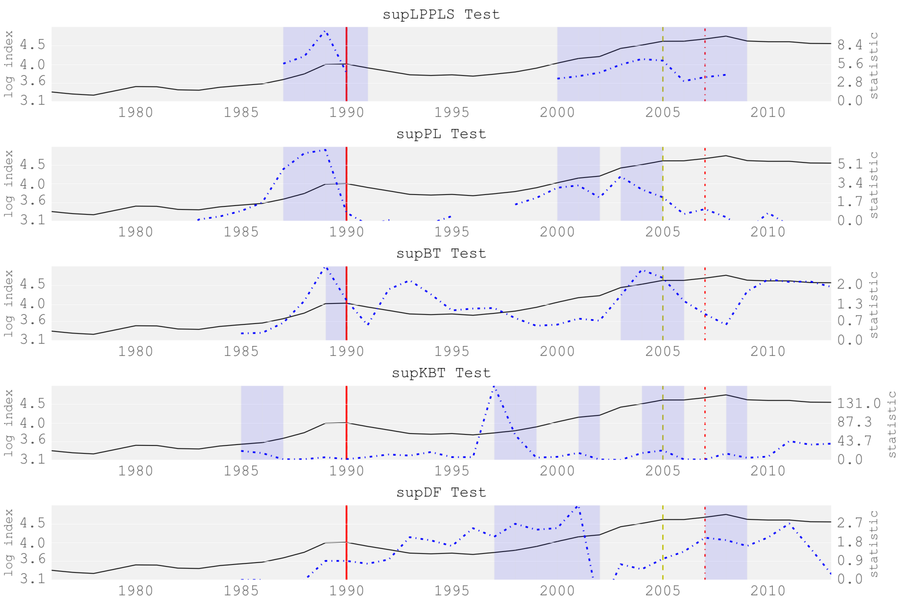
where $\Delta^y \ln p_t > 0$ denotes the yearly log returns and are set to be positive to ensure that alarms are triggered only while prices are rising.

Figure 1 presents the resulting bubble signals for the housing markets of Australia and the UK[**]. For Australia (panel 1b), timely bubble signals were triggered prior to 2004 in three cases (supDF, supPL, supLPPLS), while alarms not associated with bubbles were also detected around 1990 in all but one of the indicators (supLPPLS). For UK (panel 1b), three out of the five indicators (supBT, supPL, supLPPLS) timely identify the bubble prior to the 1990 tipping point while four detected a bubble prior to the 2005 tipping point (supDF being the exception); false alarms were triggered during the second half of the 90s in two of the indicators (supDF and supKBT).

[**]Signals for the other countries are available in the supplementary material of the paper.



(a) Australia



(b) UK

Figure 1. Bubble indicators for Australia and the UK. The indicators, active in the shaded areas, are triggered when the corresponding statistic exceeds the 95% critical values. The solid line depicts (left y-axis) the country's real log house index on which the tests are applied. The dash lines (right y-axis) correspond to the calculated statistics. The vertical dashed yellow and red lines denote the tipping dates of the bubbles and financial crises respectively, according to tables 1 and 2. The solid red line corresponds to a year with a crisis and a bubble peak.

4.2. Evaluation methodology

Following Schularick and Taylor (2012), we analyze the relationship between the bubble indicators and our dependent variable using yearly logit regressions. We use the following specification:

$$y_{i,t}^{bubble} = \Lambda(\sum_{k=0}^{MAXLAG} \beta_k BI_{i,t-k-1}^{j}) \tag{22}$$

where i, k, j denote indices for the country, the time lag, and the bubble indicator respectively, $\Lambda(\dot{)}$ is the logit function, $MAXLAG$[††] is set to 3, and $y_{i,t}^{bubble} = 1$ if there was a tipping point at time t in country i (and 0 otherwise). The quality of the models is assessed via the significance of the Area under the Receiver Operating Characteristics Curve (AUROC). The ROC curve characterizes the quality of a forecast system by describing the system's ability to anticipate correctly the occurrence or non-occurrence of predefined events (Mason and Graham, 2002). The AUROC is increasing with the predictive power of the indicator and lies between 0 and 1.

To obtain the ROC for a probabilistic forecast system, the probability at which a positive warning is issued is varied across a range of thresholds. For each threshold, the correspondence between the forecast and the observation is determined. This correspondence is described by a two component vector defined by the rate of true and false positives. Each pair of rates gives a coordinate in the ROC space that defines the ROC curve. It can be shown that when the forecast system has skill, the area under the curve will exceed 0.5 (0.5 being the case of an uninformative indicator). Furthermore, the significance of the ROC area can be objectively assessed via a rescaled Mann-Whitney U statistic, which tests the significance of forecast event probabilities for cases where events actually occur. For large samples, the significance of these measures can be assessed using a normal-distribution approximation:

$$U = AUROC \cdot \frac{e'e}{2} \sim N(\mu = \frac{e'e}{2}, \sigma^2 = \frac{e'e(n+1)}{12}) \tag{23}$$

where n is the total number of forecasts, e the total number of events (i.e. bubble tipping points), and $e' = n - e$ the total number of non-events. The null hypothesis of no forecasting skill is rejected for large values of U. As discussed by Mason and Graham (2002), the errors of these approximations tend to be small for relatively large samples (a large number of forecasts).

We study the AUROC obtained in-sample for the different bubble indicators. In addition, we study two types of out-of-sample performance: by country and by period. Out-of-sample by country calibrates the model multiple times, each time leaving one country out of the calibration in order to use it to assess the models' performance. Out-of-sample by period dynamically calibrates the models using data up to the t_n period and tests the performance in year t_{n+1} for $t_n = 1998...2012$.

5. Empirical results

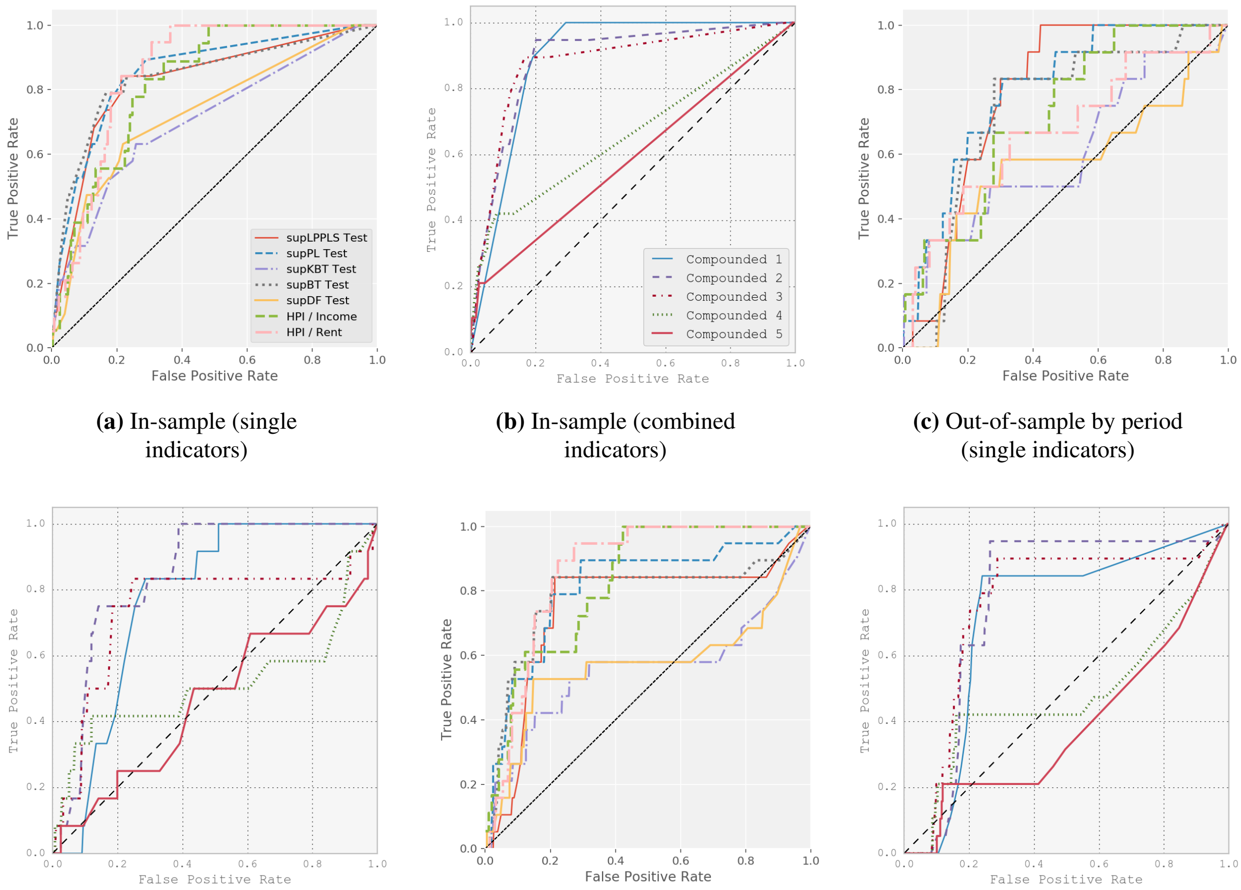
5.1. In-sample and out-of-sample performance

Figure 2 summarizes the main results, also reported numerically in table 4. AUROCs are expressed in percentage terms, while the respective U-statistic appears in parentheses. As a reference, we have

[††]We explored with values between 2 and 5 and the results did not qualitatively differ.

also included the performance of logit regressions, having the corresponding country price-to-income and price-to-rent ratios as dependent variables.

The in-sample AUROC of the different indicators is large with high statistical significance. The first row of table 4 shows that all AUROCs are above 0.7, and significant at the 99% level. The supPL test exhibits the highest performance, followed by the supBT and supLPPLS tests. The performance of the other two indicators, supKBT and supDF, is somewhat lower but still significantly positive. In-sample results are thus encouraging and suggest that the indicators contain relevant information associated to the end of the bubble. In order to test whether they translate to out-of-sample performance, we move now to the out-of-sample analysis. In 3 out of 5 cases, the values are comparable to those obtained when using the price-to-income and price-to-rent ratios, which exhibit high AUROC values, in line with the positive results for these quantities reported in the literature (Bourassa et al., 2016).



(a) In-sample (single indicators)

(b) In-sample (combined indicators)

(c) Out-of-sample by period (single indicators)

(d) Out-of-sample by period (combined indicators)

(e) Out-of-sample by country (single indicators)

(f) Out-of-sample by country (combined indicators)

Figure 2. Receiver operating characteristic (ROC) curves derived from bubble logit regressions (22). The independent variables are the bubble indicators derived from log prices (see equations (21) and (24)). The dependent variable $y_{i,t}^{bubble}$ is set to 1 if a bubble peaked at time t in country i.

Table 4. In-sample and out-of-sample AUROCs and AUROC statistics. Each column presents the AUROC (in percentage terms) and the AUROC statistic (in parentheses) for the denoted indicator. Out-of-sample leaving one country out corresponds to excluding once each country from the dataset, estimating the model, and evaluating the model's performance for the country out of the sample. Out-of-sample by period calibrates the model up to year t and evaluates the model's performance at year $t+1$ for t between 1998 and 2012. Significance levels are marked with *, **, and *** for 95%, 99%, and 99.9% respectively.

	Single bubble indicators (based on log prices)						
Test:	supLPPLS	supPL	supKBT	supBT	supDF	HPI / Income	HPI / Rent
In-sample	82.380***	85.994***	70.290**	84.529***	72.831***	82.45***	85.95***
	(4.811)	(5.348)	(3.015)	(5.131)	(3.392)	(4.682)	(5.331)
Out-of-sample by period	78.185***	79.423***	62.908	74.306**	53.306	72.522**	67.117*
	(3.305)	(3.450)	(1.514)	(2.850)	(0.388)	(2.637)	(2.004)
Out-of-sample by country	69.561**	74.428***	52.727	75.192***	52.386	82.72***	86.429***
	(2.910)	(3.634)	(0.406)	(3.748)	(0.355)	(4.722)	(5.401)
	Single bubble indicators (based on log price-to-income)						
In-sample	74.306***	78.536***	60.353	73.776***	53.481		
	(3.608)	(4.236)	(1.537)	(3.529)	(0.517)		
Out-of-sample by period	64.231*	64.615*	65.641*	69.455*	58.109		
	(1.667)	(1.712)	(1.832)	(2.278)	(0.950)		
Out-of-sample by country	58.868	64.203*	49.123	60.988	33.707		
	(1.318)	(2.111)	(−0.130)	(1.633)	(−2.422)		
	Single bubble indicators (based on log price-to-rent)						
In-sample	82.867***	79.208***	60.138	74.210***	59.265		
	(4.879)	(4.336)	(1.505)	(3.594)	(1.375)		
Out-of-sample by period	77.484***	63.734	36.875	54.936	37.019		
	(3.219)	(1.608)	(−1.537)	(0.578)	(−1.520)		
Out-of-sample by country	63.641*	63.028*	36.057	61.093*	35.373		
	(2.028)	(1.937)	(−2.073)	(1.649)	(−2.174)		
	Combined bubble indicators (based on log prices)						
Threshold:	1	2	3	4	5		
In-sample	89.177***	88.763***	87.747***	66.095**	58.804		
	(5.821)	(5.760)	(5.609)	(2.392)	(1.308)		
Out-of-sample by period	77.249***	84.617***	74.245**	51.495	46.633		
	(3.196)	(4.059)	(2.843)	(0.175)	(−0.395)		
Out-of-sample by country	71.967***	76.822***	75.520***	47.899	38.900		
	(3.268)	(3.991)	(3.797)	(−0.313)	(−1.651)		

Panels 2c and 2e show the results for the out-of-sample by period and by country tests respectively. Not surprisingly, the performance of all regressions drops visibly, from an average of 0.79 to an average of 0.69 (out-of-sample by period) and 0.65 (one-country-out). Nevertheless, three of the indicators contain significant predictive power in the out-of-sample by period test, exhibiting AUROC close to or above 0.75. The supLP test achieves the highest predictive power, followed by the supLPPLS (2nd) and the supBT tests (3rd). The performance of these cases is even higher that those of the price-to-income and price-to-rent ratios, respectively at 0.72 and 0.67, significant at the 99% level.

As for the setup consisting of leaving one country out, the results are somewhat less favorable. On one hand, The AUROCs based on the supBT, the supPL, the supLPPLS tests are still significant, but the significance of the latter dropped, while supBT, showing the best performance, is still significant at the 99.9% level. On the other hand, the AUROCs of the indicators based on the supKBT and the supDF tests have decreased more (16.1% and 1.7% respectively). The performance of the indicators have thus decreased, evidencing the indicators' difficulties in transferring the learnings from a set of countries to an unseen market. This contrasts with the price-to-income and price-to-rent ratios, which reached AUROC values above the 0.8 level, perhaps because they are better able to exploit the correlation between OECD house prices observed during the last global real estate cycle.

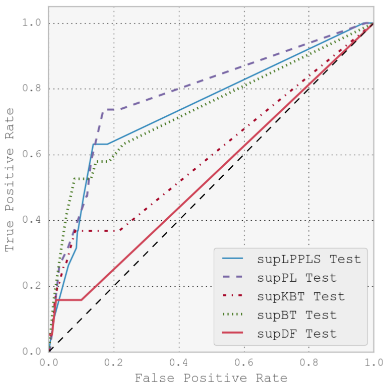
The differences between the results of the tests might be explained in several ways. The fact that the supBT statistic dominated the supKBT statistic, despite the fact that they assume the same data generating process, highlights the impact that a test's statistical power has on the bubble detection. The observed lacking forecasting skill of the supDF statistic, especially compared to those reported by Homm and Breitung (2012), might stem from our use of sliding windows of a maximum of 60 quarters to ensure that any alarm corresponds to a current bubble episode[‡‡]. The satisfactory results of the supLPPLS and the supPL tests might be revealing key features of bubbles, since they presuppose bubbles as fundamentally different market regimes, driven by slow-maturation and transient non-linear dynamics with a necessary end associated with the finite-time singularity. Other readings are possible, but we decided not to inquire further, as surely our results fall short of disentangling the origin of the variations in the results of the tests. Admittedly, the bubble analysis that we have conducted face a joint-hypothesis problem, as the differences among the tests might arise either because of the theory on which the statistics are based or because of their statistical power.

Nevertheless, beyond the specific results of each of the indicators, we argue that the emerging picture is very positive. The bubble tests have forecasting skill associated to the end of the bubble, and they can complement the ex-ante identification based on the HP filter and similar approaches, often used in the literature. As the bubble tests rely on very few parameters inferred from the past, we believe they are likely to be more robust, and therefore more appropriate, to ex-ante study future housing boom episodes. Moreover, the fact that the supLPPLS and supPL contain significant value relative to the other approaches highlights the need to further employ diagnosis that are sensitive to the super-exponential growth pattern, which is increasingly supported as being a key stylized facts of financial and housing bubbles (husler et al., 2013; Leiss et al., 2015; Gjerstad and Smith, 2014).

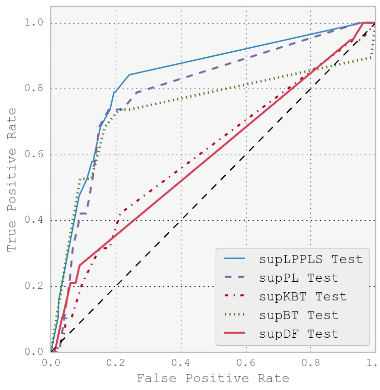
5.2. Assessment based on Price-to-Income and Price-to-Rent ratios

A common question regarding the practical application of the statistical tests that we have analyzed is whether they should be applied directly on prices, or on some measure intended to reflect the misalignment of prices with the fundamental value, such as price-to-income and price-to-rent ratios. In this section, we seek to shed some light on this issue by analyzing the performance of the analogous bubbles indicators based on these two ratios. Figure 3 summarizes the main results, also reported numerically in panels 2 and 3 of table 4.

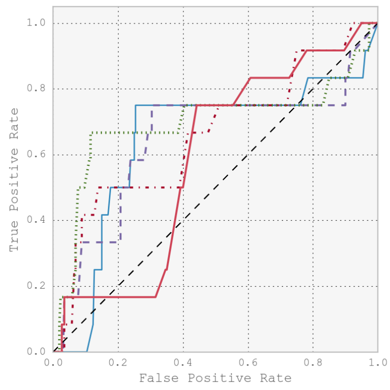
[‡‡]Homm and Breitung (2012) employed the whole sample to date the bubbles, which is difficult in an ex-ante setup. Although they also proposed the CUSUM and FLUC tests to monitor time series, these statistics require a bubble-free period as a training sample, which would introduce further complications and make difficult a comparison between the tests.



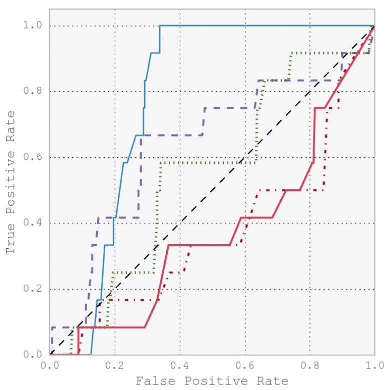
(a) In-sample (single indicators, log price to income ratio)



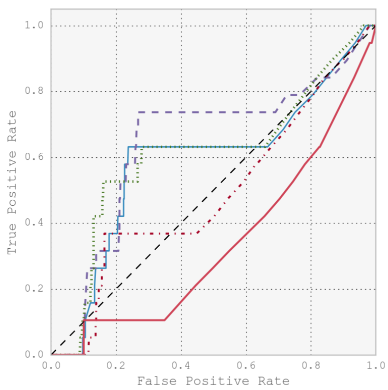
(b) In-sample (single indicators, log price-to-rent ratio)



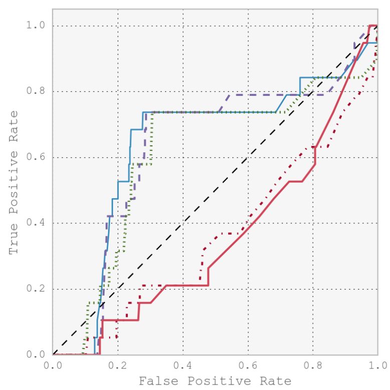
(c) Out-of-sample by period (single indicators, log price-to-income ratio)



(d) Out-of-sample by period (single indicators, log price-to-rent ratio)



(e) Out-of-sample by country (single indicators, log price-to-income ratio)



(f) Out-of-sample by country (single indicators, log price-to-rent ratio)

Figure 3. Receiver operating characteristic (ROC) curves derived from bubble logit regressions (22). The independent variables are the bubble indicators derived from log price-to-income and price-to-rent ratios (see equations (21) and (24)). The dependent variable $y_{i,t}^{bubble}$ is set to 1 if a bubble peaked at time t in country i.

The performance of the ratio-based indicators drops substantially relative to that of the price-based indicators. In-sample, average AUROCs values decrease from 0.78 to 0.68 and 0.71 for the price-to-income and price-to-rent ratios respectively. Out-of-sample by period, the corresponding average values fall from 0.69 to 0.64 and 0.54. Out-of-sample by country, the average AUROCs values plunge to the relatively low levels of 0.53 and 0.51 for the price-to-income and price-to-rent ratios respectively.

The significance of the AUROC statistics also deteriorates. For both ratios, two of the indicators (supKBT and supDF) are not in-sample significant anymore; in contrast to the highly in-sample significance levels observed by the use of price levels. Only the supLPPLS-based indicator yields a significant out-of-sample by period AUROC value when derived from the price-to-rent ratios, while the significance level of all the indicators decreases in the equivalent quantities derived from the price-to-income ratios. Similarly, in the out-of-sample by country set-up, there is only one indicator derived from price-to-income ratios with significant results (supPL), and the significance level of the three significant indicators derived from price-to-rent ratios is lower than the equivalent quantities derived from prices. As we described in the last section, these patterns contrast sharply with the performances of the regressions based directly on either of the indicators, which are relatively high and significant.

Lastly, in figure 3 we observe that the maximum attainable rate of true positive decreases out-of-sample for most of the indicators, which differs from the perfect true positive rate generally attainable by the indicators derived from price levels. In the out-of-sample by country set-up, none of the indicators reaches a perfect rate of true positives, and only the supLPPLS indicator achieves the maximum rate in the out-of-sample by period set-up, when derived from the log price to rent ratio. In other words, the bubble tests when applied on ratios, have not exhibited sufficient power to ex-ante reject the null hypothesis of no bubble. As a consequence, there is no path for the corresponding indicators to timely trigger the bubble signal, for any admissible false positive rate.

In our view, the lower performance yielded by the application of the bubble tests on these two ratios points to the endogenous nature of housing bubbles, which often co-occur as part of a more broader macroeconomic phenomenon. To the extent that rent, disposable income, and possibly other macroeconomic factors might also be endogenous to the bubble regime - as it is arguably the case of housing and credit booms coinciding with economic booms -, ratio-based tests might lack the statistical power to detect the bubble episode. Prices might be considered in line with fundamentals, because the so-called fundamental factors might also be going through an unstable phase.

5.3. *Combination of the indicators*

We now explore whether there are gains in combining the different bubble indicators (*BI*). We create a composite indicator based on the number of simultaneous alarms at a given time period. Namely:

$$I_t^C(k) = \begin{cases} 1 & \sum_{j=1}^{5} BI_{j,t} \geq k \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where j denotes one of the five indicators and k corresponds to the necessary threshold to trigger the composite bubble signal. Panels 2b, 2f and 2d of figure 2, and the second part of table 4 present the results.

We make two observations. First, combining the single indicators seems to yield more robust performance. Excluding the results for $k = 4, 5$, which correspond anyhow to very high thresholds, the in-sample and out-of-sample AUROCs are significant and above 0.87 (in-sample) and 0.72 (out-of-sample). Second, these results are positive enough to improve over the single top performing indicator (supPL or suptBT depending on the setup). The in-sample performance of the composite indicator is notably higher than the single indicator and the out-of-sample results are up to 6.5 percentage points higher. These two observations together suggest that, in the presence of model uncertainty, combining the indicators (i.e. using multiple tests for bubbles in a more general set-up) is a sensible decision. The resulting gains might derive not only in greater performance, but also in increasing robustness. Our results are thus consistent with the rather large literature, both in economics and more generally in pattern recognition, on the improved performance of compound predictors (Kittler et al., 1998; Timmermann, 2006). The gains are remarkable in light of the simplicity of the rule that we have employed.

5.4. Bubbles and systemic financial crises

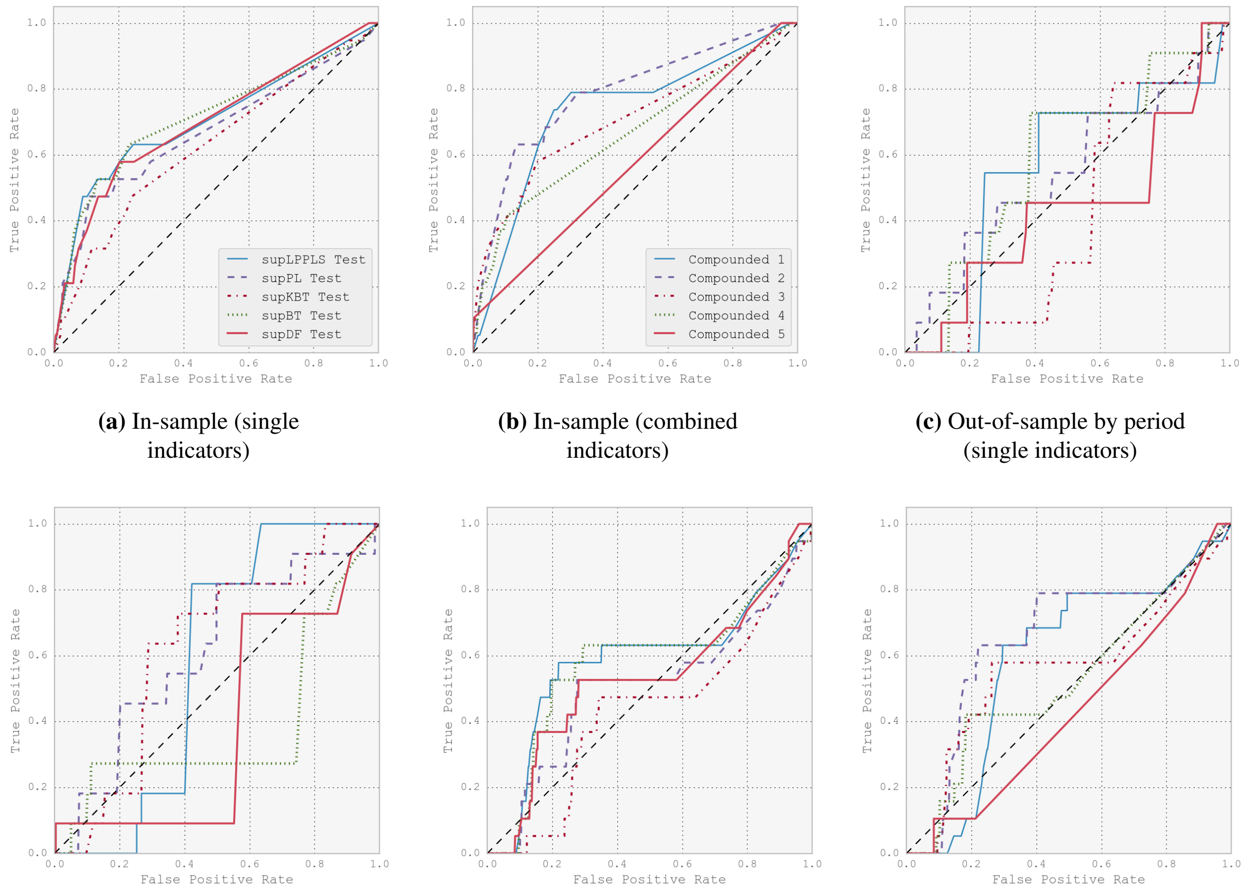
We have already presented supporting evidence that the bubble tests contain significant information to identify the end of the bubble period. We now contrast these results with those obtained by applying the same tests to forecast systemic financial crises. The goal is to further shed light on the phenomenon that the bubble tests are capturing. We use the regression

$$y_{i,t}^{crisis} = \Lambda\left(\sum_{k=0}^{MAXLAG} \beta_k BI_{i,t-k-1}^{j}\right) \tag{25}$$

where i, k, j denotes indices for the country, the time lag, and the bubble indicator respectively, $\Lambda(\cdot)$ is the logit function, $MAXLAG$ is set to 3, and $y_{i,t}^{crisis} = 1$ if a crisis started at time t in country i (and 0 otherwise). As before, we discuss the AUROCs and AUROC statistics to analyze the indicators' performance.

Table 5. In-sample and out-of-sample AUROCs and AUROC statistics. Each column presents the AUROC (in percentage terms) and the AUROC statistic (in parentheses) for the denoted indicator. Out-of-sample leaving one country out corresponds to excluding once each country from the dataset, estimating the model, and evaluating the model's performance for the country out of the sample. Out-of-sample by period calibrates the model up to year t and evaluates the model's performance at year $t + 1$ for t between 1998 and 2012. Significance levels are marked with *, **, and *** for 95%, 99%, and 99.9% respectively.

	Single bubble indicators (based on log prices)						
Test:	supLPPLS	supPL	supKBT	supBT	supDF	HPI / Income	HPI / Rent
In-sample	82.380***	85.994***	70.290**	84.529***	72.831***	82.45***	85.95***
	(4.811)	(5.348)	(3.015)	(5.131)	(3.392)	(4.682)	(5.331)
Out-of-sample by period	78.185***	79.423***	62.908	74.306**	53.306	72.522**	67.117*
	(3.305)	(3.450)	(1.514)	(2.850)	(0.388)	(2.637)	(2.004)
Out-of-sample by country	69.561**	74.428***	52.727	75.192***	52.386	82.72***	86.429***
	(2.910)	(3.634)	(0.406)	(3.748)	(0.355)	(4.722)	(5.401)
	Single bubble indicators (based on log price-to-income)						
In-sample	74.306***	78.536***	60.353	73.776***	53.481		
	(3.608)	(4.236)	(1.537)	(3.529)	(0.517)		
Out-of-sample by period	64.231*	64.615*	65.641*	69.455*	58.109		
	(1.667)	(1.712)	(1.832)	(2.278)	(0.950)		
Out-of-sample by country	58.868	64.203*	49.123	60.988	33.707		
	(1.318)	(2.111)	(−0.130)	(1.633)	(−2.422)		
	Single bubble indicators (based on log price-to-rent)						
In-sample	82.867***	79.208***	60.138	74.210***	59.265		
	(4.879)	(4.336)	(1.505)	(3.594)	(1.375)		
Out-of-sample by period	77.484***	63.734	36.875	54.936	37.019		
	(3.219)	(1.608)	(−1.537)	(0.578)	(−1.520)		
Out-of-sample by country	63.641*	63.028*	36.057	61.093*	35.373		
	(2.028)	(1.937)	(−2.073)	(1.649)	(−2.174)		
	Combined bubble indicators (based on log prices)						
Threshold:	1	2	3	4	5		
In-sample	89.177***	88.763***	87.747***	66.095**	58.804		
	(5.821)	(5.760)	(5.609)	(2.392)	(1.308)		
Out-of-sample by period	77.249***	84.617***	74.245**	51.495	46.633		
	(3.196)	(4.059)	(2.843)	(0.175)	(−0.395)		
Out-of-sample by country	71.967***	76.822***	75.520***	47.899	38.900		
	(3.268)	(3.991)	(3.797)	(−0.313)	(−1.651)		



(a) In-sample (single indicators)

(b) In-sample (combined indicators)

(c) Out-of-sample by period (single indicators)

(d) Out-of-sample by period (combined indicators)

(e) Out-of-sample by country (single indicators)

(f) Out-of-sample by country (combined indicators)

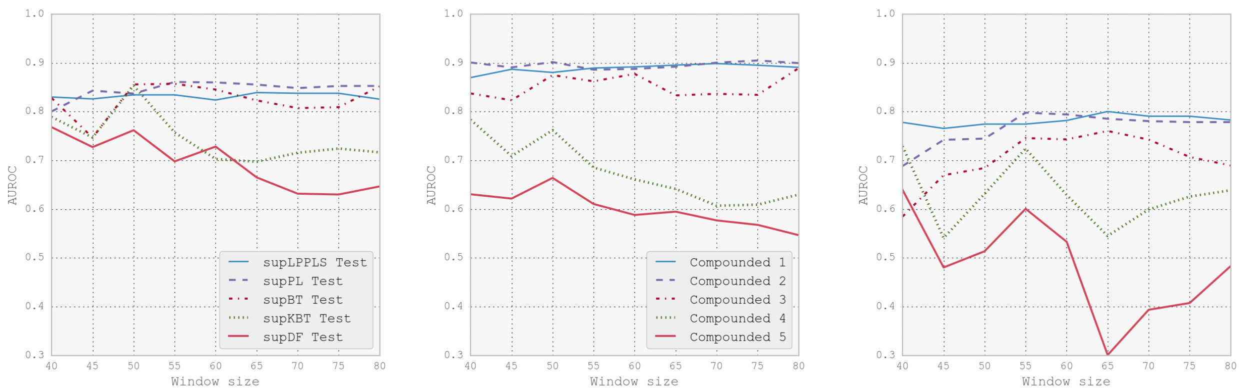
Figure 4. Receiver operating characteristic (ROC) curves derived from crisis logit regressions (25). The independent variables are the bubble indicators derived from log prices (see equations (21) and (24)). The dependent variable $y_{i,t}^{crisis}$ is set to 1 if a crisis started at time t in country i.

Figure 4 and table 5 summarize the results for the single and combined indicators. The differences in performance are appreciable. In general, only the AUROC of in-sample regressions are statistically significant. The performance of out-of-sample AUROCs of single and combined indicators drops visibly and no single indicator yields significant results. Among the out-of-sample tests, the maximum achieved AUROC slightly exceeds the 0.65 value (combined *BI*, with threshold 2) against the 0.77 value of the equivalent quantity reached by the bubble logit regressions of the previous section.

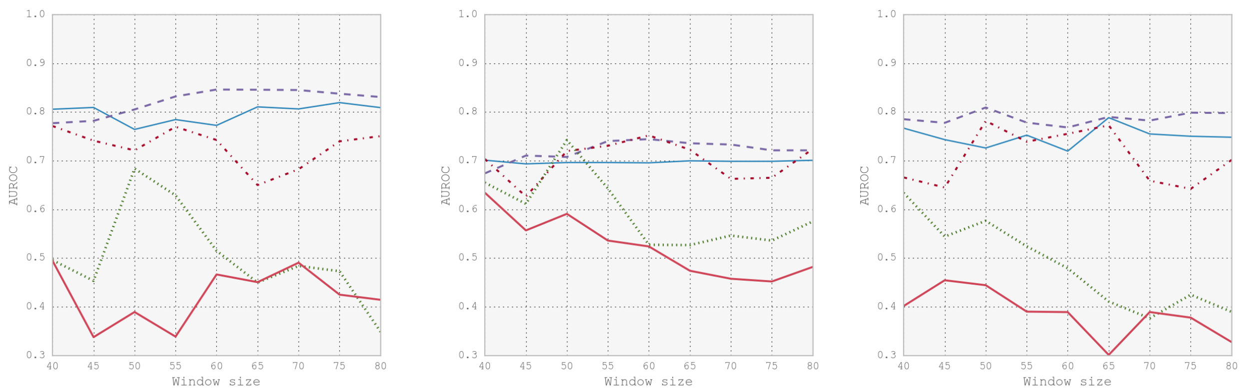
We interpret these results as clear evidence that bubbles and crises are different phenomena. This is not to deny the strong relationship that may exist among them, but to emphasize that they should be treated and analyzed differently. In addition, the poor performance of the indicators to predict crises may be interpreted as a sanity check for our results. The indicators have been able to capture well the events for which they are intended, and have underperformed when used to forecast a related, but different, type of event.

5.5. *Robustness checks*

Finally, we assess the sensitivity of our results to the length of the rolling window employed by the bubble tests. In stationary time series, a longer a window length should increase the power of the tests. However, in non-stationary time series, such as real estate prices, exhibiting commonly several structural breaks, a longer sample can complicate the bubble identification, as the price dynamics can change with changing market conditions. Intuitively, the tests might be unable to distinguish between past and current bubble periods, or label recoveries followed by protracted busts, as new exuberance periods.



(a) In-sample (single indicators) **(b)** In-sample (combined indicators) **(c)** Out-of-sample by period (single indicators)



(d) Out-of-sample by period (combined indicators) **(e)** Out-of-sample by country (single indicators) **(f)** Out-of-sample by country (combined indicators)

Figure 5. Robustness checks to the length of the rolling window used by the bubble tests. Each plot shows the area under the receiver operating characteristic curves AUROCs of the denoted set of bubble indicators, when varying the window length.

Concretely, using rolling window lengths varying between 40 and 80 quarters, with 5 quarters of increment (for a total of 9 different values), we re-calculated the bubble indicators and re-evaluated their performance for the in-sample and out-of-sample set-ups. Figure 5 presents the AUROC values of the individual and compounded indicators.

The supLPPLS, supPL, and supBT-based indicators present relatively low sensitivity to the window length. In-sample (panel 5a), the respective AUROCs remain above 0.8. Out-of-sample by period (panel 5c) and by country (panel 5e), AUROC values of the supLPPLS and supPL indicators vary only slightly, while supBT's performance peaks in the 50-70 length range and only deteriorates mildly (less than 5 percent relative to the peak) as the length becomes much shorter or larger. On the contrary, the supDF and the supKBT statistics do present substantial sensitivity to the window length. In-sample and out-of-sample, there is a visible negative and rather volatile trend as the window length increases. Their AUROC values peak at a length of 50 quarters, while bottom bellow 0.4 in the 65-70 range. Overall, these results are consistent with the relative performance of the indicators discussed in section 5.1.

Also consistent with the previous discussions, combining the indicators generates higher AUROC values relative to the single indicators regardless of the window length. Using 1 or 2 indicators to trigger the bubble signal yields AUROC values close to 0.9 in-sample (panel 5b), and above 0.75 out-of sample (panels 5d and 5f), for most of the window lengths explored. Using 3 indicators yields somewhat more volatile results, though they remain fairly consistent in the 50-65 quarters range. Lastly, combining 4 and 5 indicators clearly leads to under-performing and highly volatile values. The latter should not be surprising, as two of the indicators (those based on the supDF and supKBT tests) yield results very sensitive to the window length.

6. Conclusions

This paper analyzed the information content of statistical tests for bubble detection in the context of international real estate markets. We derived binary indicators from the causal application of selected statistical tests to log house prices, and via logit regressions we assessed the indicators' in-sample and out-of-sample performance in the forecasting of tipping points of housing bubbles and systemic financial crises. We have argued that our benchmark is more robust and objective than any benchmark relying on the ex-post identification of bubble episodes using a technical rule sizing price expansions and corrections.

Overall, our results suggest that the tests contain significant ex-ante information related to the end of the bubble period (with supPL and supBT being the tests that performed best), while combining the indicators yielded the highest performance. Robustness checks that vary the rolling window length employed by the bubble tests further support these conclusions. In addition, the application of the tests to price-to-income and rent-to-income ratios led to a substantial deterioration in the performance of the indicators, which indicates that it should be discouraged.

The patterns that we have documented contribute to the literature in several ways. First, they have a direct impact on the literature on early warning signals and financial stress indices, à la (Misina et al., 2009) and (Christensen and Li, 2014), as bubble-tests derived indicators might be integrated into broader forecast systems. Second, they suggest that bubble tests based on price dynamics, and specially those based on the super-exponential growth pattern, constitute sound tools to identify bubble regimes and study housing bubble episodes. Finally, our results also stress the need for understanding the theoretical underpinnings behind explosiveness in price dynamics, as well as the consequences of such phenomena.

Acknowledgments

We are very grateful to the anonymous reviewers for valuable communications. We would like to acknowledge support for this project from the SNF, under Grant 160305.

Conflict of interest

The authors declare no conflict of interest.

References

Antipa P, Lecat R (2010) The Housing Bubble and Financial Factors: Insights from a Structural Model of the French and Spanish Residential Markets. *Housing Markets in Europe.* Springer, Berlin, Heidelberg 161–186.

Anundsen AK (2015) Econometric regime shifts and the us subprime bubble. *J Appl Econ* 30: 145–169.

Anundsen AK, Jansen ES (2011) Self-reinforcing effects between housing prices and credit. *J Hous Econ* 22: 192–212.

Ayuso J, Restoy F (2006) House prices and rents: An equilibrium asset pricing approach. *J Empir Financ* 13: 371–388.

Barsky RB (2009) The japanese bubble: A 'heterogeneous' approach. NBER Working Paper No. 15052.

Berry M, Dalton T (2004) Housing prices and policy dilemmas: a peculiarly australian problem? *Urban Policy Res* 22: 69–91.

Black A, Fraser P, Hoesli M (2006) House prices, fundamentals and bubbles. *J Bus Finan Account* 33: 1535–1555.

Blanchard OJ, Watson MW (1982) Bubbles, rational expectations and financial markets. NBER Working Paper No. 945.

Bourassa SC, Hoesli M, Oikarinen E (2016) Measuring house price bubbles. *Real Estate Econ* 1–30.

Brunnermeier MK, Oehmke M (2012) Bubbles, financial crises, and systemic risk. *Handbook of The Econ of Financ* 2: 1221–1288.

Busetti F, Taylor A (2004) Tests of stationarity against a change in persistence. *J of Econ* 1: 33–66.

Cameron G, Muellbauer J, Murphy A (2006) Was there a British house price bubble? evidence from a regional panel. CEPR Discussion Paper No. 5619.

Campbell JY, Shiller RJ (1988) The dividend-price ratio and expectations of future dividends and discount factors. *Rev Financ Stud* 1: 195–228.

Case KE, Shiller RJ (2003) Is there a bubble in the housing market? *Brook Pap Econ Act* 299–342.

Christensen I, Li F (2014) Predicting financial stress events: A signal extraction approach. *J Financ Stab* 14: 54–65.

Connor G, Flavin T, O'Kelly B (2012) The us and irish credit crises: Their distinctive differences and common features. *J Int Money Finan* 31: 60–79.

Claussen CA (2013) Are swedish houses overpriced? *Int J Hous Mark Analysis* 6: 180–196.

Dam NA, Hvolbøl TS, Pedersen EH, et al. (2011) Developments in the market for owner-occupied housing in recent years–can house prices be explained. *Danmarks Nationalbank, Monetary Rev, 1st Quarter* 1–82.

Das S, Gupta R, Kanda R (2011) International articles: Bubbles in south african house prices and their impact on consumption. *J of Real Estate Literature* 19: 69–91.

Davidson R, MacKinnon JG (1981) Several tests for model specification in the presence of alternative hypotheses. *Econ* 49: 781–793.

Drehmann M, Borio CE, Tsatsaronis K (2012) Characterising the financial cycle: don't lose sight of the medium term! BIS Working Papers No 380.

Drehmann M, Juselius M (2014) Evaluating early warning indicators of banking crises: Satisfying policy requirements. *Int J Forecast* 30: 759–780.

Edison HJ (2003) Do indicators of financial crises work? An evaluation of an early warning system. *Int J of Financ& Econ* 8: 11–53.

Feigenbaum JA, Freund P (1996) Discrete scale invariance in stock markets before crashes. *Int J of Modern Physics B* 10: 3737–3745.

Filimonov V, Sornette D (2013) A stable and robust calibration scheme of the log-periodic power law model. *Physica A* 392: 3698–3707.

Flood RP, Hodrick RJ (1990) On testing for speculative bubbles. *J Econ Perspect* 4: 85–101.

Francke M, Vujic S, Vos GA (2009) Evaluation of house price models using an ecm approach: the case of the netherlands. OFRC Working Paper No. 2009-05.

Gençay R, Gradojevic N (2010) Crash of'87was it expected?: Aggregate market fears and long-range dependence. *J Empir Financ* 17: 270–282.

Giglio S, Maggiori M, Stroebel J (2014) No-bubble condition: Model-free tests in housing markets. NBER Working Paper No. w20154.

Gisler M, Sornette D (2009) Exuberant innovations: the Apollo program. *Society* 46: 55–68.

Gisler M, Sornette D (2010) Bubbles everywhere in human affairs. *Swiss Financ Institute Res Paper* 137–153

Gisler M, Sornette D, Woodard R (2011) Innovation as a social bubble: The example of the human genome project. *Res Policy* 40: 1412–1425.

Gjerstad SD, Smith VL (2014) *Rethinking housing bubbles: The role of household and bank balance sheets in modeling economic cycles*. Cambridge: Cambridge University Press.

Glindro ET, Subhanij T, Szeto J et al. (2011) Determinants of house prices in nine asia-pacific economies. *J Cent Bank* 7: 163–204.

Homm U, Breitung J (2012) Testing for speculative bubbles in stock markets: a comparison of alternative methods. *J of Financ Econ* 10: 198–231.

Hort K (1998) The determinants of urban house price fluctuations in sweden 1968–1994. *J of housing Econ* 7: 93–120.

Hott C (2012) The influence of herding behaviour on house prices. *J of European Real Estate Res* 5: 177–198.

Hott C, Monnin P (2008) Fundamental real estate prices: an empirical estimation with international data. *J Real Estate Financ Econ* 36: 427–450.

Hüsler A, Sornette D, Hommes CH (2013) Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price. *J Econ Behav Organ* 92: 304–316.

Jacobsen DH, Naug BE (2005) What drives house prices? *Norges Bank. Econ Bulletin* 76: 29–41

Johansen A, Ledoit O, Sornette D (2000) Crashes as critical points. *Int J of Theoretical and Applied Financ* 3: 219–255.

Johansen A, Sornette D (1999) Predicting financial crashes using discrete scale invariance. *J Risk* 1: 5–32.

Jordà Ò, Schularick M, Taylor AM (2015) Leveraged bubbles. *J Monetary Econ* 76: S1–S20.

Kim BH, Min HG (2011) Household lending, interest rates and housing price bubbles in Korea: Regime switching model and kalman filter approach. *Economic Modelling* 28: 1415–1423.

Kim JY (2000) Detection of change in persistence of a linear time series. *J of Econ* 95: 97–116.

Kittler J, Hatef M, Duin RP et al. (1998) On combining classifiers. *IEEE Trans Pattern Anal Mach Intell* 20: 226–239.

Laeven L, Valencia F (2008) Systemic banking crises: a new database. IMF Working Paper No. 08/224.

Leiss M, Nax HH, Sornette D (2015) Super-exponential growth expectations and the global financial crisis. *J Econ Dyn Control* 55: 1–13.

Lin L, Sornette D (2013) Diagnostics of rational expectation financial bubbles with stochastic mean-reverting termination times. *Eur J Financ* 19: 344–365.

Lux T, Sornette D (2002) On rational bubbles and fat tails, *Journal of Money, Credit and Banking* Ohio State University Press, 589–610.

Mack A, Martínez-García E (2011) A cross-country quarterly database of real house prices: a methodological note. Globalization and Monetary Policy Institute Working Paper NO.99

Mason SJ, Graham N (2002) Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation. *Q J R Meteorol Soc* 128: 2145–2166.

Mikhed V, Zemčík P (2009) Do house prices reflect fundamentals? aggregate and panel data evidence. *J Hous Econ* 18: 140–149.

Misina M, Tkacz G (2009) Credit, asset prices, and financial stress. *Int J Cent Bank* 5: 95–122.

Muellbauer J, Murphy A (1997) Booms and busts in the UK housing market. *The Econ J* 107: 1701–1727.

Narayan PK, Sharma SS, Phan DHB (2016) Asset price bubbles and economic welfare. *Int Rev Financ Anal* 44: 139–148.

Neal L, García-Iglesias MC (2013) The economy of spain in the euro-zone before and after the crisis of 2008. *The Quarterly Rev of Econ and Financ* 53: 336–344.

Nobili A, Zollino F (2012) A structural model for the housing and credit markets in italy. Bank of Italy Temi di Discussione Working Paper No. 887.

Palan S (2013) A review of bubbles and crashes in experimental asset markets. *J Econ Surv* 27: 570–588.

Pavlidis E, Yusupova A, Paya I, et al. (2013) Episodes of exuberance in housing markets: in search of the smoking gun. *J Real Estate Financ Econ* 53: 1–31.

Phillips PC, Wu Y, Yu J (2011) Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values?. *Int Econ Rev* 52: 201–226.

Phillips PC, Yu J (2011) Dating the timeline of financial bubbles during the subprime crisis. *Quant Econ* 2: 455–491.

Sarn L, Taylor MP (2003) An empirical investigation of asset price bubbles in latin american emerging financial markets. *Applied Financ Econ* 13: 635–643.

Scatigna M, Szemere R (2015) Bis collection and publication of residential property prices. *Irving Fisher Committee Bulletin* 39: 1–8.

Schatz M, Sornette D (2018) Inefficient bubbles and efficient drawdowns in financial markets. Swiss Finance Institute Research Paper No. 18-49.

Scheinkman J, Xiong W (2004) Heterogeneous beliefs, speculation and trading in financial markets, In: Carmona R.A., inlar E., Ekeland I., Jouini E., Scheinkman J.A., Touzi N., Author, In *Paris-Princeton Lectures on Mathematical Finance 2003*, Springer, Berlin, Heidelberg, 1847: 217–250.

Schularick M, Taylor AM (2012) Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *Am Econ Rev* 102: 1029–1061.

Shi S (2017) Speculative bubbles or market fundamentals? an investigation of us regional housing markets. *Econ Model* 66: 101–111.

Sørensen PB (2013) The swedish housing market: Trends and risks. Finanspolitiska rdet, 1–77.

Sornette D (2002) "Slimming" of power law tails by increasing market returns. *Physica A* 309: 403–418.

SornetteD, Johansen A, Bouchaud JP (1996) Stock market crashes, precursors and replicas. *J Phys I Franc* 6: 167–175.

Stevenson S (2008) Modeling housing market fundamentals: Empirical evidence of extreme market conditions. *Real Estate Econ* 36: 1–29.

Timmermann A (2006) Forecast combinations, *Handbook of economic forecasting,* Elsevier, 135–196.

Valencia F, Laeven L (2012) Systemic banking crises database: An update. IMF Working Paper No. 12/163.

Walks A (2014) Canada's housing bubble story: Mortgage securitization, the state, and the global financial crisis. *Int J Urban Reg Res* 38: 256–284.