Testing for exuberance in house prices using data sampled at different frequencies*

Jesús Otero
Facultad de Economía
Universidad del Rosario
Colombia

Theodore Panagiotidis
Department of Economics
University of Macedonia
Greece

Georgios Papapanagiotou
Department of Economics
University of Macedonia
Greece

July 23, 2021

Abstract

We undertake Monte Carlo simulation experiments to examine the effect of changing the frequency of observations and the data span on the Phillips, Shi, and Yu (2015) Generalised Supremum ADF (GSADF) test for explosive behaviour via Monte Carlo simulations. We find that when a series is characterised by multiple bubbles (periodically collapsing), decreasing the frequency of observations is associated with profound power losses for the test. We illustrate the effects of temporal aggregation by examining two real house price data bases, namely the S&P Case-Shiller real house prices and the international real house price indices available at the Federal Reserve Bank of Dallas.

JEL Classification: C15, C22

Keywords: Exuberant/explosive behaviour; bubbles; Monte Carlo; house prices.

---

*We are grateful to an anonymous reviewer for useful comments.

E-mail addresses: jesus.otero@urosario.edu.co (J. Otero); tpanag@uom.gr (T. Panagiotidis); papapgeorge@uom.edu.gr (G. Papapanagiotou).

1 Introduction

There is no doubt that real estate is the most important financial asset that individuals can acquire during their lifetimes, as it represents a major share in their wealth. The latter justifies the interest in following the dynamic path of real estate prices. There is an additional focus on the periods of time when they reach levels well beyond those that can be justified by their fundamentals, compromising the affordability of the asset. These episodes are typically reported as bubbles if the price in question is measured in relation to an economic fundamental, or viewed as evidence of exuberant/explosive behaviour otherwise (see e.g. Hu and Oxley, 2018). Authors such as Goodhart and Hofmann (2007) identify several channels through which real estate prices may affect economic activity, with the former generally leading developments in the latter. Thus, for example, Shiller (2008), Brunnermeier (2009), and Martin (2010) argue that the real estate bubble in the U.S. in the early 2000 was a major contributing factor in the economic and financial crisis that occurred some years later. However, a real estate bubble can be avoided by sound policy actions. Hence, detecting explosive behaviour in house prices is of prime importance. This is reflected in the vast literature focusing on the detection of bubbles in house prices, which includes Capozza et al. (2004), Quigley et al. (2006), Lai and Van Order (2010, 2017), Gomez-Gonzalez et al. (2018), Gomez-Gonzalez and Sanin-Restrepo (2018), Bourassa et al. (2019), and Fabozzi and Xiao (2019), among others.

The econometrics literature offers alternative methods to identify explosive behaviour in time series. To this day, the Supremum Augmented Dickey-Fuller (SADF) test put forward by Phillips, Wu, and Yu (2011), henceforth PWY, and its subsequent generalised version, denoted GSADF, developed by Phillips, Shi, and Yu (2015), henceforth PSY, have become the most commonly applied testing approaches, with the latter version being the most effective bubble detecting procedure. Indeed, extensive Monte Carlo simulation experiments carried out by PSY indicate that their testing approach, based on recursive flexible regressions of the right-tailed ADF test, provides consistent date-stamping in real time for the start and end dates of multiples bubbles; see, for instance, Phillips et al. (2014) for empirical guidelines regarding the implementation of the tests in practice. The popularity of the approach advocated by PSY is best reflected in its numerous applications. Studies of the presence of explosive behaviour include applications not only to real estate prices, but also to financial time series, such as exchange rates and stock prices; see Phillips and Shi (2020) for a list of recent applications.[1]

This paper is related to the econometric literature that examines the implications of aggregation over individuals and over time. In the context of the tests for explosive behaviour that we referred to above, Pavlidis et al. (2019) study the effect aggregation over individuals has on the likelihood of detecting periods of exuberance. These authors find that aggregation over individuals leads to power loss in both the SADF and the GSADF tests, with the effect being much more prominent in the case of the former. While these authors have already studied the effect of aggregation over individuals, the effects of temporal aggregation remain unknown. Temporal aggregation is relevant because advances in the collection, organisation, storage and retrieval of data lead to the availability of more significant amounts of information. As a result of this, policymakers and academics need to decide carefully the type of data (e.g., monthly, quarterly, annual) that they want to use in their analyses. This paper aims to fill this gap. Thus, the paper falls within the literature that studies the effects of temporal aggregation on the time-series properties of variables (see e.g. Shiller and Perron, 1985; Hooker, 1993; Lahiri and Mamingi, 1995; Ng, 1995; Haug, 2002). To this end, we perform an extensive set of Monte Carlo simulations to study the effects of changing the frequency of observations and the data span on the power of the GSADF test proposed by PSY. The simulation exercises allow for models that contain one, two, and multiple bubbles. These models have been employed extensively in the real estate literature.

The results presented in this paper offer valuable insights for practitioners and policymakers involved in analysing economic and financial data. Indeed, one of the most significant issues in applied econometric work is finding appropriate data. Statistical agencies and central banks tend to collect and publish data on a low frequency (e.g., a quarterly basis) without any information on the data generating process of the variable of interest which, in a sense, is understandable given the limited financial resources available to these institutions. However, nowadays, there are instances in which economic and financial data are aggregated over time for modelling purposes, despite being available at higher frequencies. Ahmad and Paya (2020) indicate that this poses the question of whether the chosen frequency is appropriate to characterise the "true" data generating process

[1] At the time of writing, a Google Scholar search of the PSY paper yields more than 700 citations. These include Phillips and Shi (2019), who present additional theoretical results which show that the PSY testing strategy is also capable of detecting episodes of financial collapse and price meltdowns.

The paper proceeds as follows. Section 2 briefly discusses the existing methodologies on bubble identification. Section 3 outlines the design of the Monte Carlo simulation experiments. Section 4 summarises the results. Section 5 validates the results from the simulations through the examination of the existence of exuberant/explosive behaviour in house prices using data sampled at different frequencies. The last section concludes.

2 Unit root tests for explosive behaviour: An overview

The identification of rational bubbles derives from the definition of exuberance in terms of explosive autoregressive behaviour. Following this definition and the work of Lucas (1978) and using dividend and stock price data Shiller (1981), Blanchard and Watson (1982) and West (1988) relate the existence of bubbles with an inconsistency in the efficient market hypothesis. Nelson and Plosser (1982) however, support that apparent evidence for bubbles can be reinterpreted in terms of market fundamentals that are unobserved by the researcher.

Diba and Grossman (1988) and Hamilton and Whiteman (1985) recommend another method of testing for rational bubbles by investigating the stationarity properties of asset prices and observable fundamentals. Using standard unit root tests applied to real U.S. Standard and Poor's Composite Stock Price Index data over the period 1871-1986, Diba and Grossman (1988) test levels and differences of stock prices for non-stationarity, finding support in the data for non-stationarity in levels but stationarity in differences. Since differences of an explosive process still manifest explosive characteristics, these findings appear to reject the presence of a market bubble in the data. Although the results were less definitive, further tests by Diba and Grossman (1988) provide confirmation of cointegration between stock prices and dividends over the same period, supporting the conclusion that prices do not diverge from long-run fundamentals and thereby giving additional evidence against bubble behaviour.

Evans (1991) shows through simulation methods that non-recursive unit root tests have low power and frequently cannot reject the the null of no explosive behaviour even when present in the data. Nonlinear dynamics, such as those displayed by mildly explosive processes, may lead the standard right-tailed ADF test to findings of spurious stationarity. Intuitively, this is the case because increases followed by downward corrections make the process appear mean-reverting and stationary in finite samples even when it is inherently not. An extensive literature review on the aforementioned proposed econometric methodologies can be found in Flood and Hodrick (1990) and Gürkaynak (2008).

PWY propose a test based on the supremum t-statistic from a recursive estimation of the ADF test using a forward expanding sample. Drawing on the notation employed by Phillips et al. (2015), the idea is to estimate the augmented Dickey and Fuller (1979) (ADF) regression with intercept:

$$\Delta x_t = \alpha_{r_1,r_2} + \beta_{r_1,r_2} x_{t-1} + \sum_{i=1}^{k} \delta^i_{r_1,r_2} \triangle x_{t-i} + \varepsilon_t, \tag{1}$$

where Δ is the first difference operator, x_t is the variable under consideration at time t, k is the number of lags of Δx_t included to allow for serial correlation in the residuals, and r_1 and r_2 denote the starting and ending observations used for estimation, respectively. With T denoting the number of observations in the sample, r_1 and r_2 are expressed as fractions of T such that $r_2 = r_1 + r_w$, where r_w is the window size of the regression, which is also expressed as a fraction of T. Eq. (1) is estimated using $T_w = \lfloor T r_w \rfloor$ observations, where $\lfloor \cdot \rfloor$ denotes the floor function which yields the integer part of the argument. The error term is ε_t. Under the null hypothesis there is a unit root, $H_0 : \beta_{r_1,r_2} = 0$, while the alternative the variable x_t exhibits explosive behaviour, that is $H_1 : \beta_{r_1,r_2} > 0$. To test the null hypothesis the required ADF t-statistic associated to $\beta_{r_1,r_2} = 0$ in Eq. (1) is denoted $\text{ADF}_{r_1}^{r_2}$. In this setting, PWY propose a statistic based on the supremum t-statistic that results from a forward recursive estimation of Eq. (1):

$$\text{SADF}(r_0) = \sup_{r_2 \in [r_0, 1]} \text{ADF}_0^{r_2}, \tag{2}$$

where the window size, r_w, expands from the smallest sample window width r_0, which provides the first t-statistic of the recursion, to the last observation that is available.

Homm and Breitung (2011) show that compared to others procedures the PWY approach performs satisfactorily. However, in the presence of multiple bubbles in the sample, the PWY test suffers power losses due to complex nonlinear structure involved in the multiple breaks that produce the bubble phenomena. To address this, PSY generalise the PWY testing approach, and in doing so produce a test which is not affected by multiple bubbles. The result is the generalised supremum ADF (GSADF) test which, as implied by its name, requires the estimation of a much larger number of regressions where the first observation used for estimation varies from 0 to $r_2 - r_0$, and the last observation varies from r_0 to 1, that is:

$$\mathrm{GSADF}(r_0) = \sup_{\substack{r_2 \in [r_0,1] \\ r_1 \in [0, r_2 - r_0]}} \mathrm{ADF}_{r_1}^{r_2}. \tag{3}$$

PSY recommend implementing the GSADF test by choosing a low value of k, say $k = 0, 1$, and determining the minimum window size using the rule $r_0 = (0.01 + 1.8/\sqrt{T})$. To perform inference, the right-tail SADF and GSADF statistics are tabulated using Monte Carlo simulations; in what follows, we use the critical values provided by Vasilopoulos et al. (2020) in the R package `exuber`.

One useful application of the testing approaches based on recursive and recursive flexible windows estimation of Eq. (1) is that they can be used in real time to date-stamp episodes of bubbles or exuberance if the unit-root null hypothesis is rejected. To see how this works, let us assume that there is interest in establishing whether a particular observation, say r_2, belongs to a phase of explosive behaviour. PSY recommend to perform a supADF test on a sample sequence where the endpoint is fixed at the observation of interest r_2, and expands backwards to the starting point, r_1, which varies between 0 and $(r_2 - r_0)$. In this context, the backward SADF statistic is defined as the supremum of the resulting sequence of ADF statistics:

$$\mathrm{BSADF}_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} \mathrm{ADF}_{r_1}^{r_2}. \tag{4}$$

Next, the $\mathrm{BSADF}_{r_2}(r_0)$ statistic is compared against the corresponding critical values of the $\mathrm{SADF}(r_0)$ for $\lfloor r_2 T \rfloor$ observations.[2]

3 Design of the Monte Carlo simulations

We consider two datasets. In the first one, we generate time series of monthly data for 150 years. We then obtain sample sizes of 50, 100, and 150 years of monthly, quarterly, and annual observations. In the second dataset, the generated time series contain daily data for 10 years and calculate sample sizes of 5 and 10 years of daily, weekly and fortnightly observations. In both cases, all samples contain at least one bubble. To convert to lower frequency data, we use the method of systematic sampling and the method of averaging with non-overlapping observations.

[2] Phillips et al. (2015) mention that this date-stamping procedure is more general than the earlier suggestion in Phillips et al. (2011) in which $r_1 = 0$ in (4), and so it is more effective at identifying episodes of multiple bubbles.

We first investigate whether temporal disaggregation causes size distortions (giving a false positive result). Following Phillips, Shi, and Yu (2015), we use the data generating process (DGP): $x_t = T^{-1} + x_{t-1} + \epsilon_t$, to examine how frequently the GSADF test due to PSY rejects the null of a unit root in favour of an explosive alternative.

Next, we simulate four different DGPs to examine the effect of temporal disaggregation on the power of the test. The first two (DGP1 and DGP2) are based on Blanchard (1979) and Evans (1991) and contain multiple explosive incidents (periodically collapsing). The other two (DGP3 and DGP4) are based on Phillips, Shi, and Yu (2015) and contain one and two mildly explosive bubbles, respectively. For the process based on Blanchard (1979), we construct a series consisting of two regimes which occur with probability π and $1 - \pi$. In the first regime, the bubble grows exponentially,

$$x_t = \frac{1+r}{\pi} + x_{t-1} + error \tag{5}$$

whereas in the second regime, the bubble collapses to a white noise.

The next model, proposed by Evans (1991) follows the DGP:

$$x_t = \begin{cases} (1+r)x_{t-1}u_t, & \text{if} \quad x_{t-1} \leq a \\ [\delta + \pi^{-1}(1+r)\theta_t(x_{t-1}) - (1+r)^{-1}\delta]u_t, & \text{if} \quad x_{t-1} > a \end{cases} \tag{6}$$

where δ and a are positive parameters with $0 < \delta < (1+r)a$, u_t is an exogenous *iid* positive random variable with $E_{t-1}u_t = 1$, and θ_t is an exogenous independently and identically distributed Bernoulli process (independent of u) which takes the value 1 with probability π and 0 with probability $1 - \pi$, where $0 < \pi \leq 1$. When $x_{t-1} \leq a$ the bubble grows at an average rate of $1 + r$. When $x_{t-1} > a$ the bubble expands at an increased rate of $(1 + r)\pi^{-1}$. We set $a = 1, \delta = 0.5, \tau = 0.05, \pi = 0.7$ and $r = 0.05$.

Turning to the processes based on Phillips, Shi, and Yu (2015), we construct two additional time series. The first one according to the DGP,

$$x_t = \begin{cases} x_{t-1} + \epsilon_t, & \text{if} \quad t \in [1, t_e) \\ \delta_T x_{t-1} + \epsilon_t, & \text{if} \quad t \in [\tau_e, \tau_f] \\ \sum_{k=\tau_f+1}^{t} \epsilon_k + x^*_{\tau_f}, & \text{if} \quad t \in (\tau_f, T] \end{cases} \tag{7}$$

where T is the sample size, $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$, $\delta_T = 1 + cT^{-a}$ with $c > 0$ and $a \in (0, 1)$, and $x^*_{\tau_f} = x_{\tau_e} + x^*$ with $x^* = O_p(1)$. During the pre- and post- bubble periods, x is a pure random walk process. During the bubble expansion period, $B = [\tau_e, \tau_f]$, is mildly explosive process with expansion rate given by the autoregessive coefficient δ_T. We set

$c = 1$, $a = 0.6$, $T = 1800$, $\tau_e = 240$, and $\tau_f = 480$ in the first case and $T = 3650$, $\tau_e = 560$, and $\tau_f = 1125$ in the second.

The second model extends model 7 and uses the DGP:

$$x_t = \begin{cases} x_{t-1} + \epsilon_t, & \text{if} \quad t \in [1, \tau_{1e}) \\ \delta_T x_{t-1} + \epsilon_t, & \text{if} \quad t \in [\tau_{1e}, \tau_{1f}] \cup [\tau_{2e}, \tau_{2f}] \\ \sum_{k=\tau_{1f}+1}^{t} \epsilon_k + x^*_{\tau_1 f}, & \text{if} \quad t \in (\tau_{1f}, \tau_{2e}) \\ \sum_{l=\tau_{2f}+1}^{t} \epsilon_k + x^*_{\tau_2 f}, & \text{if} \quad t \in (\tau_{2f}, T] \end{cases} \tag{8}$$

After the collapse of the first bubble, x_t resumes a martingale path until time $\tau_{2e} - 1$, and a second episode of exuberance begins at τ_{2e}. The expansion process lasts until τ_{2f} and collapses to a value of x^*_{2f}. The process then continues on a martingale path until the end of the sample period T. The expansion duration of the first bubble is assumed to be longer than that of the second bubble. We set $c = 1$, $a = 0.6$. For the case of monthly data for 150 years, we also set $T = 1800$, $\tau_{1e} = 240$, $\tau_{1f} = 480$, $\tau_{2e} = 720$, and $\tau_{2f} = 840$. For the case of daily data for 10 years, we have $T = 3650$, $\tau_{1e} = 300$, $\tau_{1f} = 1500$, $\tau_{2e} = 1800$, and $\tau_{2f} = 2170$. The plots in Figure 1 exemplify the type of dynamic behaviour that arises when one simulates the four DGPs just described using 50 years of monthly data.

The power probabilities are based on size adjusted critical values at the 5% significance level. The minimum window size r_0 of the regression is calculated using the rule $r_0 = 0.1 + 1.8/\sqrt{T}$ and the lag order is set to 0. Phillips et al. (2015) find that size increases with the lag length. The results are based on 10000 replications. All reported simulations were programmed in R version 3.6.1 using the packages `psymonitor` (Phillips et al., 2019) and `exuber` (Vasilopoulos et al., 2020). These scripts, available upon request, replicate all simulations and the results of the empirical application.

4 Monte Carlo simulation results

This section presents the Monte Carlo results. Table 1 reports size probabilities for the GSADF test. Overall, size is controlled reasonably well when using the data sampled at the monthly and daily frequencies, and also when using the skip sampling method of temporal aggregation. However, the averaging with non-overlapping method yields greater size distortions.

Table 2 summarises the main power probability results of our Monte Carlo experiments. The values reported in this table denote the probability of detecting explosive behaviour based on the GSADF test at the 5% significance level using size-corrected critical values. Considering the models that contain multiple bubbles (periodically collapsing), that is DGP1 and DGP2, we observe significant power losses as one moves from high- to low-frequency sampled data. For example, using DGP2 and a sample size of 100 years, skip sampling temporal aggregation reduces the probability of correctly rejecting the unit root null hypothesis from 100% with monthly data to 95.5% with quarterly data, and then to 7.1% with annual data; similarly, temporal aggregation through averaging with non-overlapping observations yields power probabilities of 95.7% and 32% for quarterly and annual data, respectively. These results highlight the importance of data frequency for bubble testing as moving from high- to low-frequency data increases the chance that a bubble will build and collapse between neighbouring observations.

Let us now consider the simulation results when applying the GSADF test to the two DGPs considered by PSY. Regarding the process with a single bubble, DGP3, changing the frequency has little effect on the test. Turning to the process which contains two explosive incidents, DGP4, we observe power losses only when we decrease the sample length to 50 years.[3] At this point, one could wonder whether alternative methods for identifying explosive dynamics perform better than PSY under the effect of temporal aggregation. However, for a bubble detecting strategy to be successful, it must rely on a recursive procedure, and recursive procedures suffer power losses from reducing the frequency of observations. Thus, it is unlikely that the statistical power of alternative methods based on recursive estimation is not adversely affected when using temporally aggregated data.

5 Testing for exuberance in house prices

We illustrate the effects of temporal aggregation on the power of the GSADF test by examining real house prices from two sources, the first one is the S&P Case-Shiller real house price data from US cities, and the second one is the international real house price data available at the Federal Reserve Bank of Dallas.[4]

[3]Qualitatively similar results (not reported here for brevity, but available from the authors upon request) are obtained for the SADF test of PWY, although in this case the test is less powerful than the GSADF test of PSY.

[4]The results in this section are based on the Stata command `radf` (see Otero and Baum, 2020).

5.1 S&P Case-Shiller Home Price Indices

The S&P Case-Shiller Home Price Indices are repeat-sales house price indices for selected cities in the United States. They are constructed on monthly basis, starting from January 1987. The time span covers the period between 1987m1 to 2020m6 and provides a total of 402 observations. Using this sample, we construct quarterly and annual versions of the data for ten cities. We then consider three sample sizes, 1987m1 to 2000m12, 1987m1 to 2010m12 and 1987m1 to 2020m6 of monthly, quarterly and annual data. Figure A.1 plots the indices in ten selected cities for the largest sample. The figures support the view of two incidents of explosive behaviour. The first one occurs during the period between 2000 and 2007 and the second starts around 2012 and lasts until today. Only in Denver, the magnitude of the second incident is greater than the magnitude of the first one (Boston and San Francisco are close cases as well).

Table 3 reports the results from the GSADF test on the ten indices from selected US cities for all samples. Using the monthly and quarterly data for the shortest data span, we are able detect exuberant behaviour at the 1% significance level for all cities except Las Vegas. If we apply the test on annual data we are still able to reject the null hypothesis for six out of the ten cities (five when the method of systematic sampling is used). However, in several cases, we are able to do so only at the 5% significance level. For the two longer sample periods, the test detects explosive behaviour at the 1% significance level for all cities irrespective of the frequency of observations.

To illustrate these findings graphically, Figure 2 plots the recursive GSADF statistic (black line) against the 95% critical value sequence (red line) for the city of Los Angeles, using all three data frequencies over shortest sample period (to save space we only display the results for the aggregating method of averaging with non-overlapping observations). An explosive episode occurs when the statistic exceeds the critical value. For the case of monthly data, the real time dating strategy identifies three episodes during the periods 1989m2-1989m11, 1991m3-1996m7 and 1997m12-2000m12. Applying the test on quarterly data results in the identification of only two episodes of shorter duration. These episodes occur during the periods 1992q3-1994q3 and 1998q3-2000q4. Finally, the GSADF strategy detects no episodes for the case of annual data.

5.2 World Home Price Indices

As a second example, the Federal Reserve Bank of Dallas gathers home price index series for 23 countries on a quarterly basis going back to the first quarter of 1975. The reader interested on the methodological aspects of the database is referred to Mack and Martínez-García (2011). Here, we shall analyse the sample period between 1975q1 and 2019q4 for a total of $T = 180$ observations on each individual index. Since the data are already available on a quarterly basis, in what follows we only aggregate the data series to the annual frequency. Using quarterly and annual data we employ the test on three different data spans, 1975q1 to 1999q4, 1975q1 to 2009q4 and 1975q1 to 2019q4. Plots of the quarterly time series over the longest time period are displayed in Figure B.1. As can be seen in the figures the evolution of house prices differs among the countries. Most countries experienced soaring house prices in the early 2000s. However, in Germany, Israel, Japan and South Korea this is not the case.

The results for the GSADF test are reported in Table 4. Using all available information, we are able to detect explosive incidents at the 1% significance level in all countries except South Korea and Spain. Changing the frequency to annual but keeping the data span from 1975 to 2019 affects most of the examined countries. For example, the test identifies explosive behaviour only at the 5% significance level in Australia, Belgium and New Zealand while it fails to identify such episodes in Canada, Denmark, France, Italy, South Africa and Sweden. Decreasing the sample length but using quarterly observations does not affect the performance of the test and we are still able to reject the null hypothesis at the 1% significance level. These findings are in line with the simulation results where we observed that decreasing the frequency observations affects the power of the GSADF more than decreasing the data span. In Figure 3 we illustrate these findings by plotting the recursive GSADF statistic (black line) against the 95% critical value sequence (red line) for Canada, using quarterly and annual data over the shortest sample period (once more we focus on the results obtained with the aggregation method of averaging with non-overlapping observations). For the case of quarterly data, the test detects three three episodes of exuberance. One episode from 1980q3 to 1981q4 and two consecutive episodes from 1987q3 to 1987q4 and from 1988q2 to 1990q2. No such episodes are detected when annual data are used.

6 Concluding remarks

In this paper we examine the effect of temporal aggregation on the PSY test for explosive behaviour. This test, which relies on recursive flexible regressions of the right-tailed ADF test, is employed extensively on monitoring financial, commodity, and real estate markets for speculative bubbles and exuberant dynamics. We examine whether lowering the frequency and/or decreasing the sample span affects the power of this testing strategy. The analysis considers both simulated data and actual real house price data. The Monte Carlo simulation results indicate that when a series is characterised by multiple bubbles (periodically collapsing), decreasing the frequency of observations is associated with profound power losses for the test. As for the evidence based on actual house prices, we employ data from ten U.S. cities (S&P Case-Shiller Home Price Indices) and Home price Index series from 23 countries (from the Federal Reserve Bank of Dallas). The examination of all data available revealed the presence of explosive behaviour for almost all U.S. cities and countries included in the sample. The test also identified explosive incidents in shorter sample periods. However, for shorter sample periods the lower frequency data resulted in incapability of the test to detect exuberant behaviour in several instances.

Overall, the results indicate that temporal aggregation has a substantial effect on the performance of PSY procedure. In addition, lowering the frequency of the data has a greater effect than decreasing the total sample span, especially when the time-series under examination is characterised by multiple bubbles (periodically collapsing). Although we do not suggest an optimal data frequency, we consider that using monthly data for house prices and higher frequency data for other asset prices is optimal. Our findings are relevant to policy makers and central banks since the two procedures have been proven to be the most successful methods for detecting exuberance in a wide variety of macroeconomic and financial bubbles. We leave the examination of the effect of temporal aggregation on the power of the PSY test in the presence of heteroskedasticity and leverage effect for future work.

References

Ahmad, Y. and I. Paya (2020). Temporal aggregation of random walk processes and implications for economic analysis. *Studies in Nonlinear Dynamics and Econometrics 24*(2).

Blanchard, O. (1979). Speculative bubbles, crashes and rational expectations. *Economics Letters 3*(4), 387–389.

Blanchard, O. and M. Watson (1982). Bubbles, rational expectations and financial markets. In P. Wachtel and L. Mass (Eds.), *Crises in the Economic and Financial Structure: Bubbles, Bursts and Shock*. Lexington Press.

Bourassa, S., M. Hoesli, and E. Oikarinen (2019). Measuring house price bubbles. *Real Estate Economics 47*(2), 534–563.

Brunnermeier, M. (2009). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives 23*(1), 77–100.

Capozza, D., P. Hendershott, and C. Mack (2004, 02). An anatomy of price dynamics in illiquid markets: Analysis and evidence from local housing markets. *Real Estate Economics 32*, 1–32.

Diba, B. and H. Grossman (1988). Explosive rational bubbles in stock prices? *American Economic Review 78*(3), 520–30.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366), 427–431.

Evans, G. (1991). Pitfalls in testing for explosive bubbles in asset prices. *American Economic Review 81*(4), 922–30.

Fabozzi, F. and K. Xiao (2019). The timeline estimation of bubbles: The case of real estate. *Real Estate Economics 47*(2), 564–594.

Flood, R. and R. Hodrick (1990). On testing of sprculative bubbles. *Journal of Economic Perspectives 4*(2), 85–101.

Gomez-Gonzalez, J. E., J. Gamboa-Arbeláez, J. Hirs-Garzón, and A. Pinchao-Rosero (2018). When bubble meets bubble: Contagion in OECD countries. *The Journal of Real Estate Finance and Economics 56*(4), 546–566.

Gomez-Gonzalez, J. E. and S. Sanin-Restrepo (2018). The maple bubble: A history of migration among Canadian provinces. *Journal of Housing Economics 41*(C), 57–71.

Goodhart, C. and B. Hofmann (2007). *House Prices and the Macroeconomy: Implications for Banking and Price Stability.* Oxford: Oxford University Press.

Gürkaynak, S. R. (2008). Econometric tests of asset price bubbles: Taking stock. *Journal of Economic Surveys 22*(1), 166–186.

Hamilton, J. and C. Whiteman (1985). The observable implications of self-fulfilling expectations. *Journal of Monetary Economics 16*(3), 353–373.

Haug, A. A. (2002). Temporal aggregation and the power of cointegration tests: A Monte Carlo study. *Oxford Bulletin of Economics and Statistics 64*(4), 399–412.

Homm, U. and J. Breitung (2011). Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Hooker, M. A. (1993). Testing for cointegration: Power versus frequency of observation. *Economics Letters 41*(4), 359 – 362.

Hu, Y. and L. Oxley (2018). Do 18th century 'bubbles' survive the scrutiny of 21st century time series econometrics? *Economics Letters 162*(C), 131–134.

Lahiri, K. and N. Mamingi (1995). Testing for cointegration: Power versus frequency of observation — another view. *Economics Letters 49*(2), 121 – 124.

Lai, R. and R. Van Order (2010). Momentum and house price growth in the united states: Anatomy of a bubble. *Real Estate Economics 38*(4), 753–773.

Lai, R. and R. Van Order (2017). U.S. house prices over the last 30 years: Bubbles, regime shifts and market (in)efficiency. *Real Estate Economics 45*(2), 259–300.

Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica 46*(6), 1429–1445.

Mack, A. and E. Martínez-García (2011). A cross-country quarterly database of real house prices: a methodological note. Globalization Institute Working Papers 99, Federal Reserve Bank of Dallas.

Martin, R. (2010). The local geographies of the financial crisis: from the housing bubble to economic recession and beyond. *Journal of Economic Geography 11*(4), 587–618.

Nelson, C. and C. Plosser (1982). Trends and random walks in macroeconmic time series: Some evidence and implications. *Journal of Monetary Economics 10*(2), 139–162.

Ng, S. (1995). Testing for unit roots in flow data sampled at different frequencies. *Economics Letters 47*(3-4), 237–242.

Otero, J. and C. F. Baum (2020). RADF: Stata module to calculate unit root tests for explosive behaviour.

Pavlidis, E., E. Martínez-García, and V. Grossman (2019). Detecting periods of exuberance: A look at the role of aggregation with an application to house prices. *Economic Modelling 80*(C), 87–102.

Phillips, P. C. B. and S. Shi (2019). Detecting financial collapse and ballooning sovereign risk. *Oxford Bulletin of Economics and Statistics 81*(6), 1336–1361.

Phillips, P. C. B. and S. Shi (2020). Real time monitoring of asset markets: bubbles and crises. In H. D. Vinod and C. R. Rao (Eds.), *Handbook of Statistics: Financial, Macro and Micro Econometrics Using R*, Volume 42, pp. 61–80. Amsterdam: Elsevier.

Phillips, P. C. B., S. Shi, and I. Caspi (2019). *psymonitor: Real Time Monitoring of Asset Markets: Bubbles and Crisis.* R package version 0.0.2.

Phillips, P. C. B., S. Shi, and J. Yu (2014). Specification sensitivity in right-tailed unit root testing for explosive behaviour. *Oxford Bulletin of Economics and Statistics 76*(3), 315–333.

Phillips, P. C. B., S. Shi, and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*(4), 1043–1078.

Phillips, P. C. B., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review 52*(1), 201–226.

Quigley, J., M. Hwang, and J. Son (2006, 02). The dividend pricing model: New evidence from the Korean housing market. *The Journal of Real Estate Finance and Economics 32*, 205–228.

Shiller, R. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review 71*(3), 421–436.

Shiller, R. J. (2008). *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It.* Princeton University Press.

Shiller, R. J. and P. Perron (1985). Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters 18*(4), 381 – 386.

Vasilopoulos, K., E. Pavlidis, S. Spavound, and E. Martínez-García (2020). *exuber: Econometric Analysis of Explosive Time Series*. R package version 0.4.1.

West, K. D. (1988). Bubbles, fads and stock price volatility tests: A partial evaluation. *The Journal of Finance 43*(3), 639–656.

Table 1: Size of the GSADF test

	Monthly data				Daily data	
Number of years	50	100	150		5	10
Monthly data	0.045	0.047	0.047	Daily data	0.046	0.039
Quarterly skip	0.050	0.047	0.047	Weekly skip	0.036	0.066
Annual skip	0.041	0.054	0.063	Fortnightly skip	0.041	0.037
Quarterly average	0.195	0.208	0.230	Weekly average	0.180	0.312
Annual average	0.128	0.190	0.226	Fortnightly average	0.176	0.196

Notes: Skip refers to the systematic sampling method for aggregating data and average to the averaging with non-overlapping observations method. For monthly (daily) observations the initial sample consists of 50 (5) years. The minimum window size r_0 of the regression is calculated using the rule $r_0 = 0.1 + 1.8/\sqrt{T}$.

Table 2: Size-adjusted power of the GSADF test

	Monthly data				Daily data	
Number of years	50	100	150		5	10
DGP1:						
Monthly data	0.998	1.000	1.000	Daily data	1.000	1.000
Quarterly skip	0.894	0.959	0.975	Weekly skip	0.575	0.747
Annual skip	0.095	0.117	0.153	Fortnightly skip	0.105	0.145
Quarterly average	0.884	0.961	0.984	Weekly average	0.712	0.869
Annual average	0.236	0.339	0.416	Fortnightly average	0.368	0.525
DGP2:						
Monthly data	0.999	1.000	1.000	Daily data	1.000	1.000
Quarterly skip	0.871	0.955	0.985	Weekly skip	0.577	0.766
Annual skip	0.044	0.071	0.106	Fortnightly skip	0.063	0.085
Quarterly average	0.858	0.957	0.983	Weekly average	0.735	0.864
Annual average	0.224	0.320	0.404	Fortnightly average	0.370	0.507
DGP3:						
Monthly data	0.929	0.929	0.929	Daily data	0.986	0.986
Quarterly skip	0.933	0.932	0.934	Weekly skip	0.986	0.986
Annual skip	0.934	0.937	0.936	Fortnightly skip	0.986	0.985
Quarterly average	0.931	0.933	0.934	Weekly average	0.987	0.986
Annual average	0.935	0.934	0.931	Fortnightly average	0.986	0.985
DGP4:						
Monthly data	0.730	0.957	0.958	Daily data	0.923	0.999
Quarterly skip	0.742	0.957	0.957	Weekly skip	0.925	0.999
Annual skip	0.734	0.957	0.955	Fortnightly skip	0.923	0.999
Quarterly average	0.743	0.957	0.958	Weekly average	0.921	0.999
Annual average	0.736	0.954	0.955	Fortnightly average	0.918	0.999

Notes: The data-generating processes are DGP1 and DGP2 for multiple (periodically collapsing) bubbles as in Blanchard (1979) and Evans (1991), respectively; and DGP3 and DGP4 for one and two bubbles, respectively, as in Phillips, Shi, and Yu (2015). Skip refers to the systematic sampling method for aggregating data and average to the averaging with non-overlapping observations method. The table reports the probability of rejecting the null of the GSADF test in favour of the (true) alternative of a mildly explosive process at the 5% significance level using size-corrected critical values. For monthly (daily) observations the initial sample consists of 50 (5) years.

Table 3: GSADF test on S&P/Case-Shiller home price index in selected cities

Period	City	Monthly data	Quarterly data Avg.	Quarterly data Skip	Annual data Avg.	Annual data Skip
1987m1–2000m12	Boston	12.036‡	10.518‡	8.216‡	5.825‡	6.860‡
	Chicago	3.413‡	4.153‡	3.058‡	3.961‡	3.321†
	Denver	10.494‡	10.967‡	7.109‡	5.461‡	4.818‡
	Las Vegas	0.937	0.722	0.286	0.679	0.335
	Los Angeles	5.642‡	2.867‡	2.595‡	0.741	0.831
	Miami	5.393‡	4.321‡	3.856‡	0.205	0.989
	New York	7.903‡	6.633‡	5.396‡	2.150†	2.447†
	San Diego	7.359‡	4.605‡	4.593‡	1.175	1.778†
	San Francisco	14.014‡	8.381‡	8.535‡	3.480‡	5.543‡
	Washington DC	9.458‡	8.503‡	7.100‡	−0.134	0.094
1987m1–2010m12	Boston	12.493‡	11.983‡	8.216‡	6.511‡	6.405‡
	Chicago	6.794‡	7.544‡	7.252‡	6.696‡	5.864‡
	Denver	10.494‡	9.983‡	7.109‡	3.588‡	3.563‡
	Las Vegas	26.238‡	20.301‡	14.535‡	12.602‡	13.805‡
	Los Angeles	17.608‡	12.788‡	10.252‡	6.171‡	7.049‡
	Miami	30.544‡	26.383‡	21.427‡	14.063‡	15.744‡
	New York	12.762‡	10.350‡	8.623‡	5.883‡	5.632‡
	San Diego	17.982‡	12.747‡	10.849‡	6.686‡	8.207‡
	San Francisco	14.014‡	8.867‡	8.535‡	3.254‡	2.611‡
	Washington DC	24.250‡	18.976‡	15.790‡	10.952‡	7.726‡
1987m1–2020m6	Boston	12.493‡	11.983‡	8.216‡	5.335‡	4.772‡
	Chicago	6.794‡	7.544‡	7.252‡	6.696‡	5.864‡
	Denver	10.494‡	9.983‡	7.109‡	3.588‡	3.563‡
	Las Vegas	26.238‡	20.301‡	14.535‡	11.081‡	13.805‡
	Los Angeles	17.608‡	12.788‡	10.252‡	6.171‡	7.049‡
	Miami	30.544‡	26.383‡	21.427‡	14.063‡	15.744‡
	New York	12.762‡	10.350‡	8.623‡	5.883‡	5.632‡
	San Diego	17.982‡	12.747‡	10.849‡	6.632‡	7.894‡
	San Francisco	14.014‡	8.867‡	8.535‡	3.254‡	2.406†
	Washington DC	24.250‡	18.976‡	15.790‡	9.239‡	6.835‡

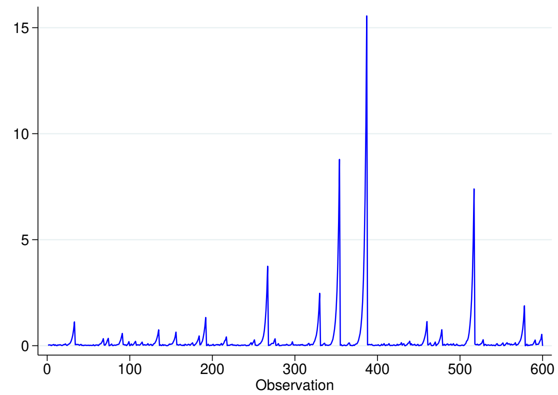
Note: Test regressions do not include lags of the dependent variable. Skip refers to the systematic sampling method for aggregating data and avg to the averaging with non-overlapping observations method. † and ‡ denote significance at the 5% and 1% levels, respectively. Inference based on critical values obtained by Monte Carlo simulations with 2000 replications (see Vasilopoulos et al., 2020).

Table 4: GSADF test on real house price indices in selected countries

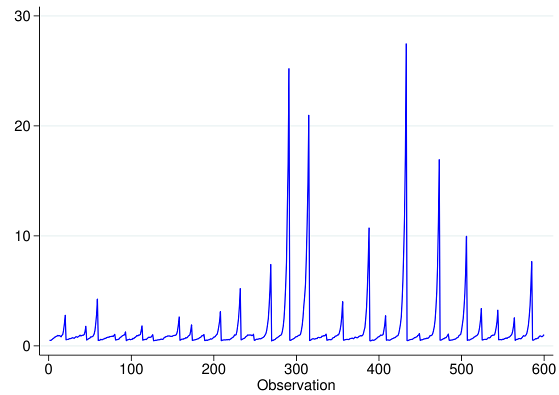
	1975q1–1999q4			1975q1–2009q4			1975q1–2019q4		
Country	Q	A avg.	A skip	Q	A avg.	A skip	Q	A avg.	A skip
Australia	$4.339^{\ddagger}$	$1.840^{\dagger}$	$2.393^{\dagger}$	$10.243^{\ddagger}$	$8.406^{\ddagger}$	$8.213^{\ddagger}$	$10.243^{\ddagger}$	$8.406^{\ddagger}$	$8.213^{\ddagger}$
Belgium	$6.287^{\ddagger}$	$1.856^{\dagger}$	$1.785^{\dagger}$	$7.953^{\ddagger}$	$4.517^{\ddagger}$	$4.216^{\ddagger}$	$7.953^{\ddagger}$	$4.517^{\ddagger}$	$4.216^{\ddagger}$
Canada	$6.988^{\ddagger}$	-0.446	-0.587	$8.441^{\ddagger}$	$7.034^{\ddagger}$	$5.171^{\ddagger}$	$8.441^{\ddagger}$	$6.812^{\ddagger}$	$5.171^{\ddagger}$
Croatia	$5.804^{\ddagger}$	$1.736^{\dagger}$	1.085	$5.804^{\ddagger}$	1.491	0.989	$5.804^{\ddagger}$	1.308	0.871
Denmark	$2.624^{\ddagger}$	1.577	1.242	$11.642^{\ddagger}$	$4.070^{\ddagger}$	$3.809^{\ddagger}$	$10.543^{\ddagger}$	$4.070^{\ddagger}$	$3.809^{\ddagger}$
Finland	$9.637^{\ddagger}$	$3.240^{\ddagger}$	$2.350^{\dagger}$	$6.663^{\ddagger}$	$3.240^{\ddagger}$	$3.014^{\ddagger}$	$6.641^{\ddagger}$	$3.240^{\ddagger}$	$3.014^{\ddagger}$
France	$5.069^{\ddagger}$	0.790	0.623	$11.151^{\ddagger}$	$8.684^{\ddagger}$	$8.335^{\ddagger}$	$11.151^{\ddagger}$	$8.380^{\ddagger}$	$8.225^{\ddagger}$
Germany	$3.425^{\ddagger}$	-0.022	-0.172	$3.191^{\ddagger}$	$2.193^{\dagger}$	1.643	$7.121^{\ddagger}$	$4.223^{\ddagger}$	$3.851^{\ddagger}$
Ireland	$13.785^{\ddagger}$	$6.096^{\ddagger}$	$6.718^{\ddagger}$	$13.785^{\ddagger}$	$6.555^{\ddagger}$	$6.936^{\ddagger}$	$13.785^{\ddagger}$	$6.555^{\ddagger}$	$6.936^{\ddagger}$
Israel	$2.241^{\dagger}$	$2.340^{\dagger}$	1.514	$2.241^{\dagger}$	$2.340^{\dagger}$	1.514	$4.018^{\ddagger}$	$2.354^{\dagger}$	$2.482^{\dagger}$
Italy	$2.892^{\ddagger}$	0.464	0.622	$7.234^{\ddagger}$	1.599	1.617	$7.234^{\ddagger}$	1.269	1.265
Japan	$6.401^{\ddagger}$	$2.771^{\ddagger}$	$2.931^{\ddagger}$	$7.542^{\ddagger}$	$7.565^{\ddagger}$	$6.997^{\ddagger}$	$7.542^{\ddagger}$	$2.887^{\ddagger}$	$3.068^{\ddagger}$
Luxembourg	$7.208^{\ddagger}$	1.398	1.421	$8.092^{\ddagger}$	$5.331^{\ddagger}$	$5.223^{\ddagger}$	$8.092^{\ddagger}$	$4.604^{\ddagger}$	$4.765^{\ddagger}$
Netherlands	$6.342^{\ddagger}$	$6.509^{\ddagger}$	$7.138^{\ddagger}$	$7.484^{\ddagger}$	$9.045^{\ddagger}$	$8.375^{\ddagger}$	$6.591^{\ddagger}$	$9.045^{\ddagger}$	$8.375^{\ddagger}$
New Zealand	$4.631^{\ddagger}$	$2.018^{\dagger}$	$1.877^{\dagger}$	$7.834^{\ddagger}$	$4.289^{\ddagger}$	$3.424^{\ddagger}$	$7.834^{\ddagger}$	$4.289^{\ddagger}$	$3.424^{\ddagger}$
Norway	$5.893^{\ddagger}$	$2.859^{\ddagger}$	$1.894^{\dagger}$	$5.893^{\ddagger}$	$3.219^{\ddagger}$	$2.831^{\ddagger}$	$5.893^{\ddagger}$	$3.219^{\ddagger}$	$2.831^{\ddagger}$
South Africa	$3.447^{\ddagger}$	-0.258	-0.115	$10.087^{\ddagger}$	$7.347^{\ddagger}$	$4.891^{\ddagger}$	$10.087^{\ddagger}$	$7.347^{\ddagger}$	$4.891^{\ddagger}$
South Korea	1.423	0.425	0.482	1.423	0.425	0.482	1.423	0.415	0.482
Spain	1.574	0.166	-0.409	$10.044^{\ddagger}$	$6.929^{\ddagger}$	$5.322^{\ddagger}$	$10.044^{\ddagger}$	$4.935^{\ddagger}$	$4.784^{\ddagger}$
Sweden	$4.722^{\ddagger}$	0.554	0.482	$7.016^{\ddagger}$	$5.121^{\ddagger}$	$4.835^{\ddagger}$	$7.016^{\ddagger}$	$5.121^{\ddagger}$	$4.835^{\ddagger}$
Switzerland	$14.823^{\ddagger}$	$4.436^{\ddagger}$	$4.209^{\ddagger}$	$13.255^{\ddagger}$	$3.961^{\ddagger}$	$3.480^{\ddagger}$	$13.255^{\ddagger}$	$4.238^{\ddagger}$	$3.799^{\ddagger}$
United Kingdom	$7.853^{\ddagger}$	$3.442^{\ddagger}$	$4.570^{\ddagger}$	$7.853^{\ddagger}$	$5.175^{\ddagger}$	$3.794^{\ddagger}$	$7.853^{\ddagger}$	$4.361^{\ddagger}$	$3.702^{\ddagger}$
United States	$3.671^{\ddagger}$	$3.112^{\ddagger}$	$2.294^{\dagger}$	$11.099^{\ddagger}$	$10.116^{\ddagger}$	$9.824^{\ddagger}$	$11.099^{\ddagger}$	$10.116^{\ddagger}$	$9.824^{\ddagger}$

Note: Test regressions do not include lags of the dependent variable. Q and A refer to annual and quarterly data, respectively. Skip refers to the systematic sampling method for aggregating data and avg to the averaging with non-overlapping observations method. † and ‡ denote significance at the 5% and 1% levels, respectively. Inference based on critical values obtained by Monte Carlo simulations with 2000 replications (see Vasilopoulos et al., 2020).

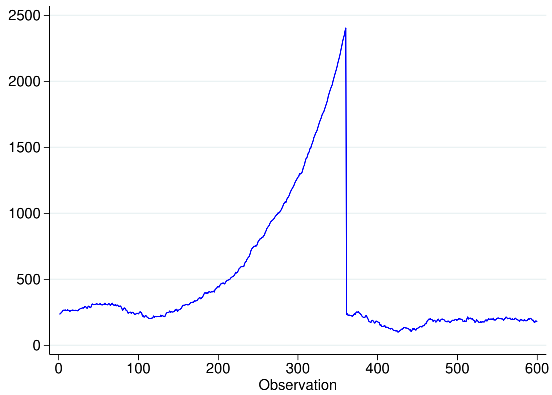
Figure 1: Example simulation of the data-generating processes used to compute power



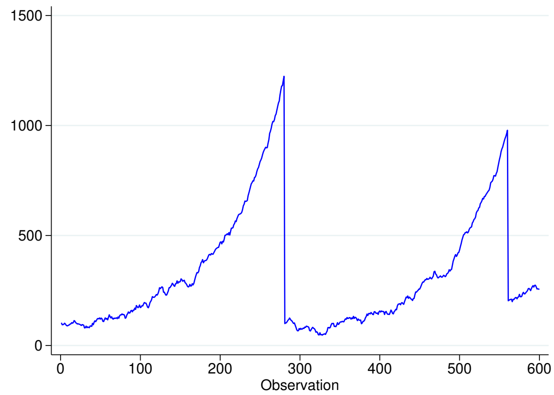
(a) DGP1: Blanchard (1979)



(b) DGP2: Evans (1991)

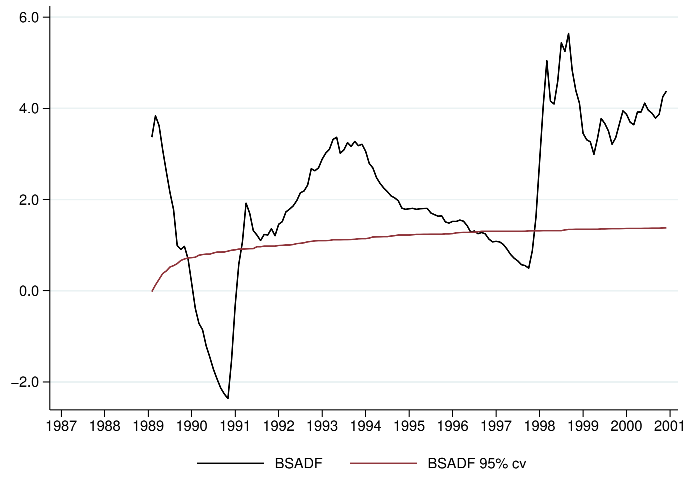


(c) DGP3: Phillips, Shi, and Yu (2015)

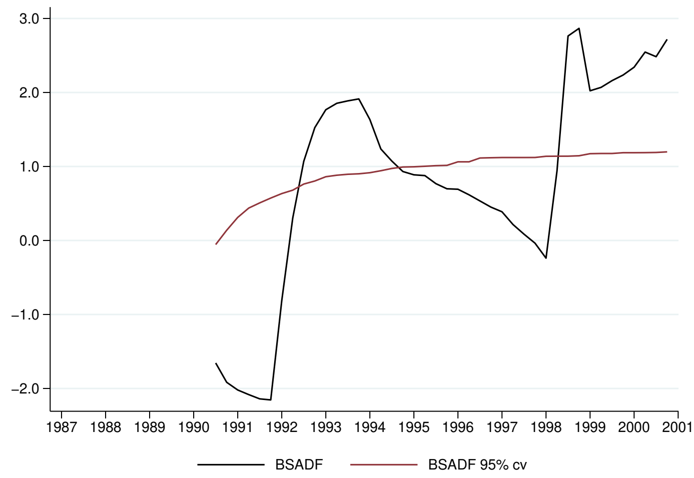


(d) DGP4: Phillips, Shi, and Yu (2015)

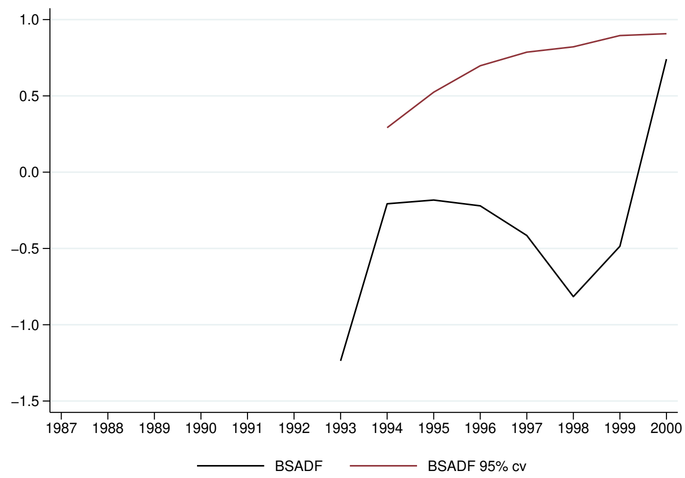
Figure 2: Date stamping real house price index in Los Angeles (1987m1 – 2000m12)



(a) Monthly data

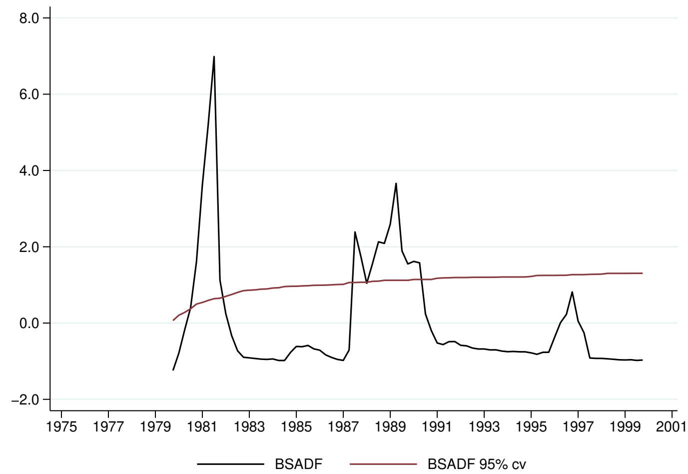


(b) Quarterly data (avg.)

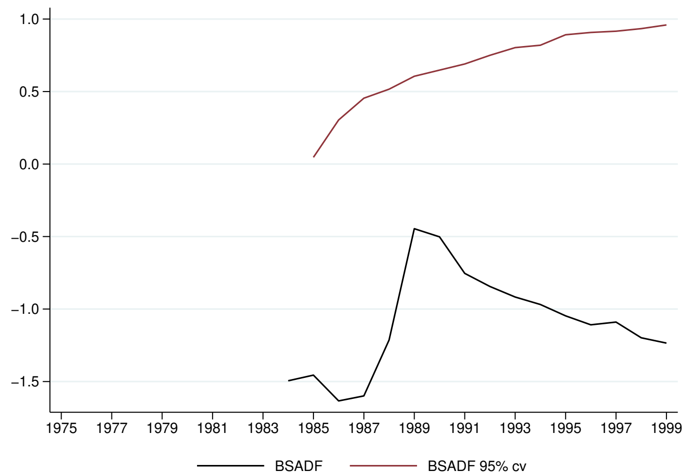


(c) Annual data (avg.)

Figure 3: Date stamping real house price index in Canada (1975q1 – 1999q4)



(a) Quarterly data (avg.)

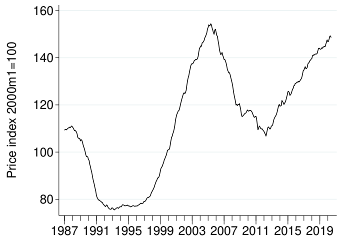


(b) Annual data (avg.)

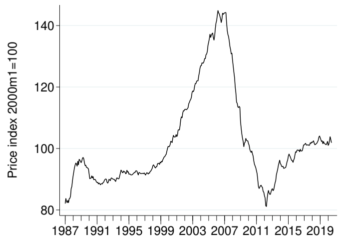
A Appendix: S&P Case-Shiller data

The price indices of Chicago, Los Angeles and New York have been deflated by the corresponding local CPI. For the other cities we used the CPI of the corresponding region as follows: Boston (North East); Denver, Las Vegas, San Diego and San Francisco (West); Miami and Washington DC (South).

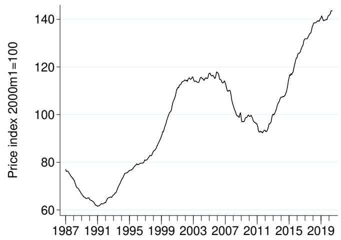
Figure A.1: S&P Case-Shiller home real price indices (1987m1 – 2020m6)



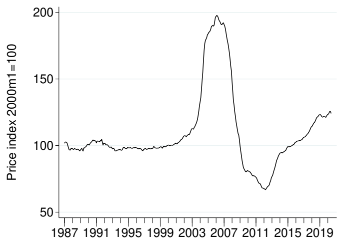
(a) Boston



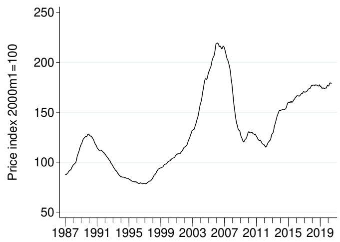
(b) Chicago



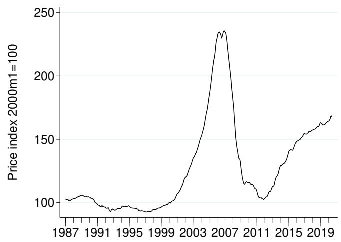
(c) Denver



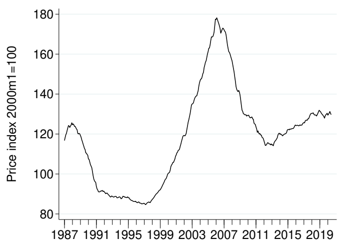
(d) Las Vegas



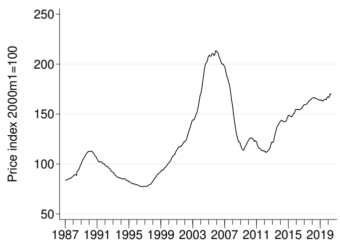
(e) Los Angeles



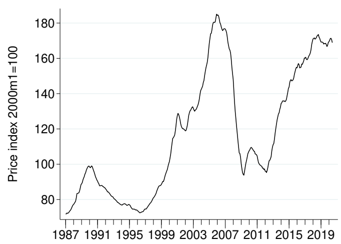
(f) Miami



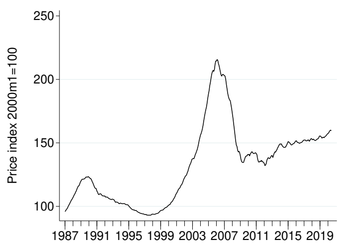
(g) New York



(h) San Diego



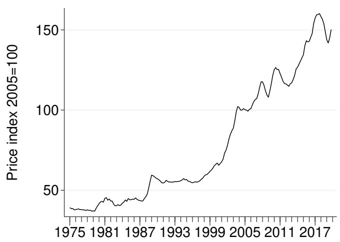
(i) San Francisco



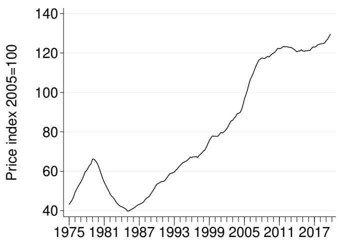
(j) Washington DC

B Appendix: Federal Reserve Bank of Dallas data

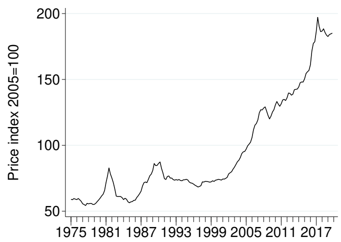
Figure B.1: Real house price indexes in selected countries (1975q1 – 2019q4)



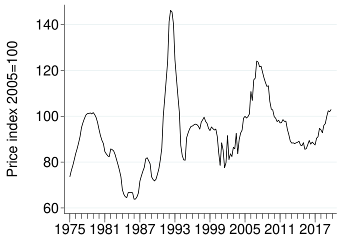
(a) Australia



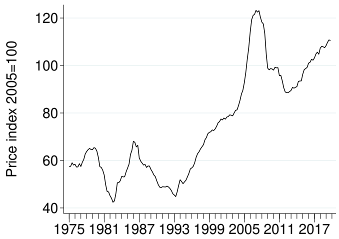
(b) Belgium



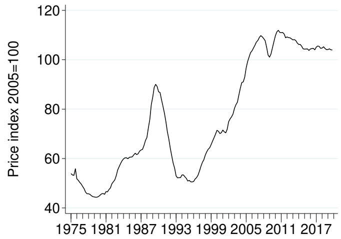
(c) Canada



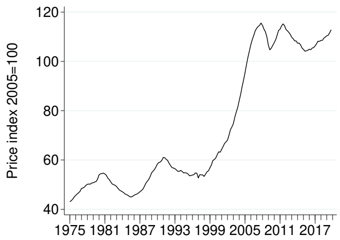
(d) Croatia



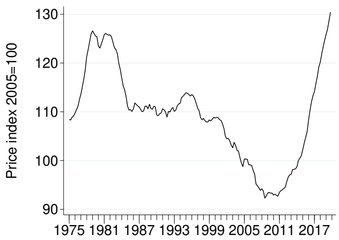
(e) Denmark



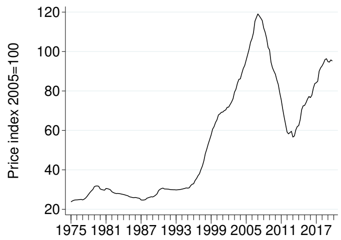
(f) Finland



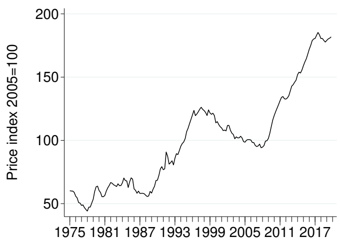
(g) France



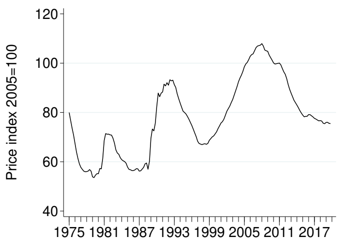
(h) Germany



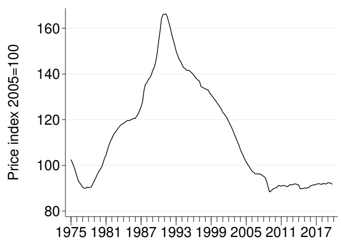
(i) Ireland



(j) Israel



(k) Italy



(l) Japan