

Detecting Periods of Exuberance: A Look at the Role of Aggregation with an Application to House Prices

Abstract

The recently developed SADF and GSADF unit root tests of Phillips and Yu (2011) and Phillips *et al.* (2015a,b) have become popular in the literature for detecting exuberance in asset prices. In this paper, we examine through simulation experiments the effect of cross-sectional aggregation on the power properties of these tests. The simulation design considered is based on simulated data and actual housing data for both U.S. metropolitan areas and international housing markets and thus allows us to draw conclusions for different levels of aggregation. Our findings suggest that aggregation lowers the power of both the SADF and GSADF tests. The effect, however, is much larger for the SADF test. We also provide evidence that tests based on panel data techniques, namely the panel GSADF test recently proposed by Pavlidis *et al.* (2016), can perform substantially better than univariate tests applied to aggregated series. Furthermore, we also illustrate the date-stamping procedure under the univariate/panel GSADF procedure uncovering novel evidence on the role of interest rates and policy uncertainty as factors explaining episodes of widespread mildly explosive dynamics in housing markets.

Keywords: Aggregation; Mildly explosive time series; Right-tailed unit-root tests; Sup ADF (SADF) test; Generalized sup ADF (GSADF) test; House prices.

JEL classification: C22; C12; G12; R30; R31.

1. Introduction

Mildly explosive behavior is modeled by an autoregressive process with a root that exceeds unity but remains within the vicinity of one and approaches unity as the sample size tends to infinity, as in Phillips and Magdalinos (2007a, 2007b) and Phillips and Magdalinos (2012). The literature refers to instances of mildly explosive behavior as periods of exuberance, a terminology which we henceforth also adopt in the paper.

Mildly explosive behavior represents a small departure from martingale behavior, but one that is consistent with the submartingale property commonly used to describe rational bubbles in the asset pricing literature. Diba and Grossman (1988a; 1988b) were among the first to argue within the standard asset pricing equation framework that, given a constant discount factor, the detection of such a departure in the data could be seen as evidence of non-fundamental (bubble-like) behavior.

Diba and Grossman (1988a; 1988b) were also among the seminal papers to propose the use of unit root and cointegration tests for detecting mildly explosive behavior. However, it is a well-known fact that standard unit root tests have extremely low power in detecting episodes of explosive behavior in asset prices that end with a large drop. As has been shown by a number of studies, this type of nonlinear dynamics, which are consistent with the presence of periodically-collapsing bubbles in asset markets, can frequently lead to finding spurious stationarity even though the underlying asset price process is inherently explosive (Evans, 1991; Gurkaynak, 2008).

To alleviate this problem, researchers have recently proposed recursive (right-tailed) unit root tests which, by utilizing subsamples of data, perform remarkably better in identifying periods of explosiveness (as shown in Homm and Breitung, 2012).¹ These recursive unit root tests—namely the supremum ADF (SADF) and the Generalized SADF (GSADF)—have been widely employed over the last decade for testing asset price dynamics (Phillips and Yu, 2011; Phillips *et*

¹ Conventional testing methods for detecting evidence consistent with the presence of rational bubbles in the time series include unit root and cointegration tests (Diba and Grossman, 1988a; 1988b), variance bound tests (LeRoy and Porter, 1981; Shiller, 1981), specification tests (West, 1987), and Chow and CUSUM-type tests (Homm and Breitung, 2012).

al., 2015a,b).² A popular application of these recursive unit root tests is in real estate markets. Following the housing boom of the early and mid-2000s and its subsequent market collapse leading to the 2008 global recession, there has been a plethora of studies that test for explosive dynamics in real estate prices (e.g., Phillips and Yu, 2011; Pavlidis *et al.* 2016; Engsted *et al.* 2016; Yusupova *et al.* 2016; Shi, 2017; Hu and Oxley, 2018).

A common feature of all the above studies is that they employ house price indices which are constructed by aggregating data across locations. A question of direct practical relevance to applied researchers is: what is the effect of the level of cross-sectional aggregation on the performance of recursive unit root tests? This question is particularly relevant for housing since there is substantial heterogeneity in the dynamics of housing prices across local/regional markets (e.g., house prices in San Francisco behave very differently than prices in Denver or Washington). Even within the same market, there is great variation across locations so that aggregate and disaggregate series may behave very differently (see, e.g., Gyourko *et al.* 2006).

Although potentially important, the effect of cross-sectional aggregation on the power of unit root tests remains unexplored and often overlooked. The existing econometric literature has mainly focused on the role of temporal aggregation and sample frequency on the performance of standard unit root tests. For instance, Shiller and Perron (1985), Perron (1991), and Campbell and Perron (1991) find through Monte Carlo simulations that the power of unit root tests is mainly affected by the time span, and much less by sampling frequency. In line with this finding, Pierse and Snell (1995) show theoretically that the asymptotic local power of a unit root test is not dependent on sample frequency. However, Boswijk and Klaassen (2012) demonstrate that this result does not hold for time series that exhibit fat tails and volatility clustering. With regard to temporal aggregation, Choi (1992) illustrates that in finite samples aggregating subinterval data can also result in substantial power losses.

² Recent studies suggest that, during the last financial crisis, macro and financial time series have been described by significant nonlinearities (Tsagkanos and Siriopoulos, 2015, and Evgenidis *et al.* 2017). The tests considered in this paper are linear but recursive in nature. By allowing for time variation in regression coefficients, the tests can capture nonlinearities in the data (such as changes in persistence or asymmetric effects). This fact is based on White's Theorem that shows that any nonlinear model can be approximated by a time-varying linear model. For a thorough discussion see Granger (2008).

In this paper, we examine the role of cross-sectional aggregation by conducting one simulation experiment based on artificial processes that deterministically switch between martingale and explosive behavior. Moreover, we also conduct two large simulation experiments based on actual real house price data. The first of the experiments with actual data uses monthly real house price data from the S&P/Case-Shiller 10-City Composite Index for the 10 largest metropolitan areas in the U.S., deflated by their corresponding local CPI. The second experiment utilizes national-level, quarterly real house price index data for 23 mostly-advanced economies, deflated with the corresponding country's PCE deflator, obtained from the Federal Reserve Bank of Dallas' International House Price Database (Mack and Martínez-García, 2011).

The main conclusion that emerges from our simulation exercises is that aggregation lowers the power of both the SADF test of Phillips and Yu (2011) and the GSADF test of Phillips *et al.* (2015a,b). The decline is substantially larger for the SADF test than for the GSADF test, which provides an important reason to prefer the latter for detecting periods of exuberance in asset prices—particularly whenever researchers have to rely on highly aggregated data (as often happens with housing data). We also provide evidence that the panel GSADF test recently proposed by Pavlidis *et al.* (2016) is preferable whenever disaggregated data is available.

Furthermore, we illustrate the date-stamping strategy under the univariate and panel GSADF methodologies to establish a chronology of episodes of global exuberance for the 23 countries covered in the International House Price Database (Mack and Martínez-García, 2011). Using a dynamic panel probit model we show explosive dynamics in the data are, consistent with theory, partly driven by fundamental behavior tied interest rate spreads and policy uncertainty. Housing demand and macro factors (real personal disposable income growth per capita, stock market valuation gains, and real GDP growth) as well as a number of financial variables (the current account ratio over GDP, real credit growth in the nonfinancial sector, and market volatility measured by the VIX index) contribute as well. Our findings also reinforce the view that the degree of aggregation can be a limiting factor on the overall usefulness of right-tailed tests for monitoring asset markets as it can impact on our empirical inferences.

The remainder of the paper is structured as follows: Section 2 outlines the SADF test of Phillips and Yu (2011), the GSADF test of Phillips *et al.* (2015a,b), and the panel GSADF test of Pavlidis *et al.* (2016). This section also discusses the potential role of aggregation on the power of the recursive (right-tailed) unit root tests. Section 3 defines our Monte Carlo simulation design and presents results for the simulated as well as for the actual data. Section 4 discusses illustrates the practical implementation of these recursive unit root tests for monitoring the dynamics of asset prices. It also shows the aggregation bias that may arise on empirical inferences based on date-stamping a chronology of periods of exuberance after rejecting the null of no exuberance in-sample. Section 5 concludes, while the Appendix provides a stylized theoretical model motivating the spillovers from monetary policy—through policy uncertainty and interest rate spreads—into real house prices that we investigate in our application in section 4 and additional technical details on the implementation of the recursive unit root tests.

2. Testing Mildly Explosive Behavior

Phillips and Magdalinos (2007a, 2007b) define a mildly explosive root using the following data-generating process (DGP) for the observed time series:

$$y_t = \delta_T y_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.d. (0, \sigma^2), \quad (1)$$

with the intercept set at zero for simplicity, where $\delta_T = 1 + \frac{c}{T^\alpha}$, $\alpha \in (0, 1)$ and T denotes the sample size. Whenever $c > 0$, such a root is explosive and approaches unity at a rate slower than $O(T^{-1})$ as $T \rightarrow \infty$.³ Subtracting y_{t-1} from both sides, the process in (1) can be expressed

³ Phillips and Magdalinos (2007a, 2007b) and Phillips and Magdalinos (2012) provide a large-sample asymptotic theory for this class of mildly explosive processes that enables econometric inference, unlike what occurs for purely explosive processes. Autoregressive processes with a purely explosive root, $y_t = \delta y_{t-1} + \epsilon_t$, $\epsilon_t \sim NID(0, \sigma^2)$, $|\delta| > 1$, were first discussed by White (1958) and Anderson (1959). Assuming a zero initial condition for y_t , an asymptotic Cauchy limit distribution theory for the OLS/ML estimator exists. However, the asymptotic distribution of the estimator is ultimately dependent on the distributional assumptions imposed on the innovations (Anderson, 1959)—the imposed Gaussianity of the errors cannot be relaxed without changing the asymptotic distribution. Hence, there is no general framework for asymptotic inference on purely explosive processes.

as $\Delta y_t = \beta_T y_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d. (0, \sigma^2)$ where Δ is the difference operator, and $\beta_T = \delta_T - 1$ is the corresponding coefficient to be tested.

If serial correlation is a concern, a standard parametric autoregressive approach to deal with it consists in extending equation (1) to an AR(k+1) process (Said and Dickey, 1984). The approach is based on generalizing the process to be $\theta_{k+1}(B)y_t = \epsilon_t$, where $\theta_{k+1}(B) = 1 - \theta^1 B - \dots - \theta^k B^k - \theta^{k+1} B^{k+1}$ defines the lag operator. A unit root in $\theta_{k+1}(B)$ corresponds to $\theta_{k+1}(1) = 0$. Then, testing for a unit root is more easily performed by rewriting the augmented regression model in the following form:⁴

$$\Delta y_t = \beta_T y_{t-1} + \sum_{j=1}^k \psi^j \Delta y_{t-j} + \epsilon_t, \quad \epsilon_t \sim i.i.d. (0, \sigma^2), \quad (2)$$

where $\beta_T = -\theta_{k+1}(1)$ and $\psi^1 = -(\theta^2 + \theta^3 + \dots + \theta^k + \theta^{k+1})$, $\psi^2 = -(\theta^3 + \dots + \theta^k + \theta^{k+1})$, ..., $\psi^k = -(\theta^{k+1})$.

The procedure studied in the paper for detecting mildly explosive behavior consists in recursively applying the Augmented Dickey-Fuller (ADF) test for the null of a unit root against the alternative of a mildly explosive root (the right tail of the distribution) based on the specification in (2).

First, some notation is required to describe the recursive implementation of the supremum ADF (SADF) of Phillips and Yu (2011) and the generalized SADF (GSADF) of Phillips *et al.* (2015a,b). We can think of the full sample as being normalized on the interval $[0,1]$ (i.e., divided by the total number of observations T). We denote r_1 and r_2 as the corresponding fractions of the sample which define the beginning and end of a given subsample such that $0 \leq r_1 < r_2 \leq 1$. We denote by $r_w = r_2 - r_1$ the window size of the regression estimation, while r_0 is the fixed initial window required by the econometrician such that the subsample ending in r_2 satisfies that $r_2 \in [r_0, 1]$ (i.e., r_0 is the minimum window size).

⁴ The ADF approach generalizes the Dickey and Fuller (1979) test by parametrically removing the structural autocorrelation in the time series, but otherwise implements the same testing procedure.

Second, the empirical specification used for testing is the following recursive formulation of the ADF auxiliary regression equation:

$$\Delta y_t = a_{r_1, r_2} + \beta_{r_1, r_2} y_{t-1} + \sum_{j=1}^k \psi_{r_1, r_2}^j \Delta y_{t-j} + \epsilon_t, \quad \epsilon_t \sim i.i.d. (0, \sigma_{r_1, r_2}^2), \quad (3)$$

where y_t denotes the generic time series tested for explosiveness, Δy_{t-j} for $j = 1, \dots, k$ are the differenced lags of the time series, and ϵ_t is an i.i.d. error term. Furthermore, k is the maximum number of lags included in the specification, while a_{r_1, r_2} , β_{r_1, r_2} , and ψ_{r_1, r_2}^j for $j = 1, \dots, k$ are the corresponding regression coefficients—the intercept, the autoregressive coefficient, and the coefficients of the lagged first differences—when estimated over the (normalized) subsample beginning in r_1 and ending in r_2 .

Finally, whenever disaggregated data is available, our statistical toolkit for the detection of periods of mildly explosive behavior also includes the novel panel GSADF test of Pavlidis *et al.* (2016). The GSADF test was designed to overcome the SADF's lack of power in identifying multiple episodes of periodically-collapsing mildly explosive behavior in-sample. The panel GSADF test, in turn, offers a more flexible alternative than the univariate GSADF test because it explicitly models the cross-sectional dependencies and heterogeneity present in the constituent series that are otherwise muddled together in the aggregated series.

2.1 Standard Right-Tailed Augmented Dickey-Fuller (ADF) Test

Setting $r_1 = 0$ and $r_2 = r_0 = 1$ yields the standard ADF test statistic over the full sample,

$$ADF_0^1 = \frac{\hat{\beta}_{0,1}}{s.e.(\hat{\beta}_{0,1})}. \text{ Under the } I(1) \text{ null, the limit distribution of } ADF_0^1 \text{ is given by } \frac{\int_0^1 \tilde{W}(r) d\tilde{W}(r)}{\left(\int_0^1 \tilde{W}(r)^2 dr\right)^{\frac{1}{2}}},$$

where \tilde{W} is a demeaned Wiener process (Brownian motion). With this, we test the null hypothesis of a unit root in y_t , $H_0: \beta_{0,1} = 0$, against the alternative of mildly explosive behavior, $H_1: \beta_{0,1} > 0$. Whenever ADF_0^1 exceeds the corresponding right-tailed critical value from its limit distribution, the unit root hypothesis is rejected in favor of the alternative of mildly explosive behavior.

Evans (1991) shows through simulation methods that non-recursive unit root tests like ADF_0^1 (and cointegration tests as well) have low power and frequently cannot reject the null of no explosive behavior 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 its DGP is not.

2.2 Sup ADF (SADF) Test

In order to deal with the effect of a collapse occurring within sample on the performance of the standard right-tailed ADF test (ADF_0^1), Phillips and Yu (2011) proposed a procedure based on the recursive estimation of the ADF regression equation in (3) on subsamples of the data. The approach uses a forward expanding estimation subsample with the end of the subsample r_2 increasing from $r_0 \in (0,1)$ (the fixed minimum size for the initial window) to one (the last available observation). The starting point of each estimation is kept fixed at $r_1 = 0$, so the expanding window size of the regression (over the normalized sample) is simply given by $r_w = r_2$. Then, incrementing the window size $r_2 \in [r_0, 1]$ with one additional observation at a time, the recursive estimation of the ADF regression equation in (3) over the forward expanding subsample yields a sequence of $ADF_0^{r_2} = \frac{\hat{\beta}_{0,r_2}}{s.e.(\hat{\beta}_{0,r_2})}$ statistics.

The Phillips and Yu (2011) test statistic, called sup ADF (SADF), is defined as the supremum value of the sequence of $ADF_0^{r_2}$ statistics expressed as follows:

$$SADF(r_0) = \sup_{r_2 \in [r_0, 1]} ADF_0^{r_2}. \quad (4)$$

Under the $I(1)$ null, the limit distribution of the $SADF(r_0)$ statistic is given by

$$\sup_{r_2 \in [r_0, 1]} \frac{\int_0^{r_2} \tilde{W}(r) d\tilde{W}(r)}{\left(\int_0^{r_2} \tilde{W}(r)^2 dr \right)^{\frac{1}{2}}}. \text{ Whenever } SADF(r_0) \text{ exceeds the corresponding right-tailed critical value}$$

from its limit distribution, the unit root hypothesis is rejected in favor of mildly explosive behavior.

The rolling-window structure of the $SADF(r_0)$ test leads to improved power in detecting mildly explosive behavior relative to what can be achieved with the standard ADF_0^1 test alone.

Furthermore, Hogg and Breitung (2012) show through simulation experiments also that the $SADF(r_0)$ test generally outperforms alternative testing methods commonly used to detect a single structural break in the persistence of the process (from $I(1)$ to explosive).

The alternative tests considered by Hogg and Breitung (2012) aim to detect a permanent structural break in the persistence of the process and, as a consequence, perform well only when the series becomes explosive but never bursts in-sample. Intuitively, the $SADF(r_0)$ test performs better than those alternatives because it deals with series where at most one episode of explosiveness occurs and collapses in-sample. However, the power of the $SADF(r_0)$ test and its performance deteriorate in the presence of multiple (more than one) recurring and periodically-collapsing episodes of exuberance, as established in Phillips *et al.* (2015a,b).

2.3 Generalized SADF (GSADF) Test

Phillips *et al.* (2015a,b) proposed another recursive (right-tailed) unit root test, the Generalized SADF (GSADF), covering a larger number of subsamples than the $SADF(r_0)$ test by relaxing the requirement that the starting point of the subsample r_1 be kept fixed. This additional margin of flexibility on the estimation window of the $GSADF(r_0)$ results in substantial power gains, consistent with multiple and periodically-collapsing episodes of explosiveness in the data.

The GSADF approach builds on the forward expanding estimation subsample strategy of the SADF procedure, but instead allows the starting point of the subsample r_1 to change. The initial window size r_0 satisfies that $r_0 < r_2$, while the expanding window size of the regression (over the normalized sample) is defined as $r_w = r_2 - r_1$. Incrementing the window size $r_2 \in [r_0, 1]$ with one additional observation at a time over each starting point of the sample $r_1 \in [0, r_2 - r_0]$, the recursive estimation of the ADF regression equation in (3) yields a sequence of

$$ADF_{r_1}^{r_2} = \frac{\hat{\beta}_{r_1, r_2}}{s.e.(\hat{\beta}_{r_1, r_2})} \text{ statistics.}$$

The Phillips *et al.* (2015a,b) test statistic, called Generalized SADF (GSADF), is defined as the supremum value of the sequence of $ADF_{r_1}^{r_2}$ statistics expressed as follows:

$$GSADF(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} \left\{ \sup_{r_2 \in [r_0, 1]} ADF_{r_1}^{r_2} \right\}. \quad (5)$$

Under the $I(1)$ null, the limit distribution of the $GSADF(r_0)$ statistic is given by

$$\sup_{r_1 \in [0, r_2 - r_0], r_2 \in [r_0, 1]} \left\{ \frac{\frac{1}{2} r_w [W(r_2)^2 - W(r_1)^2 - r_w] - \int_{r_1}^{r_2} W(r) dr [W(r_2) - W(r_1)]}{r_w^{\frac{1}{2}} \left\{ r_w \int_{r_1}^{r_2} W(r)^2 dr - \left[\int_{r_1}^{r_2} W(r) dr \right]^2 \right\}^{\frac{1}{2}}} \right\}. \text{ Whenever } GSADF(r_0)$$

exceeds the corresponding right-tailed critical value from its limit distribution, the unit root hypothesis is rejected in favor of mildly explosive behavior. The rolling-window structure of the $GSADF(r_0)$ test leads to improved power in detecting recurring episodes of mildly explosive behavior relative to what can be achieved with the standard ADF_0^1 and the $SADF(r_0)$ tests.

2.4 Panel GSADF Test

Pavlidis *et al.* (2016) developed an extension of the GSADF test procedure for heterogeneous panels based on the panel data techniques developed by Im *et al.* (2003). Consider the panel version of the ADF regression equation in (3):

$$\Delta y_t^s = \alpha_{r_1, r_2}^s + \beta_{r_1, r_2}^s y_{t-1}^s + \sum_{j=1}^k \psi_{r_1, r_2}^{s, j} \Delta y_{t-j}^s + \epsilon_t^s, \quad \epsilon_t^s \sim i.i.d. (0, \hat{\sigma}_{r_1, r_2}^2), \quad (6)$$

where $s = 1, \dots, N$ indexes the N constituent time series to be tested $\{y_t^s\}_{s=1}^N$, and the aggregated series y_t is a linear combination of those disaggregated series. The procedure then tests the null hypothesis of a unit root in all N disaggregated series $\{y_t^s\}_{s=1}^N$, $H_0: \beta_{r_1, r_2}^s = 0, \forall s$, against the alternative of mildly explosive behavior, $H_1: \beta_{r_1, r_2}^s > 0$ for at least one s .

Let $ADF_{r_1}^{s, r_2} = \frac{\hat{\beta}_{r_1, r_2}^s}{s.e.(\hat{\beta}_{r_1, r_2}^s)}$ denote the test statistic for a given series s over the subsample

beginning in r_1 and ending in r_2 . Over the sequence of $ADF_{r_1}^{s, r_2}$ for all possible r_1 and r_2 , we

define the Backward SADF (BSADF) as $BSADF_{r_2}^s(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} \left\{ \sup_{r_2 \in [r_0, 1]} ADF_{r_1}^{s, r_2} \right\}$. The panel BSADF statistic can now be defined by taking the average at each time period of the BSADF statistics of each constituent series s as follows: $Panel\ BGSADF_{r_2}(r_0) = \frac{1}{N} \sum_{s=1}^N BSADF_{r_2}^s(r_0)$. Then, the definition of the panel GSADF test statistic follows naturally as the supremum of the panel BSADF, i.e.,

$$Panel\ GSADF(r_0) = \sup_{r_2 \in [r_0, 1]} \{Panel\ BGSADF_{r_2}(r_0)\}. \quad (7)$$

The distribution of the panel unit root test based on mean unit root statistics is not invariant to cross-sectional dependence of the error terms. To deal with that, the procedure of Pavlidis *et al.* (2016) adopts a sieve bootstrap approach that is designed specifically to allow for cross-sectional error dependence. If the panel GSADF statistic is larger than the corresponding critical value, we reject the null hypothesis of a unit root in favor of mildly explosive behavior.

The specification of the alternative hypothesis proposed by Pavlidis *et al.* (2016) allows for the coefficients β_{r_1, r_2}^s to vary across each of the constituent time series and, in that sense, is a more flexible alternative for testing than the homogenous alternative hypothesis. Hence, the panel GSADF test specification better captures cases in which the aggregated data inherits its explosiveness from some (but not all) of its constituent series. Furthermore, while the univariate GSADF test on aggregated data imposes a linear restriction on the H_0 and H_1 hypothesis being tested, the panel GSADF approach tests the null that all constituent series are I(1) directly against the alternative that at least one displays mildly explosive behavior.⁵ Not

⁵ As an illustration of the specification of the panel and univariate GSADF testing hypothesis, consider a simple arithmetic mean aggregator where the N disaggregated series are $\{y_t^s\}_{s=1}^N$. Therefore, the aggregated series is given by $y_t = \frac{1}{N} \sum_{s=1}^N y_t^s$. The null hypothesis of a unit root in the aggregate series y_t for a given subsample can be expressed as $H_0: \beta_{r_1, r_2} = \frac{1}{N} \sum_{s=1}^N \beta_{r_1, r_2}^s = 0$, while the null hypothesis of a unit root in all disaggregated series $\{y_t^s\}_{s=1}^N$ imposes the stricter requirement that $H_0: \beta_{r_1, r_2}^s = 0, \forall s$ for the panel test. In turn, the alternative of mildly explosive behavior in y_t for any subsample is expressed as $H_1: \beta_{r_1, r_2} = \frac{1}{N} \sum_{s=1}^N \beta_{r_1, r_2}^s > 0$ while the alternative tested on disaggregated data $\{y_t^s\}_{s=1}^N$ simply requires that $H_1: \beta_{r_1, r_2}^s > 0$ for some (but not necessarily all) of the constituent series s . Therefore, if we reject the null with the panel GSADF test when it is false, then it must be the case that $\beta_{r_1, r_2} = \frac{1}{N} \sum_{s=1}^N \beta_{r_1, r_2}^s > 0$ is true and accordingly we should reject the null with the univariate GSADF test on aggregated data as well. In comparing the univariate and panel GSADF tests in finite samples, however, we often find that the ability or power of the panel test to detect mildly explosive behavior if such behavior truly occurs in-sample (rejecting the null when it is incorrect) tends to be higher than that of the univariate GSADF test.

surprisingly, due to the additional flexibility of the framework to incorporate all relevant cross-sectional information (not just the time series information) for testing, we find that such a panel statistic—when data is available to implement it—performs better in terms of power than a simple univariate GSADF test on the series that aggregates all constituent series.

2.5 Univariate SADF/GSADF and Panel GSADF: Implementation

The computation of the SADF, GSADF, and panel GSADF test statistics requires the selection of the minimum window size r_0 and the maximum autoregressive lag length k . Regarding the minimum window size, we follow the rule of thumb of Phillips *et al.* (2015a,b), $r_0 = 0.01 + \frac{1.8}{\sqrt{T}}$ where T is the total number of observations. With respect to the autoregressive lag length, this is known for our exercise with simulated data. Accordingly, we set k equal to its true value of zero. For the other two simulation experiments that use actual house price data, the true lag length is unknown and we evaluate our results for the case where k is equal to four. Our findings do not appear very sensitive to the lag length specification, though.

The implementation of the univariate test procedures also requires the limit distributions of the SADF, GSADF, and panel GSADF test statistics. These distributions are non-standard and depend on the minimum window size r_0 . For the former two tests, finite-sample critical values are obtained through Monte Carlo simulations by generating 2000 replications of a driftless random walk process with $N(0,1)$ errors. For the panel test, we set the number of bootstrap replications to 1000.

2.6 Power Losses Due to Aggregation

Time series that display boom-bust episodes, like real house prices, display two important properties: First, they are nonlinear (because they burst) and, second, they are explosive during their boom phase. Blanchard and Watson (1982), Diba and Grossman (1988a; 1988b), and Evans (1991), among others, provide theoretical DGPs consistent with the standard asset

pricing equation (in the presence of rational bubbles) that display such nonlinear and explosive dynamics.

Regarding the first property, the effects of aggregating nonlinear processes (or linear with nonlinear processes) are, in general, unknown, as discussed in Granger and Lee (1999). This implies that, even if we knew the true DGP for each individual constituent series, we may still not be able to infer the exact form and features for the aggregate index constructed from those constituent series.

We can, nevertheless, draw inferences about the integration properties of an aggregated series by exploiting the second property of periodically-collapsing episodes of exuberance (i.e., mildly explosive dynamics). It is well known that the combination of explosive processes with other explosive, unit root, and/or stationary processes generally results in an explosive process.⁶ This point is particularly important for unit root testing procedures because it implies that as long as one of the constituent series is explosive, so will be the aggregated series.

However, we argue that aggregation affects the performance of the recursive (right-tailed) unit root tests to detect mildly explosive behavior. We often find that the empirical evidence to detect explosiveness is a lot stronger if we can test it directly on the affected constituent series than if one has to make inferences on the basis of an observed aggregate series that mixes it with other constituent series that may not display similar patterns of explosiveness.

We evaluate the performance of the recursive (right-tailed) unit root tests discussed in this section on the basis of their statistical power. The power of a test is defined by the probability with which the test correctly rejects the null hypothesis (H_0) when the alternative hypothesis (H_1) is true—in our case, that determines the ability of the test to detect deviations from unit root behavior whenever the data actually displays mildly explosive behavior. As the power of

⁶ In theory, the combination of explosive processes can be stationary or $I(1)$ if the processes under consideration are co-explosive (see, e.g., Nielsen, 2010). Therefore, the effect of an upward explosive period in one of the constituent series could offset the effect of a simultaneous downward explosive period in another constituent series of the aggregate. Although theoretically possible, this scenario seems unlikely from a practical point of view in most conventional applications (a knife-edge case).

the test increases, the probability of a Type II error (a false negative) declines under the alternative hypothesis.

Along this metric, our simulation findings reveal that the power of the recursive (right-tailed) unit root ADF test improves when we exploit the recursive nature of the SADF specification of Phillips and Yu (2011) and especially that of the GSADF specification of Phillips *et al.* (2015a,b). Simply put, the power loss due to aggregation worsens the already low power problem of the standard (right-tailed) unit root test. The recursive nature of the SADF and GSADF tests improves the power and, accordingly, gives these tests a better chance at detecting exuberance in the aggregated series. The panel GSADF test applied to disaggregated data introduces a much richer specification that captures the heterogeneity and cross-sectional dependencies of the constituent series and further improves the power of the GSADF approach.⁷

To understand this, we recognize that the statistical power of the (right-tailed) unit root ADF test depends on a number of factors.⁸ First, we note that the recursive implementation of the SADF test and, particularly, that of the GSADF test keep the full sample size invariant but lead to higher power. The recursive SADF and GSADF procedures lift the restriction of testing over the full sample and instead define their statistics in terms of the supremum of an ADF statistic sequence over many subsamples (including the full sample itself). Intuitively, the probability of rejecting the null when it is false (the power) should not be lower with the SADF/GSADF tests than the power of the standard (right-tailed) ADF test over the full sample as the performance of the standard (right-tailed) ADF test can always be achieved within the given sequence.

Second, the magnitude of the deviation from unit root to mildly explosive behavior present in some of the disaggregated series can become diluted due to aggregation.⁹ In turn, the smaller

⁷ However, the practical implementation of panel testing is not always possible when aggregates exist but we lack the actual disaggregated data.

⁸ The statistical significance is defined in terms of the probability of a Type I error (α)—which is the probability of rejecting the null when the null is correct. Naturally, another factor that influences the power of a test is its statistical significance which is customarily set at $\alpha = 5\%$. However, while increasing α lowers the probability of a Type II error (β) and therefore increases the power of the test, it also means accepting a higher risk of rejecting the null when the null is true (Type I error).

⁹ Even when the processes for the disaggregated series under consideration are not co-explosive to begin with (see, e.g., Nielsen, 2010).

effect on the aggregate series lowers the finite-sample power for the standard and for any of the recursive (right-tailed) unit root ADF tests studied in the paper. In order to illustrate this point, consider a simple arithmetic mean aggregator where the N disaggregated series are $\{y_t^s\}_{s=1}^N$. Therefore, the aggregated series is given by $y_t = \frac{1}{N} \sum_{s=1}^N y_t^s$. Assume the following simple DGP for the constituent series (with only one of them displaying mildly explosive behavior):

$$\Delta y_t^s = \epsilon_t^s, \forall s = 1, \dots, N-1, \quad (8)$$

$$\Delta y_t^N = \beta y_{t-1}^N + \epsilon_t^N,$$

where $\epsilon_t^s \sim i.i.d. (0, \sigma^{s,2})$ for all $s = 1, \dots, N$ with possibly cross-sectional dependence in the error terms, and where $\beta = \delta - 1$ is the coefficient to be tested. Hence, in this scenario, we see that the dynamics for the aggregate series can be expressed as:

$$\Delta y_t = \frac{\beta}{N} \left(\frac{y_{t-1}^N}{y_{t-1}^s} \right) y_{t-1} + \epsilon_t = \left(\frac{y_{t-1}^N}{\sum_{s=1}^N y_{t-1}^s} \right) \beta y_{t-1} + \epsilon_t, \quad (9)$$

where $\epsilon_t = \frac{1}{N} \sum_{s=1}^N \epsilon_t^s$ is the average error (whose distribution depends on the assumptions imposed on the errors for each series s).

With this example we show that the aggregated series inherits the mildly explosive behavior of the constituent series $s = N$, but we also see that the deviation from unit root behavior is different in the aggregated series. To be more precise, we observe that the relative contribution of the constituent series $s = N$ to the aggregate series weighs down the magnitude of the effect given by the coefficient $\beta > 0$. Intuitively, the smaller effect incorporated into the aggregated series tends to lower the statistical power of the recursive ADF tests in finite samples.

In this sense, we could achieve better performance by testing for mildly explosive behavior directly on the disaggregated series $s = N$ or over the entire cross-section (along the lines of what the panel GSADF of Pavlidis *et al.* 2016 does) than testing for it indirectly through the aggregated series. Needless to say, this is not an option when the disaggregated series are not

available for testing. Then, the empirical question becomes to quantify the extent of the power loss due to aggregation and the mitigation effects that can be obtained with the recursive SADF and GSADF procedures. The simulation exercises that we discuss in the remainder of the paper aim to answer formally that question.

3. A Monte Carlo Examination of the Power of Recursive Unit Root Tests

What is the effect of aggregation on the power of the recursive SADF and GSADF tests in finite samples? To shed light on this issue, we run three sets of Monte Carlo simulation experiments. The first set of experiments is based on a family of time series processes examined by Phillips *et al.* (2015a,b). The second and third sets are based on actual real house price data.

3.1 Illustration with Artificial Data

Phillips *et al.* (2015a) consider two DGPs that deterministically switch between martingale and explosive behavior. The first DGP is given by:

$$y_t = y_{t-1} 1\{t < \tau_e\} + \delta_T y_{t-1} 1\{\tau_e \leq t \leq \tau_f\} + \left(\sum_{k=\tau_f+1}^t \epsilon_k + y_{\tau_f}^* \right) 1\{t > \tau_f\} + \epsilon_t 1\{t \leq \tau_f\}. \quad (10)$$

This process exhibits martingale behavior in the first part of the sample, it then enters a mildly explosive phase during which it grows exponentially and, finally, it collapses and resumes martingale behavior until the end of the sample (T). The origination and termination dates of the explosive episode are denoted by τ_e and τ_f , respectively. The autoregressive coefficient $\delta_T = 1 + cT^{-\alpha}$ with $c > 0$ and $\alpha \in (0,1)$ determines the rate of expansion, $\epsilon_t \sim N(0, \sigma^2)$, and $y_{\tau_f}^* = y_{\tau_e} + y^*$ equals y_{τ_e} plus a small perturbation $y^* = O_p(1)$.

The second DGP considered by Phillips *et al.* (2015a,b) constitutes an extension of the previous one that allows for two episodes of mildly explosive behavior:

$$y_t = y_{t-1}1\{t \in T_0\} + \delta_T y_{t-1}1\{t \in B_1 \cup B_2\} + \left(\sum_{k=\tau_{1f}+1}^t \epsilon_k + y_{\tau_{1f}}^* \right) 1\{t \in T_1\} + \left(\sum_{l=\tau_{2f}+1}^t \epsilon_l + y_{\tau_{2f}}^* \right) 1\{t \in T_2\} + \epsilon_t 1\{t \in T_0 \cup B_1 \cup B_2\}, \quad (11)$$

where $T_0 = [1, \tau_{1e})$, $B_1 = [\tau_{1e}, \tau_{1f}]$, $T_1 = (\tau_{1f}, \tau_{2e})$, $B_2 = [\tau_{2e}, \tau_{2f}]$, and $T_2 = (\tau_{2f}, T]$. The observations τ_{1e} and τ_{1f} mark the origination and termination dates of the first episode of exuberance, while τ_{2e} and τ_{2f} are the origination and termination dates of the second exuberance episode to occur in-sample. After the collapse of the first period of explosiveness, y_t resumes a martingale path until $\tau_{2e} - 1$. Then, a second episode of exuberance begins at τ_{2e} which lasts until it collapses at time τ_{2f} . The process y_t then continues on a martingale path until the end of the sample period.

In this section, we examine the power of the recursive (right-tailed) unit root tests whenever we aggregate the above processes—the one exuberance episode case in (10) and the two exuberance episodes case in (11)—with pure random walk processes. Given the computationally intensive nature of the panel GSADF test, in the simulation experiments we restrict the cross-sectional dimension to five series, and we allow the number of processes that display explosive dynamics to increase from $s = 1, \dots, 5$. With regard to parameter values, we set $\delta_T = 1.05$, $\sigma^2 = 0.2$, $T = 100$, and we assume that the cross-correlation of errors equals to 0.25. The number of Monte Carlo simulations is 1000 and the number of bootstrap replications for the panel GSADF is also 1000. For the one-episode-of-exuberance case, we set the origination date to $\tau_e = 50$ and the termination date to $\tau_f = \{60, 70, 80\}$. For the two-episodes-of-exuberance case, we consider a ‘short’ first episode that originates at $\tau_{1e} = 30$ and terminates at $\tau_{1f} = 40$ and a ‘long’ second episode that originates at $\tau_{2e} = 60$ and terminates at $\tau_{2f} = \{80, 90\}$.

Tables 1 and 2 report the power of the SADF, GSADF, and panel GSADF tests in the presence of at least one series displaying one- and two-episodes of explosiveness, respectively. Overall, we observe that the power of all tests increases with the number of series that display exuberance, the duration of the exuberance episode(s), and the number of episodes. The latter two findings

are not surprising and are in line with those of previous studies (Phillips *et al.*, 2015a,b; Pavlidis *et al.*, 2016). The former finding is new and highlights the impact of aggregation on the power of the various tests. Most notably, for a single, short episode of exuberance, the power of the tests varies from around 5 percent, when only one out of the five processes displays exuberance and the rest are pure random walks, to around 50 percent, when all processes display exuberance.

In terms of ranking, we observe that the panel GSADF is superior to both the SADF and its univariate counterpart, and that the differences in power across tests can be substantial. For instance, when only one series displays a single, long episode of exuberance, the power of the SADF and GSADF tests is less than 35 percent, while the power of the panel GSADF exceeds 52 percent. Similarly, when two out of the five processes display an episode of exuberance from $t = 30$ to 40 and a second episode from $t = 60$ to 80, the SADF and GSADF tests can detect exuberance in the aggregate series 30 and 39 percent of the cases, respectively. In turn, the panel GSADF test rejects the null hypothesis of no exuberance in any of the constituent series 54 percent of the time. Thus, the panel GSADF test can result in substantial power gains in finite samples. A comparison of the results for the two univariate tests does not reveal significant differences in the presence of a single episode of exuberance. However, as expected given its known consistency in the presence of multiple boom-bust episodes, the GSADF is superior to the SADF procedure when there are two episodes of exuberance.

Table 1. Artificial Data with One Episode of Exuberance Simulation Results: Power of the Tests

(τ_e, τ_f)	Test	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$
(50,60)	SADF	0.061	0.114	0.226	0.320	0.476
	GSADF	0.057	0.110	0.238	0.339	0.527
	Panel GSADF	0.063	0.161	0.323	0.427	0.569
(50,70)	SADF	0.151	0.402	0.629	0.735	0.841
	GSADF	0.144	0.405	0.641	0.768	0.870
	Panel GSADF	0.252	0.571	0.779	0.873	0.940
(50,80)	SADF	0.347	0.725	0.855	0.911	0.941
	GSADF	0.312	0.736	0.875	0.927	0.953
	Panel GSADF	0.525	0.879	0.946	0.980	0.990

Notes: The table reports the power of the SADF, GSADF, and Panel GSADF tests. s denotes the number of processes that display exuberance, and τ_e and τ_f denote the origination and termination dates of the episode of exuberance.

Table 2. Artificial Data with Two Episodes of Exuberance Simulation Results: Power of the Tests

(τ_{2e}, τ_{2f})	Test	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$
(60,80)	SADF	0.130	0.302	0.545	0.683	0.794
	GSADF	0.155	0.388	0.683	0.816	0.883
	Panel GSADF	0.238	0.540	0.766	0.861	0.915
(60,90)	SADF	0.317	0.653	0.794	0.872	0.922
	GSADF	0.344	0.709	0.871	0.927	0.959
	Panel GSADF	0.515	0.833	0.933	0.971	0.989

Notes: The table reports the power of the SADF, GSADF, and Panel GSADF tests. s denotes the number of processes that display exuberance, and τ_e and τ_f denote the origination and termination dates of the second episode of exuberance.

3.2 Illustration with House Price Data

Given the wide use of aggregated indices, the issue of aggregation is a very important topic in financial econometrics with applications to the monitoring of asset markets (not just housing markets). It is appropriate to note here that our findings are generally valid to applications based on any asset price index (or, for that matter, any time series) where the issue of aggregation may be pertinent in testing the occurrence of episodes of mildly explosive behavior. The application to housing market data is particularly relevant for illustration purposes because housing is notoriously heterogeneous across many characteristics (including their location). It is also an interesting application in itself because of the significance of the

housing cycle and its broad impact on real economic activity—the widespread boom-bust housing cycle that preceded the 2008-09 global recession being a prime example of that.

Our empirical investigation is then based on two real house price datasets that showcase at least two levels of aggregation across heterogeneous locations. First, we look at the construction of national house price indices that aggregate price data across a number of heterogeneous major statistical metropolitan areas of the U.S. Second, we also look at the construction of international real house price indices aggregating representative national indices. Our analysis of the disaggregated housing data suggests that there is very strong evidence of exuberance within sample in both datasets. We find that by exploiting disaggregated data, we can more successfully detect the occurrence of such episodes of exuberance.

The two simulation experiments based on actual data on real house prices resemble conventional applications of the recursive (right-tailed) tests to monitor housing markets. Specifically, both experiments under consideration consist of the following steps:

Step 1a. Fit sequentially to the corresponding real house price index the Augmented Dickey-Fuller (ADF) regression equation:

$$\Delta y_t = a_{r_1, r_2} + \beta_{r_1, r_2} y_{t-1} + \sum_{j=1}^k \psi_{r_1, r_2}^j \Delta y_{t-j} + \epsilon_t, \quad \epsilon_t \sim i.i.d. (0, \sigma_{r_1, r_2}^2), \quad (12)$$

where r_1 and r_2 denote fractions of the total sample size that specify the starting and ending points of a subsample and satisfy the rule proposed by Phillips *et al.* (2015a,b), $r_2 - r_1 \geq 0.01 + 1.8/\sqrt{T}$, and k is the maximum number of lags. Obtain, for each time period t , the vector of estimated coefficients $(\hat{a}_{r_1, r_2}, \hat{\beta}_{r_1, r_2}, \hat{\psi}_{r_1, r_2}^j, j = 1, \dots, k)$ and the estimated variance of residuals $(\hat{\sigma}_{r_1, r_2}^2)$ that correspond to the BSADF statistic.¹⁰

¹⁰ For further details on the BSADF, SADF, and GSADF statistics, see Phillips *et al.* (2015a,b) and Pavlidis *et al.* (2016).

Step 1b. Use the matrix of estimated coefficients and draw randomly from a normal distribution with mean zero and variance equal to the estimated error variance at each time period t to generate artificial series from:

$$\Delta y_t^S = \hat{a}_{r_1, r_2} + \hat{\beta}_{r_1, r_2} y_{t-1}^S + \sum_{j=1}^k \hat{\psi}_{r_1, r_2}^j \Delta y_{t-j}^S + \epsilon_t^S, \quad \epsilon_t^S \sim N(0, \hat{\sigma}_{r_1, r_2}^2). \quad (13)$$

Note that, due to its sequential nature and the flexible window size, this procedure allows for very rich dynamics in the simulated constituent series that closely resemble the dynamics of actual asset prices.

Step 2. Run the SADF, GSADF, and Panel GSADF tests on the artificial constituent series and obtain the corresponding test statistics.

Step 3. Run the SADF and GSADF tests on the aggregate series (for simplicity, the aggregate is set equal to the average of the constituent simulated series) and obtain the corresponding test statistics.

Step 4. Repeat steps (1b) to (3) a thousand times and compute the power of each test as the number of times the value of the test statistic is greater than the 95% critical value.

3.2.1 Aggregation from Local to National Real House Prices

We first apply the above procedure to the U.S. housing market and its major metropolitan statistical areas. For this exercise, we use data on monthly house prices for the S&P/Case-Shiller 10-City Composite Home Price Index and its constituents (corresponding to the 10 largest metropolitan areas in the U.S.), deflated by its corresponding local CPI. The sample covers the period from January 1987 to February 2016, giving a total of 350 observations. The data sources are S&P Dow Jones Indices and the Bureau of Labor Statistics. The data is illustrated in Chart 1.

Chart 1. U.S. Real House Prices by Major MSA

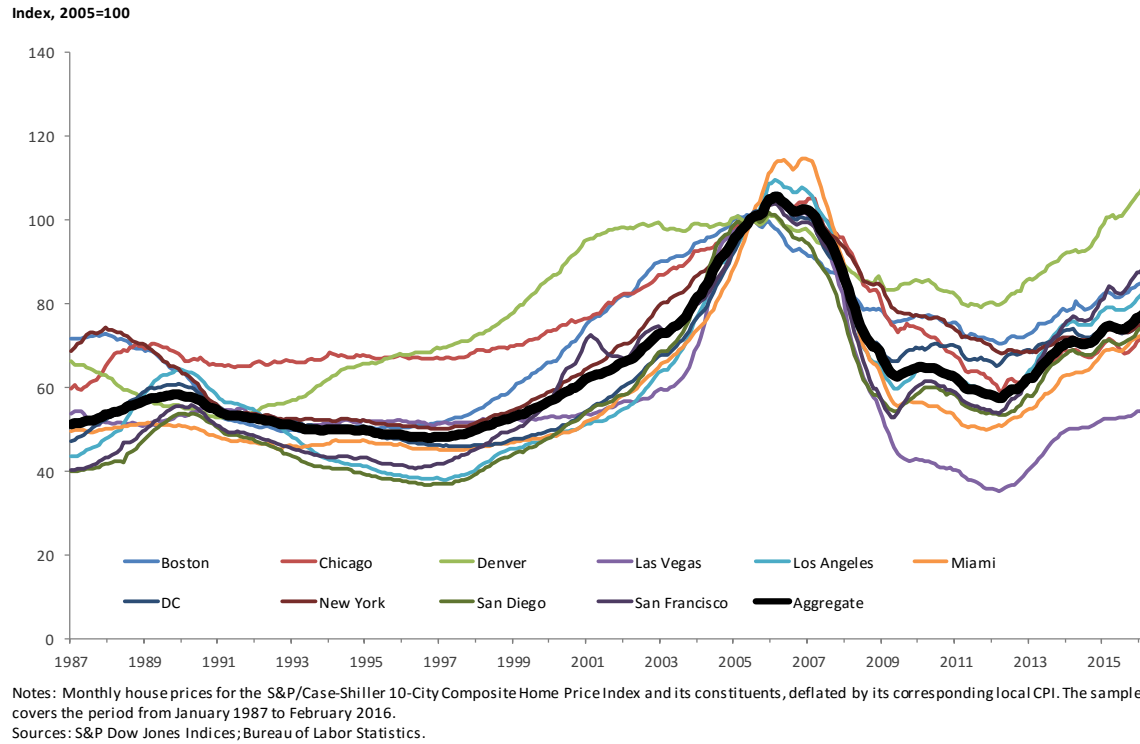


Table 3 reports the SADF and GSADF statistics for the S&P/Case-Shiller 10-City Composite Home Price Index and its constituents, as well as the finite-sample critical values. According to the SADF statistic results, the null hypothesis of a unit root can be rejected at the 5% significance level for all metropolitan areas except for Los Angeles (LA). Thus, 9 out of the 10 metropolitan areas appear to exhibit exuberance during (at least) part of the sample period under consideration. Using the GSADF test, all metropolitan areas show evidence of exuberance at the 1% significance level. Moreover, we observe that the estimated test statistics for the aggregated series fall inside the range of values for the constituent metropolitan area series.

In accordance with our evidence, the simulation results reported in Chart 2 show that the power of the SADF and GSADF tests for the aggregate series lies between the lowest and highest power for the constituent metropolitan area series. This finding shows how aggregation leads to power losses in practice, since if there was no power loss then every time a test detected explosive dynamics in a constituent metropolitan area, it should also detect explosive dynamics in the aggregate. A striking contrast comes from looking at Boston for which the

power of the SADF is 97 percent or Miami for which the SADF power is 96 percent, while the SADF power for the 10-city aggregate that includes them is merely 16 percent. By comparing the results for the two tests, the SADF and the GSADF tests, we observe that the GSADF test performs remarkably better than the SADF with very small power losses due to aggregation. In this application, the power achieved is somewhat lower but very close to that of the Panel GSADF test based on the disaggregated data.

We summarize the geographical dispersion of the power gains achieved by the GSADF test over the SADF test for each one of the metropolitan area series in the map in Chart 3. Comparing the results for the two tests—the SADF and the GSADF tests—we observe that the GSADF test performs much better than the SADF but with significant variation across metropolitan areas. For Boston and Miami the power gain of the GSADF test is under 20 percentage points (0.2 in units) while the gain exceeds 80 percentage points (0.8 in units) for Los Angeles (LA).

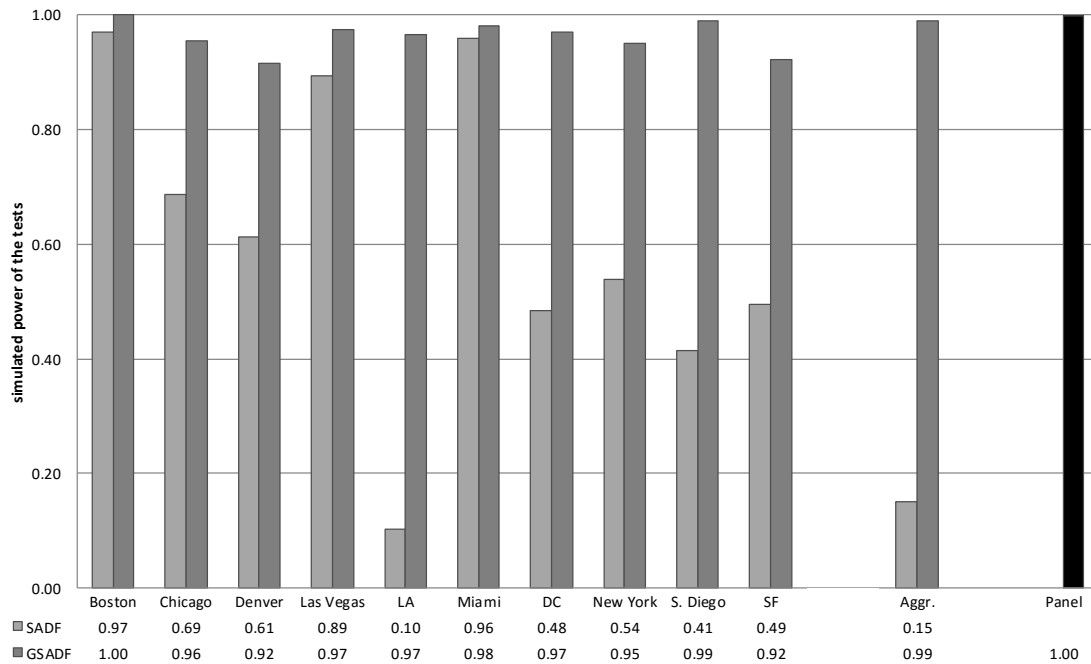
Table 3. S&P/Case-Shiller 10-City Composite Home Price Index, Unit Root Test Results (Actual Data)

Panel A: Test Statistics	SADF	GSADF	Panel GSADF
Boston	2.264***	5.253***	
Chicago	2.719***	4.946***	
Denver	1.748**	7.372***	
Las Vegas	4.791**	5.321***	
Los Angeles	1.098	3.803***	
Miami	4.703***	5.987***	
Washington	1.793**	4.329***	
New York	1.520**	4.742***	
San Diego	1.959***	3.740***	
San Francisco	2.493***	3.947***	
U.S. Composite 10	2.772***	4.095***	
Full Panel of 10 Cities			3.295***
Panel B: Critical Values	SADF	GSADF	Panel GSADF
90%	1.145	1.911	1.800
95%	1.433	2.167	2.025
99%	1.942	2.712	2.431

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent significance levels respectively. All results are for autoregressive lag length $k=4$. The aggregate series is computed as an unweighted average of the real house price indices of all the metropolitan areas.

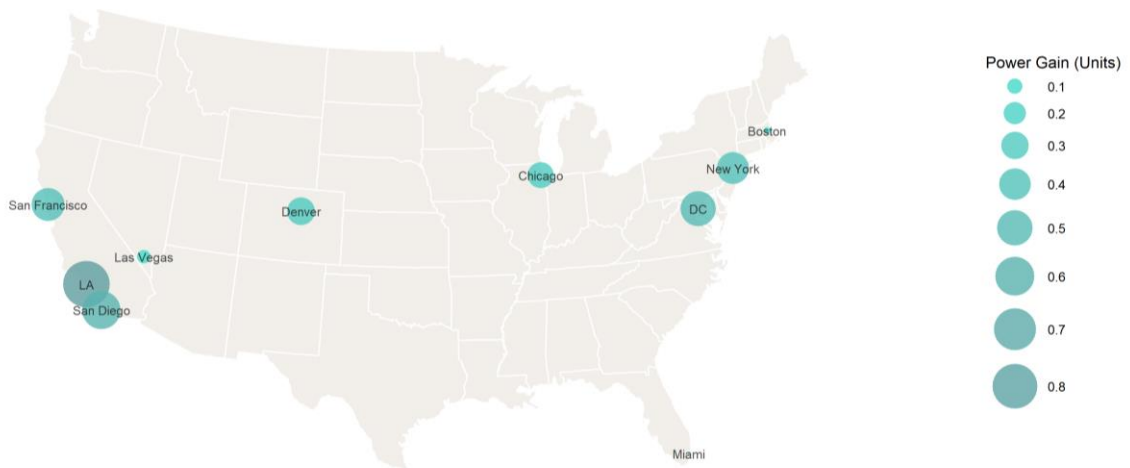
Sources: Standard & Poor's, Bureau of Labor Statistics, authors' calculations.

**Chart 2. S&P/Case-Shiller 10-City Composite Home Price Index
Simulation Results: Power of the Tests**



Notes: The figure and table display the power of the SADF and GSADF tests. The nominal significance level is set at 5%.
Sources: Standard & Poor's, Bureau of Labor Statistics, authors' calculations.

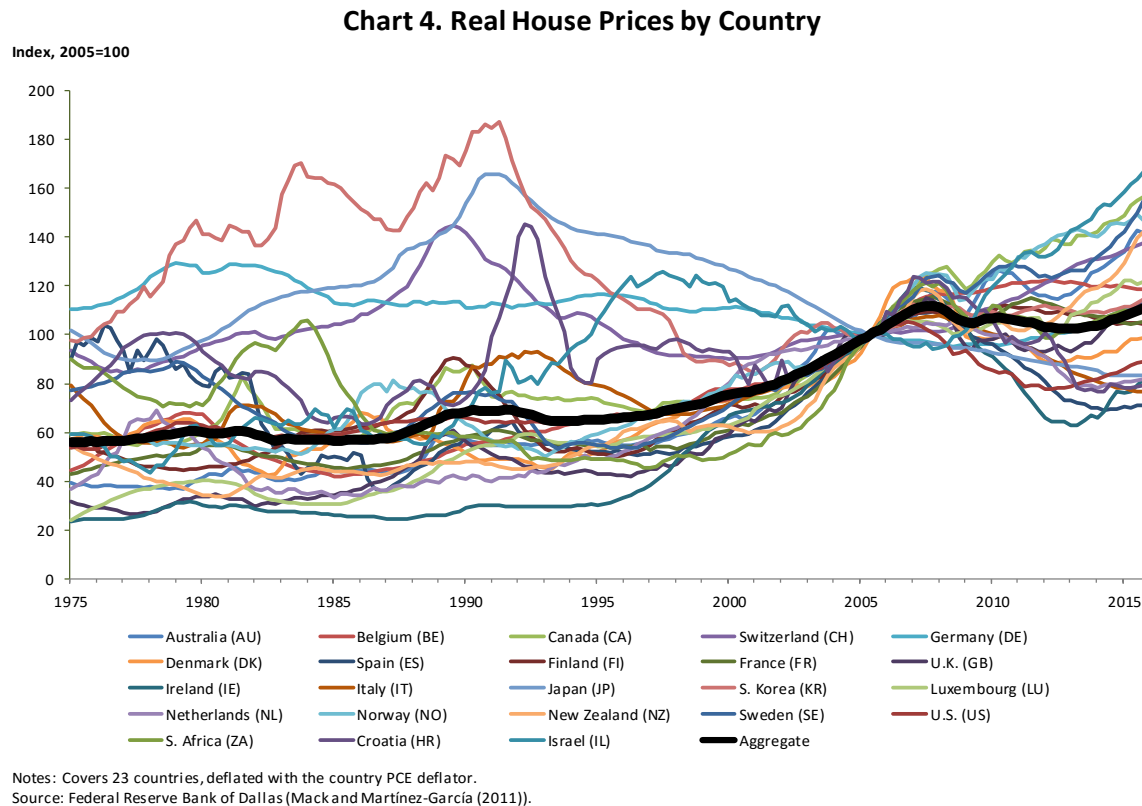
**Chart 3. S&P/Case-Shiller 10-City Composite Home Price Index
Simulation Results: Geographical Variation of Power Gains (GSADF vs. SADF Tests)**



Notes: The map displays the power gain of the GSADF test over the SADF test by location. The nominal significance level is set at 5%.
Sources: Standard & Poor's, Bureau of Labor Statistics, authors' calculations.

3.2.2 Aggregation from National to International Real House Prices

To provide further evidence on the effect of aggregation on the performance of both recursive (right-tailed) unit root tests, in our second experiment, we repeat the same Monte Carlo simulation exercise using the cross-country real house price data from the Federal Reserve Bank of Dallas' International House Price Database (Mack and Martínez-García, 2011). The database comes at quarterly frequency and covers 23 (mostly advanced) countries between the first quarter of 1975 and the fourth quarter of 2015, deflated with the corresponding country PCE deflator.¹¹ This data is illustrated graphically in Chart 4.



¹¹ The quarterly national house price indices are those most consistent with the quarterly U.S. house price index for existing single-family houses produced by the Federal Housing Finance Agency. This database can be accessed publicly at: <http://www.dallasfed.org/institute/houseprice/>

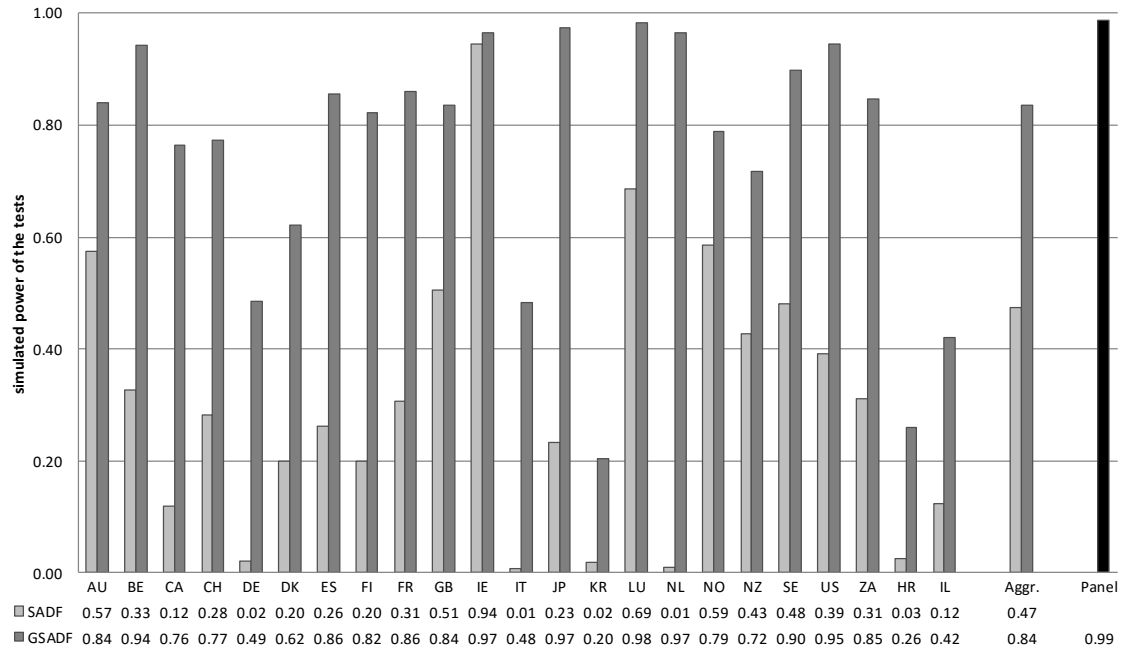
Table 4. Dallas Fed's Real House Prices, Unit Root Test Results (Actual Data)

Panel A: Test Statistics	SADF	GSADF	Panel GSADF
Australia	2.296***	6.110***	
Belgium	0.904	3.450***	
Canada	0.699	4.061***	
Switzerland	1.848**	4.091***	
Germany	-0.595	3.515***	
Denmark	1.279*	3.186***	
Spain	0.885	2.408**	
Finland	1.255*	2.357**	
France	1.065	2.055*	
U.K.	1.629**	3.143***	
Ireland	2.793***	6.781***	
Italy	-1.369	2.800***	
Japan	1.621**	5.013***	
S. Korea	-0.541	-0.130	
Luxembourg	2.383***	5.278***	
Netherlands	-1.140	4.064***	
Norway	1.518**	2.533**	
New Zealand	1.911**	3.051***	
Sweden	1.125*	5.178***	
U.S.	1.686**	4.243***	
S. Africa	-0.474	3.807***	
Croatia	-0.002	2.244**	
Israel	0.936	1.849*	
Aggregate	1.612**	3.327***	
Full Panel of 22 Countries			1.641***
Panel B: Critical Values	SADF	GSADF	Panel GSADF
90%	1.076	1.766	0.678
95%	1.380	2.065	0.749
99%	1.936	2.670	0.942

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent significance levels respectively. All results are for autoregressive lag length k=4. The aggregate series is computed as an unweighted average of the real house price indices of all the countries.

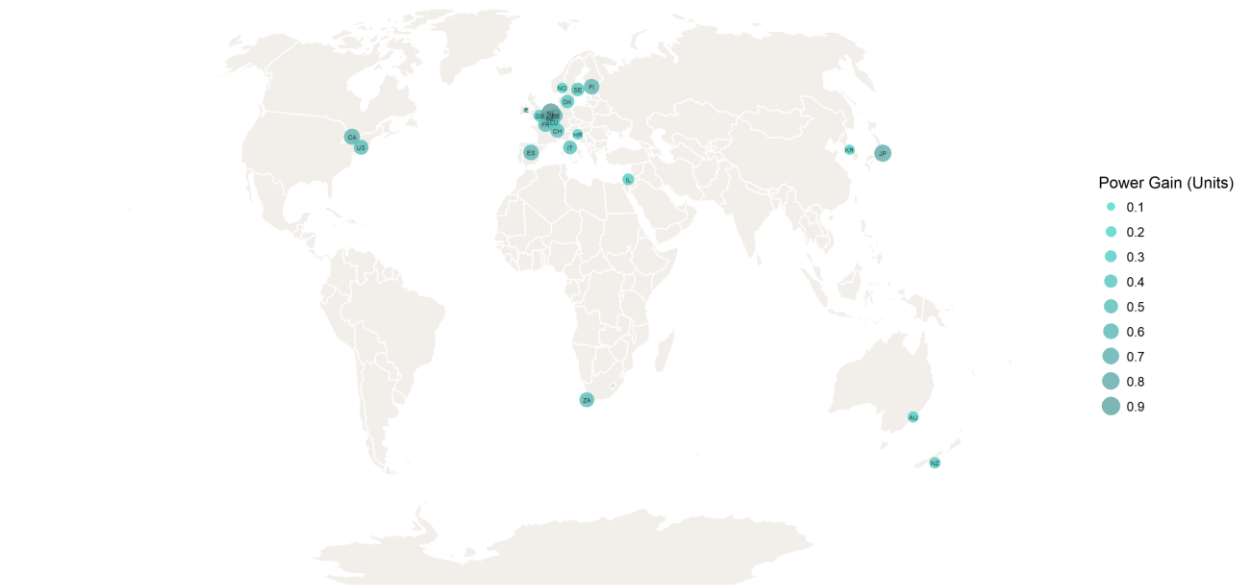
Sources: Federal Reserve Bank of Dallas' International House Price Database and authors' calculations.

Chart 5. Dallas Fed's Real House Price Series
Simulation Results: Power of the Tests



Notes: The figure and table display the power of the univariate SADF and GSADF tests and the panel GSADF. The stat. significance level is set at 5%.
Sources: Federal Reserve Bank of Dallas' International House Price Database and authors' calculations.

Chart 6. Dallas Fed's Real House Price Series
Simulation Results: Geographical Variation of Power Gains (GSADF vs. SADF Tests)



Notes: The map displays the power gain of the GSADF test over the SADF test by location. The nominal significance level is set at 5%.
Sources: Federal Reserve Bank of Dallas' International House Price Database and authors' calculations.

Table 4 reports the SADF and GSADF statistics for the aggregate real house price index and its constituent country indices, as well as the corresponding finite-sample critical values. Chart 5 displays the simulation results. This second set of results shows even more clearly than in the previous exercise with house price data by metropolitan area the severe power loss of the SADF test due to aggregation, and the superior performance of the GSADF test over the SADF test. However, it also illustrates that power losses relative to the panel GSADF test can be significant due to aggregation when having to rely on the univariate GSADF test on the aggregated series.

For our second experiment based on real house prices for the 23 countries covered by the Federal Reserve Bank of Dallas' International House Price Database, we summarize the power gains achieved by the GSADF test over the SADF test for each one of the constituent country series in the map in Chart 6. A striking contrast emerges from looking at Ireland (IE), South Korea (KR), and Norway (NO) for which the power gain of the GSADF test is under 20 percentage points (0.2 in units) while the power gain exceeds 80 percentage points (0.8 in units) for the Netherlands (NL). By comparing the results across tests, we confirm that the GSADF test performs much better than the SADF test with national indexes, but we still find significant variation in the realized gains across locations.

4. Implications of Aggregation for Empirical Inference

To the best of our knowledge, the role that aggregation plays on detecting periods of mildly explosive behavior in time series—let alone the effects of aggregation on the empirical inferences that can be made about boom-bust episodes—have not been examined in the literature. In this paper, we investigate the role of aggregation with simulated data but also with real house price data at different levels of geographical aggregation. We specifically explore aggregation from local real house prices to national real house prices and the aggregation from national real house prices to global real house prices.

The main empirical contribution of our paper is to examine the effect of cross-sectional aggregation on the performance of recursive (right-tailed) unit root tests—the power of the

univariate SADF and GSADF tests developed by Phillips and Yu (2011) and Phillips *et al.* (2015a,b), respectively. Furthermore, we explore the panel GSADF test proposed by Pavlidis *et al.* (2016) to exploit the cross-sectional variation in disaggregated data too. The panel GSADF provides a useful benchmark to quantify the power loss associated with aggregation. Our findings reveal a decline in power of both the SADF and GSADF tests due to aggregation and very large differences in the performance of both tests when applied to aggregated data—favoring the recursive implementation of the GSADF test over than of the SADF test.

We argue that this may warrant increased utilization not only of disaggregated data but also of panel data techniques, such as the panel GSADF proposed by Pavlidis *et al.* (2016), whenever possible. Moreover, our findings show that the GSADF test applied to aggregated data is substantially more powerful than the SADF test. Thus, we conclude that the GSADF test should be preferred to the SADF test whenever the researcher is restricted to using aggregated data for detecting explosive behavior in time series. In this section, we now illustrate the date-stamping and empirical inferences that we draw—using aggregated vs. panel data—under the GSADF methodology.

4.1 Exploring the Predictive Ability of Fundamentals

The implications of rejecting the null under any of the recursive (right-tailed) unit root tests—the univariate GSADF and SADF tests, and the panel GSADF test that exploits the cross-sectional dimension of the disaggregated data—must still be interpreted with caution because they are only indicative, not conclusive of non-fundamental behavior (bubbles). This is because factors other than bubbles can also give rise to explosive dynamics: for instance, in subsection 4.1 we highlight the role of risk-spreads and uncertainty in housing exuberance. This renders the interpretation of the results of these tests inconclusive.

A common approach to empirically backout evidence of non-fundamental behavior in the data is to derive the fundamental price of housing and to test whether deviations between the actual and the fundamentals-based price ($P_t - P_t^*$) display mildly explosive behavior. The

drawback of this approach is that it is heavily dependent on having the correct housing model to describe the fundamental-based price (P_t^*). More generally, omitted variables, measurement error in fundamentals, or simply model misspecification can bias our interpretation of (mildly) explosive behavior tests simply because rejecting the null may indicate the presence of rational bubbles in the data, some form of misspecification, or both (Flood and Garber, 1980; Hamilton and Whiteman, 1985; Gurkaynak, 2008). In other words, such a model-based approach has to contend with the well-known joint-hypothesis testing problem.

We adopt an alternative strategy to investigate what underlies (mildly) explosive dynamics in real house prices. Our identifying assumption does not impose a particular model of housing tying fundamentals to house prices, but instead relies on a generic implication that holds true for a large class of asset pricing models: that is, that (mildly) explosive dynamics that arise from non-fundamental behavior ought to be unpredictable while those that spill over from fundamentals must be by their very own nature predictable.

To implement our strategy, we do as follows: First, we use the date-stamping approach recommended by Pavlidis *et al.* (2016) on the panel and the related one recommended by Phillips *et al.* (2015a,b) on the aggregate series under the GSADF methodology in order to establish a timeline of episodes of exuberance (periods of mildly explosive behavior) in house prices. Second, we employ a dynamic probit model to assess the in-sample predictive ability of well-known housing-market and macro/financial fundamentals. This strategy allows us to evaluate the predictive ability of a given set of fundamentals.¹²

We illustrate this strategy with the panel of 23 countries from the Federal Reserve Bank of Dallas' International House Price Database (Mack and Martínez-García, 2011). The dependent variable of the probit model is a dummy variable EXU_t that takes the value of 1 during an episode of global exuberance in real house prices and 0 otherwise, i.e., $EXU_t =$

¹² With this strategy, we evaluate empirically the contribution—if any—of specific fundamentals to explain the observed patterns of (mildly) explosive behavior. Still, evidence that observed fundamentals are not statistically-significant does not establish that episodes of exuberance in-sample must arise from rational bubbles due to possible omitted variables, measurement error, etc. We only argue that observed fundamentals that are not statistically significant do not appear to explain much of the evidence of (mildly) explosive behavior that we detect and date-stamp in the available data.

$\begin{cases} 0, & \text{if Panel } BSADF_t(r_0) < scv_t^\alpha, \\ 1, & \text{if Panel } BSADF_t(r_0) > scv_t^\alpha, \end{cases}$ where inference about the explosiveness of the process at observation $t = \lfloor r_2 T \rfloor$ is based on comparing the panel Backward Sup ADF (BSADF) statistic (panel $BSADF_t(r_0)$) for that observation against the corresponding $100(1 - \alpha)\%$ critical value of the panel Sup ADF (scv_t^α). The critical value scv_t^α is based on $t = \lfloor r_2 T \rfloor$ observations, a sample size of T observations, and a significance level of α .¹³ Similarly, we also construct an alternative exuberance indicator using the univariate BSADF on the aggregated series.¹⁴

The panel BSADF statistic is particularly appealing because it measures the degree of overall exuberance in international housing markets taking into account the cross-sectional heterogeneity. With this panel approach we detect exuberance in-sample and date its occurrence second quarter 1999 till fourth quarter 2007. Interestingly, the chronology obtained using the univariate BSADF statistic on aggregated real house price data is similar going from second quarter 1999 till second quarter 2007. The dynamic probit $P(EXU_t = 1) = \Phi(EXU_{t-1}\beta_{exu} + x_t'\beta_x)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, includes a dummy for one-quarter lagged exuberance EXU_{t-1} and a vector x_t' of aggregate predictors for the 23 countries for which we have house price data.

We augment a simple dynamic probit model with the vector of observed aggregate fundamentals x_t' to evaluate the extent to which fundamental-based explosiveness occurs.¹⁵ As

¹³ The Backward Sup ADF (BSADF) statistic for country $i=1, \dots, N$, $BSADF_{i,t}(r_0) = \sup_{t=\lfloor r_2 T \rfloor, r_1 \in [0, r_2 - r_0]} ADF_{i,r_1}^{r_2}$, relates to the country i 's GSADF statistic as follows: $GSADF_i(r_0) = \sup_{r_2 \in [r_0, 1]} BSADF_{i,t=\lfloor r_2 T \rfloor}(r_0)$. The panel counterpart of the BSADF given by Panel $BSADF_t(r_0) = \frac{1}{N} \sum_{i=1}^N BSADF_{i,t}(r_0)$ and the corresponding panel GSADF statistics are described in detail in section 2.4. The distributions of these test statistics are non-standard and depend on the minimum window size r_0 . Hence, finite-sample critical values are obtained through Monte Carlo simulations by generating 2000 replications of a driftless random walk process with $N(0,1)$ errors. In the panel case, we adopt a sieve bootstrap approach designed specifically to allow for cross-sectional error dependence in the distribution. Additional information on the bootstrap procedure to derive the critical values can be found in Phillips *et al.* (2015a,b) and Pavlidis *et al.* (2016).

¹⁴ Following the implementation strategy discussed in subsection 2.5, we set α at 95% and the lags in the ADF regression to $k = 4$. The minimum window size r_0 ensure a minimum window of $\lfloor r_0 T \rfloor = 25$ quarterly observations for real house price series with a sample size T of 164 observations. Finally, we also follow the recommendation of Phillips *et al.* (2015a,b) and Pavlidis *et al.* (2016), we set a minimum duration for each episode of exuberance that corresponds to a fraction $\frac{\ln(T)}{T}$ of quarterly series of sample size T —which, in our case, implies a minimum duration of 5 quarters—in order to exclude very short episodes.

¹⁵ We construct aggregates for the covariates as well as for real house prices in our dataset based on a simple arithmetic average (consistent with the empirical exploration in section 3).

suggested by the standard asset pricing model of housing laid out in the Appendix, we investigate the predictive ability of measured policy uncertainty (from PolicyUncertainty.com) and the spread between the long- and short-term interest rates (which proxies the slope of the yield curve and indicates market expectations of future policy rates).

Among the other housing/macro fundamentals in x'_t we include: quarter-over-quarter changes in real personal disposable income per capita from the Federal Reserve Bank of Dallas' International House Price Database (capturing a key demand-side fundamental that is viewed as anchoring the housing market over the long-run); quarter-over-quarter changes in stock market valuations deflated with the CPI (to incorporate an alternative asset class which influences households' financial wealth); and real GDP growth (quarter-over-quarter) (as key indicator of the state of the business cycle).

Furthermore, we also incorporate a number of financial variables in the specification—whenever those are found to be statistically-significant—among a set that includes: Quarter-over-quarter changes in credit to the private sector and in the current account deflated with the CPI as well as the current-account-to-GDP ratio (since an expansion of private credit or capital inflows from abroad can lead to asset price and housing boom and busts); market volatility measured by the Chicago Board Options Exchange' VIX and, to account for commodity prices, oil prices based on the West Texas Intermediate deflated with U.S. headline CPI.¹⁶

4.2 Empirical Findings

To specify the benchmark panel dynamic probit model, we start with a general specification including all statistically-significant explanatory variables and the exuberance indicator under the panel BSADF approach as the dependent variable (Table 5, Column M1). This model is reduced by removing the financial variables (Table 5, Column M2) and all explanatory variables other than lagged exuberance, policy uncertainty and the spread (Table 5, Column M3).

¹⁶ Due to data availability, we proxy market volatility (VIX) as well as policy uncertainty from PolicyUncertainty.com with the available U.S. series for each which allows us to cover from first quarter 1986 till fourth quarter 2015 in our subsequent probit exercise.

Similarly, we estimate a probit model for the exuberance indicator under the univariate BSADF approach on the aggregated real house price series including all the relevant covariates (Column M4), the covariates excluding the statistically-significant financial variables in the specification (Column M5), and lagged exuberance, policy uncertainty, and the spread alone (Column M6) in Table 5. We report all the results with a constant in the specification.

These results largely support the view that policy uncertainty and the spread of the yield curve appear to be associated with global exuberance in-sample, but with opposite sign (as suggested by theory). Based on these findings, the higher spreads across countries tend to be associated with a higher likelihood of global exuberance while policy uncertainty can tame the emergence of such an episode of widespread explosiveness in international housing markets. We observe that apart from lagged exuberance, policy uncertainty and the spread, key explanatory variables include standard macro/housing fundamentals such as the real personal disposable income per capita and stock market valuation gains (which play a substantial role in mapping housing demand) as well real GDP growth (that captures the aggregate business cycle).

Furthermore, financial variables also appear to play a role in the emergence of global exuberance. Here, we show that our empirical inferences depend on the degree of aggregation applied to real house prices in order to derivate the dummy variable EXU_t . The role of market volatility and real credit growth in the private sector appear most significant under the chronology of exuberance obtained from aggregated real house prices while the extent of foreign financing proxied by the current account over GDP ratio is the dominant financial force whenever the chronology of exuberance is set under our preferred panel technique.

Both univariate and panel BSADF methodologies give similar chronologies and both are consistent with theory in pointing out that spreads and policy uncertainty matter. However, empirical inferences can still differ even when the resulting chronologies are close like in this exercise. As noted here, the mechanism that is highlighted by the evidence is somewhat different depending on the degree of aggregation of the underlying house price data (specially on the role of financial variables).

Table 5. Estimation Results for the Dynamic Probit Model

	Panel BSADF			Univariate BSADF		
	M1	M2	M3	M4	M5	M6
Lagged exuberance	40.56***	26.84***	7.11***	21.83***	4.85***	4.32***
Policy Uncertainty	-0.24***	-0.18**	-0.07***	-0.18**	-0.01	-0.02**
Interest rate spread	3.98***	2.20***	0.87	5.03***	0.53*	0.50
RPDI per capita growth	5.33**	0.89	--	5.39**	0.17	--
Real GDP growth	5.86***	3.29**	--	9.22**	1.04***	--
Real stock growth	0.66***	0.58**	--	0.44***	0.04**	--
Current account ratio	4.83**	--	--	--	--	--
Real credit growth	--	--	--	-3.13**	--	--
Market volatility (VIX)	--	--	--	0.94**	--	--
AIC	0.18	0.18	0.17	0.24	0.25	0.21
McFadden's R^2	0.96	0.95	0.91	0.93	0.88	0.88

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent significance levels respectively. All results underlying the detection and date-stamping of exuberance are for autoregressive lag length $k=4$ with the initial window set from first quarter 1975 to first quarter 1982. Our panel is strongly balanced and includes dummies for global exuberance based on the 23 countries in our sample and nearly complete data on a large selection of covariates covering the period from first quarter 1986 to fourth quarter 2015. The covariates include interest rate spreads, the U.S. economic policy uncertainty index from PolicyUncertainty.com, growth in real per capita personal disposable income (RPDI), changes in real stock market valuations, changes in the West Texas Intermediate (WTI) index of oil prices deflated with U.S. CPI, the CBOE market volatility index (VIX extended with the VOX, old method data), real GDP growth, private real credit growth to the nonfinancial sector, the current-account-to-GDP ratio, and the growth rate in the real current account. All covariates are aggregated with a simple arithmetic mean.

We fit a dynamic panel probit model with constant term (M1-M3) with lagged exuberance where the dependent variable is constructed with the panel BSADF statistic. We report the results of a parsimonious specification of the dynamic panel model that includes the statistically significant variables. This benchmark M1, apart from lagged exuberance, includes policy uncertainty, interest rate spreads, real per capita personal disposable income (RPDI) growth, real GDP growth, changes in real stock market valuations, and additional financial variables (the current account to GDP ratio). We also report results after dropping the financial variables (M2) and all the explanatory variables except the spread and policy uncertainty (M3). Standard errors are robustified. We also report the results of three analogous specifications (benchmark in M4 where the relevant financial variables are credit growth and the VIX index, benchmark excluding the financial variables in M5, and benchmark excluding all covariates except the spreads and policy uncertainty in M6) where episodes of global exuberance are dated based on an aggregates real house price index obtained as the arithmetic mean of the 23 countries in the sample. Standard errors are robustified.

Sources: Federal Reserve Bank of Dallas' International House Price Database, National Sources (Central Banks and Statistical Offices), Bank for International Settlements, Wall Street Journal, Chicago Board Options Exchange (CBOE), PolicyUncertainty.com, Haver Analytics, and authors' calculations.

Based solely on aggregated data, we would conclude that credit growth is an important factor explaining exuberance—and that stock market volatility (VIX) has also effects above and beyond those priced into the interest rate spreads. In turn, when we exploit the panel

dimension, we see that is financing fuelled from the rest-of-the-world—instead of overall credit growth—the key financial factor. Therefore, these findings reinforce our main conclusion that the degree of aggregation of the data can be a limiting factor on the overall usefulness of right-tailed tests for monitoring asset markets—not solely because of the power losses that we show elsewhere in this paper, but also because of the potential impact of an aggregation bias (whose direction or magnitude are not well-known) on our empirical inferences.

5. Concluding Remarks

In this paper, we explore the relevance of aggregation for detecting periods of exuberance (mildly explosive) with simulated data and with two applications based on real house price data disaggregated by location. We find that by exploiting disaggregated data in a panel setting, we can more successfully identify such episodes. Our findings also suggest that aggregation bias can be a limiting factor to interpret the deep causes of mildly explosive behavior detected by recursive (right-tailed) unit root tests.

Therefore, we recognize that the collection and analysis of locational data across different housing markets may be of crucial importance to monitoring housing market developments. We argue more generally that this may warrant increased utilization not only of disaggregated data but also of panel data techniques, such as the panel GSADF proposed by Pavlidis *et al.* (2016), whenever possible.

Our findings indicate that the GSADF test applied to aggregated data is substantially more powerful than the SADF test and, even more so, than the standard (right-tailed) ADF test implemented over the full sample only. Our findings suggest that the recursive implementation of the (right-tailed) unit root test over a large sequence of subsamples is key to lessening the power loss associated with the use of aggregated data. Thus, the univariate GSADF test should be preferred to the univariate SADF test (and also to the standard ADF test) whenever the researcher is restricted to using aggregated data only.

References

- Anderson, T.W., 1959. On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676–687.
- Branch, W.A., McGough, B., 2009. A new Keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control* 33 (5), 1036–1051.
- Blanchard, O., Watson, M., 1982. Crisis in the economic and financial structure: Bubbles, bursts, and shocks. Chapter: Bubbles, rational expectations, and financial markets, 295–315. Watchel ed., Lexington Books, Lexington, Mass.
- Boswijk, H.P., Klaassen, F., 2012. Why frequency matters for unit root testing in financial time series. *Journal of Business and Economic Statistics* 30 (3), 351–357.
- Campbell, J.Y., Perron, P., 1991. Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomics Annual* (vol. 6), eds. O. J. Blanchard and S. Fisher, Cambridge, MA: The MIT Press.
- Choi, I., 1992. Effects of data aggregation on the power of tests for a unit root. *Economics Letters* 40 (4), 397–401.
- Clayton, J., 1996. Rational expectations, market fundamentals and housing price volatility. *Real Estate Economics* 24(4), 441–470.
- Diba, B.T., Grossman, H.I., 1988a. Explosive rational bubbles in stock prices? *American Economic Review* 78 (3), 520–530.
- Diba, B.T., Grossman, H.I., 1988b. The theory of rational bubbles in stock prices. *Economic Journal* 98 (392), 746–754.
- Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74 (366a), 427–431.
- Engsted, T., Hviid, S.J., Pedersen, T.Q., 2016. Explosive bubbles in house prices? Evidence from the OECD countries. *Journal of International Financial Markets, Institutions and Money* 40, 14–25.
- Evans, G.W., 1991. Pitfalls in testing for explosive bubbles in asset prices. *American Economic Review* 81 (4), 922–930.
- Evgenidis, A., Tsagkanos, G.A., Siriopoulos, C., 2017. Towards an asymmetric long-run equilibrium between economic uncertainty and the yield spread. A multi-economy view. *Research in International Business and Finance* 39 (A), 267–279.
- Flood, R.P., Garber, P.M., 1980. An economic theory of monetary reform. *Journal of Political Economy* 88(1), 24–58.

- Gordon, M.J., Shapiro, E., 1956. Capital equipment analysis: the required rate of profit. *Management Science* 3 (1), 102–110.
- Granger, C.W.J., 2008. Non-linear models: Where do we go next - Time varying parameter models? *Studies in Nonlinear Dynamics & Econometrics* 12 (3), pp. 1–11.
- Granger, C., Lee, T., 1999. The effect of aggregation on nonlinearity. *Econometric Reviews* 18 (3), 259–269.
- Gurkaynak, R., 2008. Econometric tests of asset price bubbles: Taking stock. *Journal of Economic Surveys* 22 (1), 166–186.
- Gyourko, J., Mayer, C., Sinai, T., 2006. Superstar cities. National Bureau of Economic Research, working paper no. 12355.
- Hamilton, J.D., Whiteman, C.H., 1985. The observable implications of self-fulfilling expectations. *Journal of Monetary Economics* 16(3), 353–373.
- Hiebert, P., Sydow, M., 2011. What drives returns to euro area housing? Evidence from a dynamic dividend-discount model. *Journal of Urban Economics* 70 (2–3), 88–98.
- Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics* 10 (1), 198–231.
- Hu, Y., Oxley, L., 2018. Bubbles in U.S. regional house prices: evidence from house price-income ratios at the state level. *Applied Economics* 50 (29), 3196–3229.
- Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115 (1), 53–74.
- Knight, F.H., 1921. Risk, uncertainty, and profit. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.
- LeRoy, S., Porter, R., 1981. The present-value relation: Tests based on implied variance bounds. *Econometrica* 49, 555–577.
- 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. Federal Reserve Bank of Dallas.
- Grossman, V., Martínez-García, E., 2018. Explosive Dynamics in House Prices? An Exploration of Financial Contagion in International Housing Markets. Globalization and Monetary Policy Institute Working Paper no. 342. Federal Reserve Bank of Dallas.
- Nielsen, B., 2010. Analysis of coexplosive processes. *Econometric Theory* 26 (3), 882–915.

- Pavlidis, E.G., Yusupova, A., Paya, I., Peel, D., Martínez-García, E., Mack, A., Grossman, V., 2016. Episodes of exuberance in housing markets: In search of the smoking gun. *The Journal of Real Estate Finance and Economics* 53 (4), 419–449.
- Perron, P., 1991. Test consistency with varying sampling frequency. *Econometric Theory* 7, 341–368.
- Phillips, P.C.B., Magdalinos, T., 2007a. Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
- Phillips, P.C.B., Magdalinos, T., 2007b. The Refinement of econometric estimation and test procedures: Finite sample and asymptotic analysis, Chapter: Limit theory for moderate deviations from a unit root under weak dependence, pp. 123–162. Cambridge University Press, Cambridge.
- Phillips, P.C.B., Magdalinos, T., 2012. Mildly explosive autoregression under weak and strong dependence. *Journal of Econometrics* 169 (2), 179–187.
- Phillips, P.C.B., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* 2 (3), 455–491.
- Phillips, P.C.B., Shi, S.-P., Yu, J., 2015a. 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., Shi, S.-P., Yu, J., 2015b. Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review* 56 (4), 1079–1134.
- Pierse, R.G., Snell, A.J., 1995. Temporal aggregation and the power of tests for a unit root. *Journal of Econometrics* 65 (2), 333–345.
- Said, S.E.; Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71 (3), 599–607.
- Sargent, T.J., 1987. *Macroeconomic theory*, 2nd edition. Boston: Academic Press.
- Shi, S., 2017. Speculative bubbles or market fundamentals? An investigation of U.S. regional housing markets. *Economic Modelling* 66 (November), 101–111.
- Shiller, R.J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.
- Shiller, R.J., Perron, P., 1985. Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters* 18, 381–386.
- Tsagkanos, G.A., Siriopoulos, C., 2015. Stock markets and industrial production in North and South of Euro-zone: Asymmetric effects via threshold cointegration approach. *Journal of Economic Asymmetries* 12 (2), 162–172.

West, K.D., 1987. A specification test for speculative bubbles. *Quarterly Journal of Economics* 102 (3), 553–580.

White, J.S., 1958. The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.

Yusupova, A., Pavlidis, E.G., Paya, I., Peel, D.A., 2017. Exuberance in the U.K. regional housing markets. Lancaster University Working Paper 012.

Appendix. Mildly Explosive Behavior, Uncertainty, and Risk-Spreads

To motivate our subsequent empirical analysis and the challenges posed by aggregation, we start with a theoretical distinction between non-fundamental behavior (bubbles) in episodes of exuberance and fundamental-based explosiveness in the housing data. We adopt the asset pricing framework (in line with Clayton, 1996; Hiebert and Sydow, 2011; Pavlidis *et al.* 2016) under standard arbitrage conditions for the study of financial contagion and explosive behavior in house prices (whose formal derivation from first principles can be found in Martínez-García and Grossman, 2018). The asset pricing model yields the following difference equation for house prices

$$P_t = \frac{1}{\varepsilon_t[I_{t+1}^S]} \varepsilon_t[P_{t+1} + (1 - b)Y_{t+1}], \quad (A1)$$

where $\varepsilon_t[I_{t+1}^S]$ is the discount factor, P_{t+1} is the asset-value of a housing unit at $t+1$ and $(1 - b)Y_{t+1}$ is the housing rental income (either actual if the housing unit is rented or imputed if owner-occupied) at $t+1$. Here, rental income is defined as the payout ratio $(1 - b)$ times the flow of earnings on the housing unit (actual or imputed) Y_{t+1} ; in this way, housing rents net out expenses such as maintenance costs, taxes, reinvestment, etc. from total earnings.

The difference equation in (A1) shows that the price of a housing unit today must be equal to the discounted value of tomorrow's expected rental income (the payout component of housing earnings) plus tomorrow's resale price of housing. We discount using an appropriate rate—the expected return on an alternative risky investment—given by:

$$\varepsilon_t[I_{t+1}^S] = I^f + f(\sigma_m, \sigma_s; \cdot). \quad (A2)$$

This discount rate adds a risk-premium component ($f(\sigma_m, \sigma_s; \cdot)$) to the risk-free rate (I^f) which prices macro risks such as those arising from macro volatility (σ_m) or from market volatility associated to the alternative asset that could be used for investment (σ_s).¹⁷

By replacing (A2) into (A1) and recursively substituting forward, we obtain the following present-value specification:

$$P_t = (1 - b)\varepsilon_t \left[\sum_{j=1}^{+\infty} \left(\frac{1}{I^f + f(\sigma_m, \sigma_s; \cdot)} \right)^j Y_{t+j} \right] + \lim_{T \rightarrow +\infty} \varepsilon_t \left[\left(\frac{1}{I^f + f(\sigma_m, \sigma_s; \cdot)} \right)^T P_{t+T} \right]. \quad (A3)$$

We allow earnings from investing in a housing unit (Y_t) to follow a general autoregressive process of order 1,

$$Y_t = \varphi Y_{t-1} + \epsilon_t, \epsilon_t \sim WN(0, \sigma_\epsilon^2), \quad (A4)$$

where φ is unrestricted to take values in the $I(0)$, $I(1)$, and mildly-explosive regions, and ϵ_t is an independent and identically-distributed (i.i.d.) white noise stochastic process.

¹⁷ Whenever it follows that $f(\sigma_m, \sigma_s; \cdot) > 0$ and $\frac{\partial f(\sigma_m, \sigma_s; \cdot)}{\partial \sigma_m} > 0, \frac{\partial f(\sigma_m, \sigma_s; \cdot)}{\partial \sigma_s} > 0$, *ceteris paribus* the risk-premium is higher the higher macro volatility (σ_m) or the higher market volatility (σ_s) are.

Heterogenous beliefs and uncertainty. An important economic concept is that of Knightian uncertainty (Knight, 1921) which arise when the probabilities of outcomes cannot be accurately measured or are simply unknown—unlike risk which is related to outcomes whose odds are known or can be learned. Knightian uncertainty goes beyond situations in which we cannot measure accurately the odds of events and includes cases for which we don't know all possible outcomes, but here we highlight monetary policy uncertainty in the former case assuming that heterogenous beliefs (and disagreements across investors) reflect the difficulty in assessing the odds on the risk-free rate (I^f).

For simplicity, we assume that there is a finite number of possible outcomes for the risk-free rate which we narrow down to just two: a high rate $I^{f,H}$ and a low rate $I^{f,L}$ such that $0 < I^{f,L} \leq I^{f,H}$. Furthermore, we assume two types of agents indexed $\tau = L, H$ that put a mass of probability one in either the low or the high rate outcomes, respectively. In this simple setup, uncertainty about the risk-free rate reflects disagreements on the stance of monetary policy and leads to heterogeneous beliefs on the economically-relevant discount factor for pricing housing investments.

Heterogeneous expectations are introduced into the asset-pricing model in (A1)-(A4) following the axiomatic approach in Branch and McGough (2009). In particular, we assume that agents' expectations satisfy the law of iterated expectations at the aggregate level and rule out wealth distribution dynamics that otherwise could influence the forecasts by agent type ensuring a structure on higher-order beliefs that allows for aggregation. Thus, we denote the aggregate (subjective) expectations operator on the generic variable g_{t+1} at $t + 1$ based on information up to time t as $\varepsilon_t[g_{t+1}]$ and we define it as a weighted average of the subjective expectations across agent types τ , i.e., $\varepsilon_t[g_{t+1}] = \sum_{\tau \in \{L,H\}} \theta_\tau \varepsilon_t^\tau[g_{t+1}]$. The mass of each agent type θ_τ satisfies that $\sum_{\tau \in \{L,H\}} \theta_\tau = 1$ and $0 \leq \theta_L \leq \frac{1}{2}$.

We assume that both agent types are rational and form their expectations with all available information up to the present, except because uncertainty on the stance of monetary policy shows up in their disagreement about the likely state in which the (constant) risk-free rate falls. Accordingly, we can rewrite equations (A1) and (A2) as follows

$$P_t = \frac{1}{\theta_L \varepsilon_t^L[I_{t+1}^s] + \theta_H \varepsilon_t^H[I_{t+1}^s]} [\theta_L \varepsilon_t^L[P_{t+1} + (1-b)Y_{t+1}] + \theta_H \varepsilon_t^H[P_{t+1} + (1-b)Y_{t+1}]], \quad (\text{A5})$$

$$\varepsilon_t^\tau[I_{t+1}^s] = I^{f,\tau} + f(\sigma_m, \sigma_s; \cdot), \tau \in \{L, H\}, \quad (\text{A6})$$

where we assume that the expectations of each agent type $\varepsilon_t^\tau[\cdot]$ are rational albeit under heterogenous beliefs on the risk-free rate. From here we obtain the counterpart of the present-value specification in (A3) as:

$$P_t = (1-b) \left[\sum_{j=1}^{+\infty} \left(\frac{1}{\theta_L I^{f,L} + \theta_H I^{f,H} + f(\sigma_m, \sigma_s; \cdot)} \right)^j \varepsilon_t(Y_{t+j}) \right] + \dots$$

$$\lim_{T \rightarrow +\infty} \varepsilon_t \left[\left(\frac{1}{\theta_L I^{f,L} + \theta_H I^{f,H} + f(\sigma_m, \sigma_s; \cdot)} \right)^T P_{t+T} \right]. \quad (\text{A7})$$

Moreover, given that the expectations about the future fundamentals (Y_{t+j}) do not depend on the agents' beliefs about the risk-free rate and are rationally formed, it follows that $\varepsilon_t(Y_{t+j}) = \varepsilon_t^L[Y_{t+j}] = \varepsilon_t^H[Y_{t+j}] = \varphi^j Y_t$.

Asset-pricing under heterogenous beliefs. We define the weighted risk-free rate for the aggregate present-value specification in (A7) as $I^f = \theta_L I^{f,L} + \theta_H I^{f,H} = I^{f,H} - I^{f,Diff}$ and measure uncertainty by the disagreement amongst agent types given by $I^{f,Diff} \equiv \theta_L (I^{f,H} - I^{f,L})$. In this setting, so long as $I^{f,H} > I^{f,L}$ ($I^{f,H} \neq I^{f,L}$), uncertainty reaches its lowest when $\theta_L = 0$ and there is no disagreement $I^{f,Diff} = 0$. In turn, uncertainty is at its highest whenever both agent types reach the highest disagreement point $I^{f,Diff} = \frac{1}{2} (I^{f,H} - I^{f,L})$ and $\theta_L = \frac{1}{2}$. Hence, the higher the uncertainty becomes, the smaller the weighted risk-free rate I^f would be.

The rational-expectations solution to the present-value model in (A4)-(A7) under heterogenous beliefs consists of a fundamental component (P_t^*) and a periodically-collapsing rational bubble (B_t) such that $P_t = P_t^* + B_t$ (similar to Sargent 1987; Diba and Grossman 1988a, 1988b). There exist infinite forward solutions to the house price P_t implied by the present-value model in (A7) where the non-fundamental component (rational bubble) B_t satisfies the submartingale property:

$$\varepsilon_t[B_{t+1}] = (I^f + f(\sigma_m, \sigma_s; \cdot)) B_t, \quad (A8)$$

if the transversality condition $\lim_{T \rightarrow +\infty} \varepsilon_t \left[\left(\frac{1}{I^f + f(\sigma_m, \sigma_s; \cdot)} \right)^T P_{t+T} \right] = 0$ does not hold. Since the discount factor is positive ($I^f + f(\sigma_m, \sigma_s; \cdot) > 0$), the term B_t is thus expected to be explosive.

Imposing the transversality condition $\lim_{T \rightarrow +\infty} \varepsilon_t \left[\left(\frac{1}{I^f + f(\sigma_m, \sigma_s; \cdot)} \right)^T P_{t+T}^* \right] = 0$ to rule out non-fundamental behavior (rational bubbles), the unique (nonexplosive) solution to (A8) whenever $I^f + f(\sigma_m, \sigma_s; \cdot) - \varphi > 0$ yields a variant of the discounted dividend model (Gordon and Shapiro, 1956) for real house prices, i.e.,

$$P_t^* = \left(\frac{\varphi}{I^f + f(\sigma_m, \sigma_s; \cdot) - \varphi} \right) (1 - b) Y_t, \quad (A9)$$

and

$$P_t^* = \varphi P_{t-1}^* + \left(\frac{\varphi}{I^f + f(\sigma_m, \sigma_s; \cdot) - \varphi} \right) (1 - b) \varepsilon_t, \quad (A10)$$

where the fundamental dynamics of the house price in (A10) follow from (A9) and (A4).

Implications for explosiveness in house prices. The fact that B_t is explosive has important implications for house prices, according to this model. If the economic fundamentals in (A4) follow either a stationary ($|\varphi| < 1$) or an integrated process of order 1 (a martingale with $\varphi = 1$), then the dynamics of fundamental-based prices given by (A10)—which inherit the autoregressive coefficient φ directly from (A4)—also display non-explosive behavior.

Furthermore, house prices and fundamentals would be cointegrated in the absence of the bubble term B_t . Hence, a possible reason why house prices display explosive dynamics in this framework is because of the emergence of rational bubbles.

Often the observable housing fundamentals of the housing markets show non-explosive behavior in the data. However, factors other than bubbles can give rise to explosive dynamics in house prices too even when housing earnings (or housing rents) are non-explosive. Within the asset pricing framework laid out here, time-variation in the discount rate can lead to explosive behavior even when rational bubbles are ruled out and fundamentals are known to be non-explosive (see, e.g., Pavlidis *et al.*, 2016). In fact, it can be shown that unexpected changes in the risk-premium $rp \equiv f(\sigma_m, \sigma_s; \cdot)$ that raise the discount rate can lead to episodes of explosive behavior within sample, while greater policy uncertainty lowering the weighted risk-free rate I^f has the opposite effect.

For instance, let's say that investors price in $rp = \sigma^L > 0$, but at time t they wake up to discover that $rp^H > rp^L$ before risk-spreads return to rp^L from $t+1$ onwards. This unexpected (short-lived and temporary) spike in riskiness alone implies that equation (A10) no longer holds and the dynamics of fundamental-based house prices become:

$$P_{t+j+1}^* = \delta_j(\varphi, rp^H, rp^L; \cdot) P_{t+j}^* + \left(\frac{\varphi}{I^f + rp^L - \varphi} \right) (1 - b) \epsilon_{t+j}, \forall j \geq 0, \quad (\text{A11})$$

where $\delta_j(\varphi, rp^H, rp^L; \cdot) = \begin{cases} \varphi \left(\frac{I^f + rp^H}{I^f + rp^L} \right), & \text{if } j = 0 \\ \varphi, & \text{if } j > 0 \end{cases}$. Under a standard parameterization ensuring

that $I^f + rp^L > 0$, it follows that $\left(\frac{I^f + rp^H}{I^f + rp^L} \right) > 1$. Hence, this illustration shows that unexpected shifts on the risk-spread can generate an episode of explosive behavior whereby

$\delta_j(\varphi, rp^H, rp^L; \cdot) > 1$ for $j = 0$ when housing fundamentals are $I(1)$ ($\varphi = 1$) and even when fundamentals are stationary ($|\varphi| < 1$) if it holds that $\varphi \left(\frac{I^f + rp^H}{I^f + rp^L} \right) > 1$. Analogously, we can infer that an increase in uncertainty that increases the disagreement amongst agent types $I^{f, Diff}$ and lowers the weighted risk-free rate I^f would have the opposite effect.

Supplementary Appendix. Recursive Implementation of the Right-Tailed ADF Tests

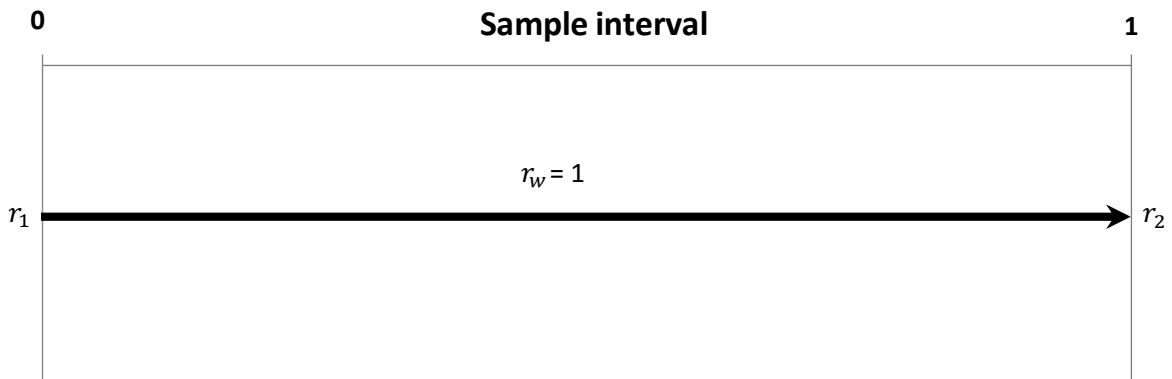
In this paper we implement a number of right-tailed unit root tests designed to detect the presence of periods of mildly explosive behavior within sample. These tests include:

1. ADF
2. sup ADF (SADF), see Phillips and Yu (2011)
3. Generalized SADF (GSADF), see Phillips *et al.* (2015a,b)
4. Panel GSADF, see Pavlidis *et al.* (2016)

These right-tailed unit root tests differ crucially on the recursion mechanism used in their implementation. In order to illustrate this, we need to review some notation first. The full sample of T observations is normalized on the interval $[0,1]$. Here, we denote r_1 and r_2 as the corresponding fractions of the sample which define the beginning and end of a given subsample such that $0 \leq r_1 < r_2 \leq 1$. We denote by $r_w = r_2 - r_1$ the window size of the regression estimation, while r_0 is the required fixed initial window which satisfies that the subsample ending in r_2 is such that $r_2 \in [r_0, 1]$ (i.e., r_0 is the required minimum window size).

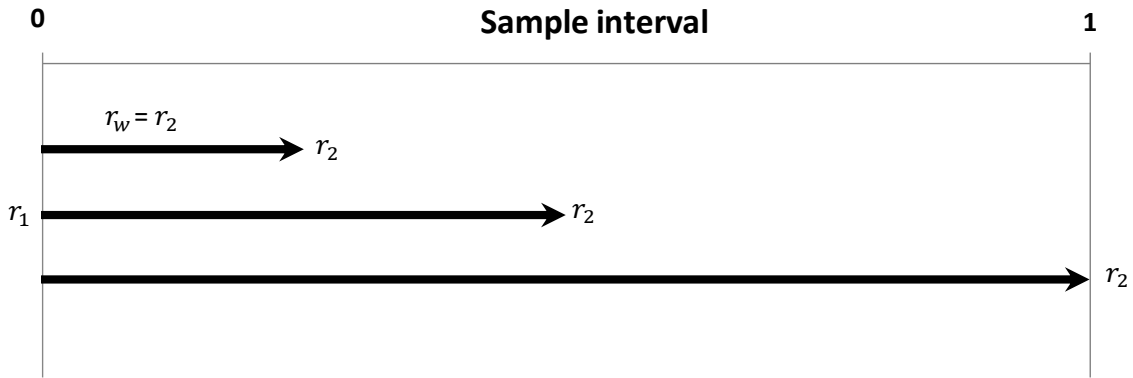
The first test is a right-tailed version of the standard ADF unit root test. With the given notation for a generic recursive mechanism, the implementation of the ADF test can be represented graphically simply as follows:

Illustration of the ADF Procedure



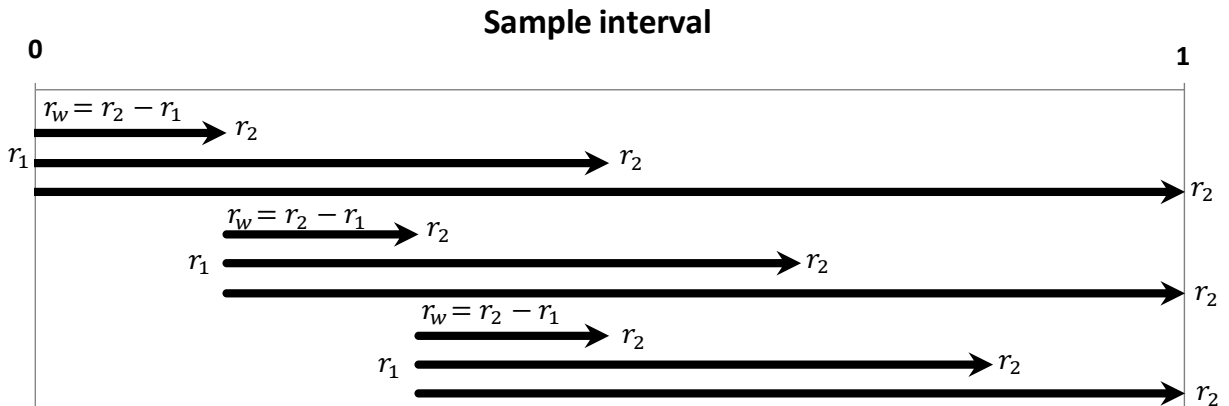
The SADF test is based on a proper recursion mechanism based of the ADF test statistics with an expanding window. The recursion mechanism goes as follows in this case:

Illustration of the SADF Procedure



The SADF test suffers from a loss of power in the presence of multiple periodically-collapsing occurrences of mildly explosive behavior. As a preferred alternative, Phillips *et al.* (2015a,b) suggest the GSADF test procedure which is a generalization of the SADF test that allows a more flexible recursion mechanism where the starting point r_1 varies within the range $[0, r_2 - r_0]$. The same recursion mechanism is applied in the panel GSADF procedure of Pavlidis *et al.* (2016). Formally, the GSADF test recursion can be illustrated as follows:

Illustration of the GSADF Procedure



The GSADF recursion mechanism adopts the following strategy: set $r_1 \in [0, r_2 - r_0]$ and $r_2 \in [r_0, 1]$; use $[r_1, r_2]$ as a moving window where $r_w = r_2 - r_1$ is the corresponding window width for each subsample; and, then, vary r_1 and r_2 over the full sample.