Federal Reserve Bank of Dallas
Globalization and Monetary Policy Institute
Working Paper No. 325
https://doi.org/10.24149/gwp325

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

Efthymios Pavlidis
Lancaster University Management School

Enrique Martinez-Garcia
Federal Reserve Bank of Dallas

Valerie Grossman
Federal Reserve Bank of Dallas

August 2017

Abstract

The recently developed SADF and GSADF unit root tests of Phillips *et al.* (2011) and Phillips *et al.* (2015) 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 actual housing data for both U.S. metropolitan 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.* (2015), can perform substantially better than univariate tests applied to aggregated series.

JEL codes: C22, C12, G12, R30, R31

* Enrique Martinez-Garcia, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75201. 214-922-5262. enrique.martinez-garcia@dal.frb.org. We would like to thank Celso Brunetti, Marc Joëts, María Teresa Martínez-García, and participants at the 3rd International Workshop on "Financial Markets and Nonlinear Dynamics" for helpful discussions and suggestions. We acknowledge the support of the Federal Reserve Bank of Dallas and the research assistance provided by Kelvinder Virdi. All remaining errors are ours alone. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

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 it is consistent with the submartingale (explosive) property commonly used to define 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 can be seen as evidence of non-fundamental (bubble-like) behavior.

Diba and Grossman (1988a; 1988b) are 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 (Homm and Breitung, 2012). These recursive unit root tests—namely the supremum ADF (SADF) and the Generalized SADF (GSADF)–have been widely employed for testing for speculative dynamics in asset prices over the last decade (Phillips and Yu (2011); Phillips *et al.*, 2015). Perhaps their most popular application is in real estate markets.[1] Following

[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

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*. 2015; Engsted *et al.* 2016; Yusupova *et al.* 2016).

A common feature of all the above studies is that they employ house prices 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 growth of housing prices across local/regional markets (e.g., house prices in San Francisco behave very differently than prices in Denver or Washington) and, 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).[2]

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.

---

and Porter, 1981; Shiller, 1981), specification tests (West, 1987), and Chow and CUSUM-type tests (Homm and Breitung, 2012).

[2]During the second half of the 20th century, house prices in the ten U.S. metropolitan areas with the highest growth rates (adjusted for inflation) increased by 2.2 to 3.5 percent. The increase in housing prices in the ten metropolitan areas with the lowest rates was more than three times smaller, between 0.5 and 1.1 percent.

In this paper, we examine the role of cross-sectional aggregation by conducting two large simulation experiments based on actual real house price data. The first experiment 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 two simulation exercises is that aggregation lowers the power of both the SADF test of Phillips *et al.* (2011) and the GSADF test of Phillips *et al*. (2015). 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. We also provide evidence that the panel GSADF recently proposed by Pavlidis *et al.* (2015) is preferable whenever disaggregated data is available.

The structure of the remainder of the paper is as follows: Section 2 outlines the SADF test of Phillips and Yu (2011), the GSADF test of Phillips *et al*. (2015), and the panel GSADF test of Pavlidis *et al.* (2015). 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 actual and the simulated data. Section 4 argues about the implications of our simulation analysis on the role of aggregation and about the practical implementation of these novel tests for monitoring the dynamics of asset prices. Section 5 concludes.

2. Tests of 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:

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

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 $T \to \infty$.[3] Subtracting y_{t-1} from both sides, the process in (1) can be expressed 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 $\theta_{k+1}(B) = 1 - \theta^1 B - \cdots - \theta^k B^k - \theta^{k+1} B^{k+1}$ 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]

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

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

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, our statistical toolkit for the detection of periods of mildly explosive behavior includes: the supremum ADF (SADF) of Phillips *et al.* (2011) and the generalized SADF (GSADF) of Phillips

[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.

[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.

et al. (2015), together with the novel panel GSADF test of Pavlidis *et al.* (2015). The GSADF test was designed to overcome the SADF's lack of power in identifying multiple episodes of periodically-collapsing mildly explosive behavior within sample. Whenever disaggregated data is available, the panel GSADF test offers a more flexible and powerful alternative than the univariate GSADF test. This is because the panel GSADF test explicitly models the cross-sectional dependencies and heterogeneity present in the constituent series that are otherwise muddled together in the aggregated series. However, whenever disaggregated data is not available, we find that the GDSAF test tends to outperform the SADF test in terms of power thanks to the more flexible 'recursive' specification used in the implementation of the ADF regression equation (which we dissect further in the remainder of this section).

Second, some notation is required to describe the recursive implementation of the SADF and GSADF tests. 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 required by the econometrician).

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

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

where y_t denotes the generic time series tested for explosiveness, Δy_t 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.

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, $ADF_0^1 = \frac{\hat{\beta}_{0,1}}{s.e.(\hat{\beta}_{0,1})}$. Under the I(1) null, the limit distribution of ADF_0^1 is given by $\frac{\int_0^1 \widetilde{W}(r) d\widetilde{W}(r)}{\left(\int_0^1 \widetilde{W}(r)^2 dr\right)^{\frac{1}{2}}}$, where $\widetilde{W}$ 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 it is inherently 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 *et al.* (2011) proposes a recursive 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 *et al.* (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 $\sup_{r_2 \in [r_0,1]} \frac{\int_0^{r_2} \widetilde{W}(r) d\widetilde{W}(r)}{\left(\int_0^{r_2} \widetilde{W}(r)^2 dr\right)^{\frac{1}{2}}}$. Whenever $SADF(r_0)$ 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 $SADF(r_0)$ 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, Homm and Breitung (2012) show through simulation experiments 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 as well.

The alternative tests considered by Homm 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 $SADF(r_0)$ test power and its performance deteriorate in the presence of recurring (more than one) and periodically-collapsing episodes of exuberance, as established in Phillips *et al.* (2015).

2.3 Generalized SADF (GSADF) Test

Phillips *et al.* (2015) 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 (while the $SADF(r_0)$ test is only consistent with a single such episode in-sample).

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

The Phillips *et al.* (2015) 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:

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

Under the I(1) null, the limit distribution of the $GSADF(r_0)$ statistic is given by $\underset{r_1 \in [0, r_2 - r_0], r_2 \in [r_0, 1]}{sup} \left\{ \frac{\frac{1}{2} r_w [W(r_2)^2 - W(r_1)^2 - r_w] - \int_{r_1}^{r_2} W(r) dr [W(r_2) - W(r_1)]}{r_w^{\frac{1}{2}} \left\{ r_w \int_{r_1}^{r_2} W(r)^2 dr - \left[\int_{r_1}^{r_2} W(r) dr \right]^2 \right\}^{\frac{1}{2}}} \right\}$. 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 test and with the $SADF(r_0)$ test.

2.4 Panel GSADF Test

Pavlidis *et al.* (2015) 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):

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

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 $\{y_t^s\}_{s=1}^N$, $H_0\colon \beta_{r_1,r_2}^s = 0, \forall s$, against the alternative of mildly explosive behavior, $H_1\colon \beta_{r_1,r_2}^s > 0$ 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.,

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

The distribution of panel unit root tests 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.* (2015) 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 a mildly explosive process.

The specification of the alternative hypothesis proposed by Pavlidis *et al.* (2015) 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 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 surprisingly due to the additional flexibility of the framework that incorporates all relevant cross-sectional information for testing (not just the time series information), 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 Power Loss 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. The reader is referred to Blanchard and Watson (1982), Diba and Grossman (1988a; 1988b), and Evans (1991) for theoretical DGPs consistent with the fundamental asset pricing equation that display such nonlinear and explosive dynamics (in the presence of rational bubbles).

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

[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 y_t for any subsample can be expressed as $H_0\colon \beta_{r_1,r_2} = \frac{1}{N}\sum_{s=1}^N \beta_{r_1,r_2}^s = 0$, while the null hypothesis of a unit root in all disaggregated series $\{y_t^s\}_{s=1}^N$ imposes the stricter requirement that $H_0\colon \beta_{r_1,r_2}^s = 0, \forall s$. In turn, the alternative of mildly explosive behavior in y_t for any subsample is expressed as $H_1\colon \beta_{r_1,r_2} = \frac{1}{N}\sum_{s=1}^N \beta_{r_1,r_2}^s > 0$, while the alternative hypothesis tested on disaggregated data $\{y_t^s\}_{s=1}^N$ simply requires that $H_1\colon \beta_{r_1,r_2}^s > 0$ for some (but not necessarily all) of the constituent series s. 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. However, in comparing the two univariate and panel GSADF tests, we often find that the ability or power of the panel test to detect mildly explosive behavior if such behavior truly occurs in the sample (rejecting the null when it is incorrect) tends to be higher than that of the univariate GSADF test.

implies that, even if we knew the true DGP for each individual constituent series, we may still not be able to infer the exact DGP 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., of mildly explosive dynamics). It is well known that the combination of explosive processes with other explosive, unit root, and/or stationary processes 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—any one of the constituent real house price indices in our illustrations, for example—is explosive, so will be the aggregate series.

However, we argue that aggregation affects the performance of the (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 do not display similar patterns of explosiveness.

We evaluate the performance of the (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 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 (right-tailed) unit root ADF test improves when we exploit the recursive nature of the SADF specification of Phillips

[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).

and Yu (2011) and especially that of the GSADF specification of Phillips *et al.* (2015). 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 aggregate 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, the generalized recursive implementation 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 (right-tailed) ADF test over the full sample as the performance of the (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 data can become diluted due to aggregation. In turn, the smaller effect on the aggregate series lowers the power for any of the non-recursive and 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 $\{y_t^s\}_{s=1}^N$.

[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).

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):

$$\Delta y_t^s = \epsilon_t^s, \forall s = 1, \dots, N-1, \tag{8}$$

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

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 illustration, we see that the dynamics for the aggregate series can be expressed as:

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

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 magnitude of the deviation from unit root behavior can get diluted in the aggregated series. In this case, we observe that the relative contribution of the constituent series $s = N$ which displays mildly explosive behavior to the aggregate series weighs down the magnitude of the effect given by the coefficient β. Intuitively, the smaller effect incorporated into the aggregate series can lead to lower statistical power 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 the panel GSADF of Pavlidis *et al.*, 2015) 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. Hence, it remains an empirical question 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 two simulation exercises applied to real house price data that we discuss in the remainder of the paper aim to answer formally those questions.

3. Simulation Experiments

The empirical question of interest for the detection of periods of exuberance is: what is the effect of aggregation on the power of the SADF and GSADF tests in finite samples? To shed light on this issue, our Monte Carlo simulation procedure under the assumption of normality of the errors consists of the following steps:

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

$$\Delta y_t = a_{r_1,r_2} + \beta_{r_1,r_2} y_{t-1} + \sum_{j=1}^{k} \psi_{r_1,r_2}^{j} \Delta y_{t-j} + \epsilon_t,\ \ \epsilon_t \sim i.i.d.\left(0, \sigma_{r_1,r_2}^2\right), \tag{10}$$

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.* (2015), $r_2 - r_1 \geq 0.01 + 1.8/\sqrt{T}$, 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. For further details on the BSADF, SADF, and GSADF statistics, see Phillips *et al.* (2015) and Pavlidis *et al.* (2015).

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:

$$\Delta y_t^s = \hat{a}_{r_1,r_2} + \hat{\beta}_{r_1,r_2} y_{t-1}^s + \sum_{j=1}^{k} \hat{\psi}_{r_1,r_2}^{j} \Delta y_{t-j}^s + \epsilon_t^s,\ \ \epsilon_t^s \sim N\left(0, \hat{\sigma}_{r_1,r_2}^2\right). \tag{11}$$

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 the actual 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.

Technical Details: 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.* (2015), $r_0 = 0.01 + \frac{1.8}{\sqrt{T}}$ where T is the total number of observations. With respect to the autoregressive lag length k, we evaluate our results for the case where k is equal to 4. Our findings do not appear very sensitive to the lag length specification.

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.

3.1 Aggregation from Local to National Real House Prices

Our first Monte Carlo simulation exercise focuses on the U.S. housing market and its major metropolitan statistical areas. For the exercise, we use data on monthly house prices for the S&P/Case-Shiller 10-City Composite Home Price Index and its constituents, 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.

Table 1 reports the SADF and GSADF statistics for the actual S&P/Case-Shiller 10-City Composite Home Price Index and its constituents, as well as finite-sample critical values. According to the results, the null hypothesis of a unit root can be rejected at the 5% significance level for all metropolitan areas except for Los Angeles, for which the SADF statistic does not exceed its critical value. Thus, all series but one appear to exhibit exuberance during (at least) part of the sample period under consideration. Using the GSADF test, all series appear to display 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 series.

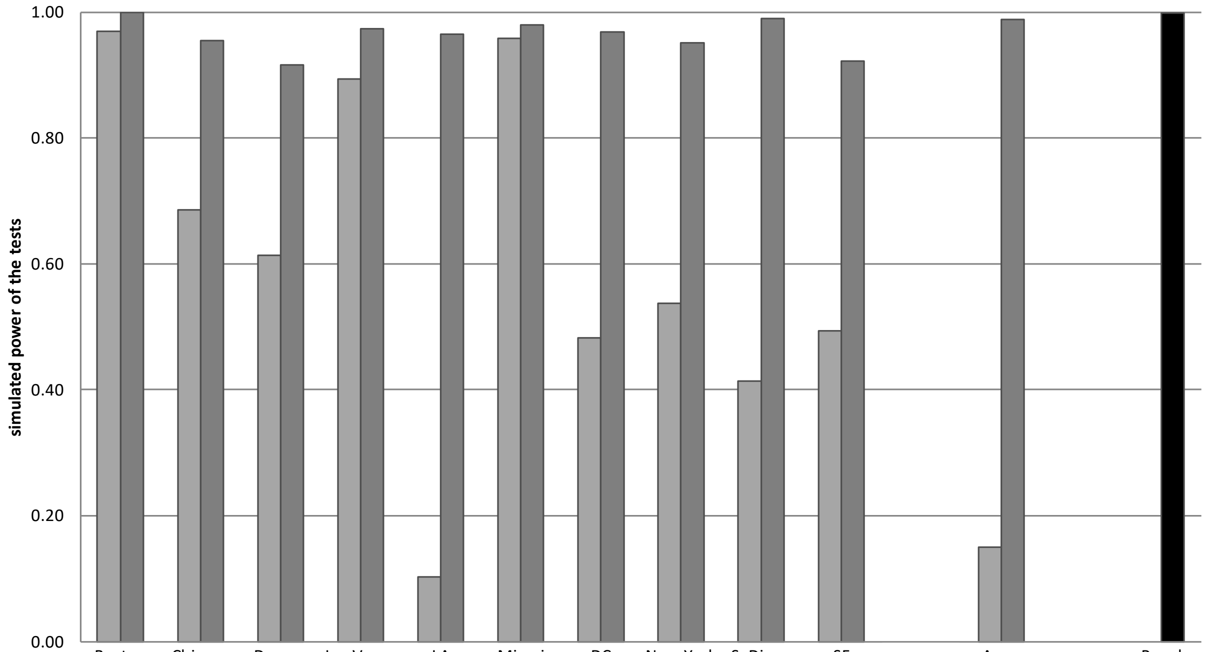
In accordance with the results for the actual series, the simulation results reported in Chart 1 show that the power of the SADF and GSADF tests for the aggregate series lies between the lowest and highest power for the constituent series. This result implies that aggregation leads to power losses, since if there was no power loss then every time a test detected explosive dynamics in a constituent, it would 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 power is 96 percent, while the power for the 10-city aggregate that includes them is merely 16 percent with the SADF test. 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.

Table 1. 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	2.417***	4.026***	
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.
Sources: Standard & Poor's, Bureau of Labor Statistics, authors' calculations.

Chart 1. S&P/Case-Shiller 10-City Composite Home Price Index, Simulation Results



	Boston	Chicago	Denver	Las Vegas	LA	Miami	DC	New York	S. Diego	SF	Aggr.	Panel
SADF	0.97	0.69	0.61	0.89	0.10	0.96	0.48	0.54	0.41	0.49	0.15	
GSADF	1.00	0.96	0.92	0.97	0.97	0.98	0.97	0.95	0.99	0.92	0.99	1.00

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.

3.2 Aggregation from National to International Real House Prices

To provide further evidence on the effect of aggregation on the performance of both sequential 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 (Mack and Martínez-García, 2011). The database comes at quarterly frequency and covers 23 countries between the first quarter of 1975 and the fourth quarter of 2015, deflated with the corresponding country PCE deflator.[9] Table 2 reports the SADF and GSADF statistics for the actual aggregate real house price index and its constituent country indices, as well as the corresponding finite-sample critical values; and Chart 2 displays the simulation results. This

[9] The 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. All data can be accessed publicly at: http://www.dallasfed.org/institute/houseprice/

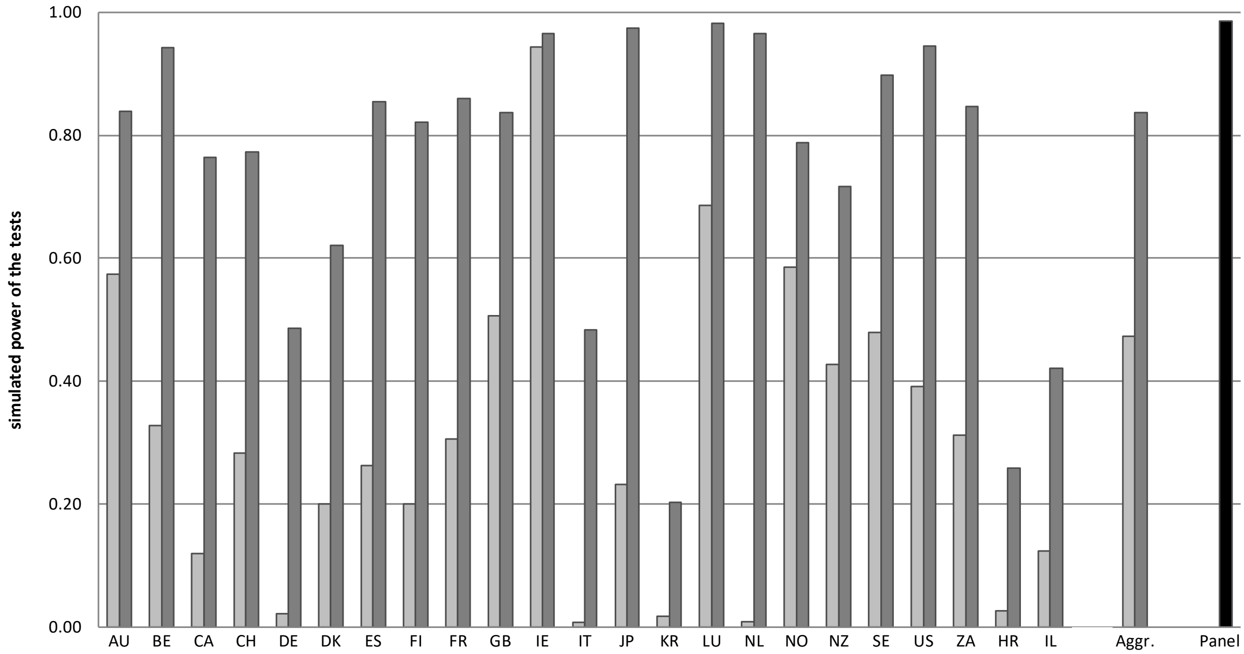
second set of results shows even more clearly than in the previous exercise the severe power loss of the SADF test due to aggregation, and the superior performance of the GSADF test over the SADF. However, it also points out that power losses relative to the panel GSADF test can be non-negligible even when using the more robust GSADF test.

Table 2. 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	-0.242	2.910***	
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.
Sources: Federal Reserve Bank of Dallas' International House Price Database and authors' calculations.

Chart 2. Dallas Fed's Real House Price Series, Simulation Results



	AU	BE	CA	CH	DE	DK	ES	FI	FR	GB	IE	IT	JP	KR	LU	NL	NO	NZ	SE	US	ZA	HR	IL	Aggr.	Panel
SADF	0.57	0.33	0.12	0.28	0.02	0.20	0.26	0.20	0.31	0.51	0.94	0.01	0.23	0.02	0.69	0.01	0.59	0.43	0.48	0.39	0.31	0.03	0.12	0.47	
GSADF	0.84	0.94	0.76	0.77	0.49	0.62	0.86	0.82	0.86	0.84	0.97	0.48	0.97	0.20	0.98	0.97	0.79	0.72	0.90	0.95	0.85	0.26	0.42	0.84	0.99

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.

4. Implications for Detecting Explosiveness

The 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 SADF and GSADF tests, developed by Phillips and Yu (2011) and Phillips *et al.* (2015) respectively. Without loss of generality, we do so using available real house price data that is reported at different levels of aggregation across locations. We specifically investigate aggregation of local real house prices to national prices and aggregation of national real prices to international real prices. To the best of our knowledge, the role that aggregation plays on detecting periods of mildly explosive behavior in time series has not been examined in the literature. Furthermore, we explore the panel GSADF test proposed by Pavlidis *et al.* (2015) to exploit the cross-sectional variation in disaggregated data and to provide a benchmark to quantify the power loss associated with aggregation.

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

Asset prices in general, and real house prices in particular, are most likely characterized by a nonlinear DGP, displaying I(1) dynamics in some periods and explosive dynamics in others. Under such a scenario, the standard full-sample ADF model coefficient estimates are biased due to model-misspecification, and simulations based on those coefficients cannot replicate the behavior of actual house prices. However, the regression specification that we focus on is not the full-sample ADF but the more flexible recursive ADF regression specifications proposed by Phillips and Yu (2011) and Phillips *et al.* (2015).

The beta estimates in the recursive ADF regression change over time, taking values in the stationary, I(1), and mildly explosive regions—these changes, in turn, can be detected to conventional significance levels with the SADF and GSADF tests proposed by the same authors. It is in this sense that this recursive ADF regressions are more flexible than the full-sample ADF regression, allowing our simulation exercise to better replicate very rich dynamics such as the ones displayed by real house prices.

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.

More specifically, our findings reveal a decline in power of both the SADF and GSADF due to aggregation and very large differences in the performance of the two tests when applied to aggregated data. 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.* (2015), whenever possible. Our findings also indicate 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 whenever the researcher is restricted to using aggregated data for detecting explosive behavior in time series.

5. Concluding Remarks

In this paper, we explore the relevance of aggregation for detecting periods of exuberance (mildly explosive) with two applications based on real house price data disaggregated by location. We find that by exploiting disaggregated data, we can more successfully identify such episodes. Therefore, we also 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.* (2015), 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. Our findings suggest that the recursive implementation of the (right-tailed) unit root test is key to lessening the power loss associated with the use of aggregated data. Thus, the GSADF test should be preferred to the SADF test (and especially 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.

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.

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.

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.

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.

Im, K. S., Pesaran, M. H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115 (1), 53–74.

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. December, Federal Reserve Bank of Dallas.

Nielsen, B., 2010. Analysis of coexplosive processes. Econometric Theory 26 (3), 882–915.

Pavlidis, E., Yusupova, A., Paya, I., Peel, D., Martínez-García, E., Mack, A., Grossman, V., 2015. Episodes of exuberance in housing markets: In search of the smoking gun. The Journal of Real Estate Finance and Economics, pp. 1-31.

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

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.

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.

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. Supplementary Materials

A. Recursive Implementation of the Tests: A Graphical Illustration

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 *et al.* (2011)

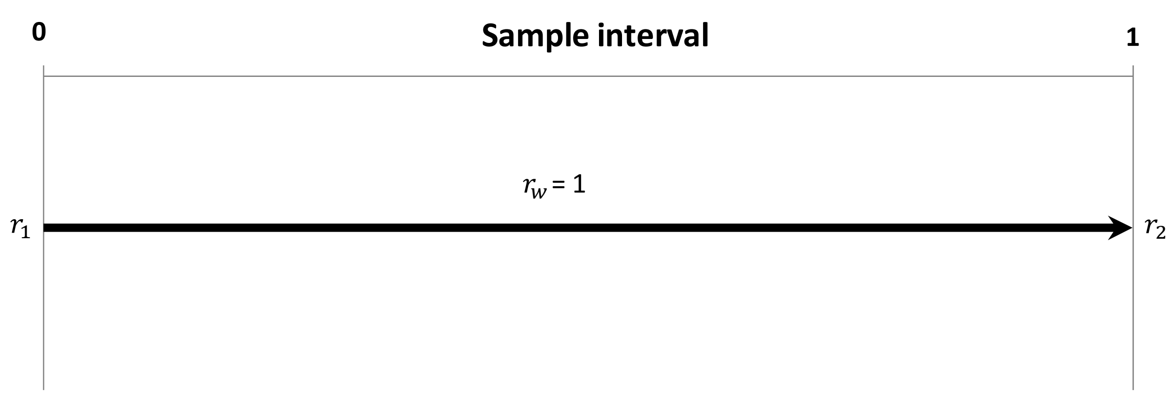
3. Generalized SADF (GSADF), see Phillips *et al.* (2015)

4. Panel GSADF, see Pavlidis *et al.* (2015)

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).

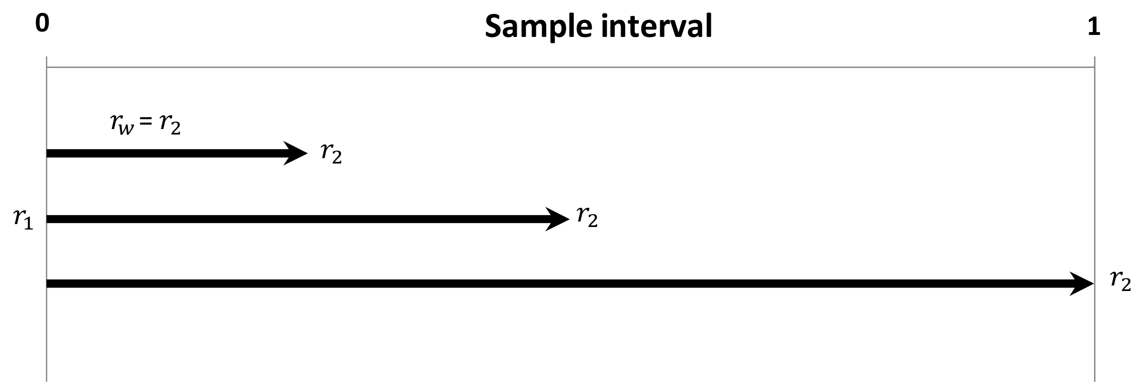
The first test is a right-tailed version of the standard ADF unit root test. With the given notation for a generic recursive mechanism, the implementation of the ADF test can be represented graphically simply as follows:

Illustration of the ADF Procedure



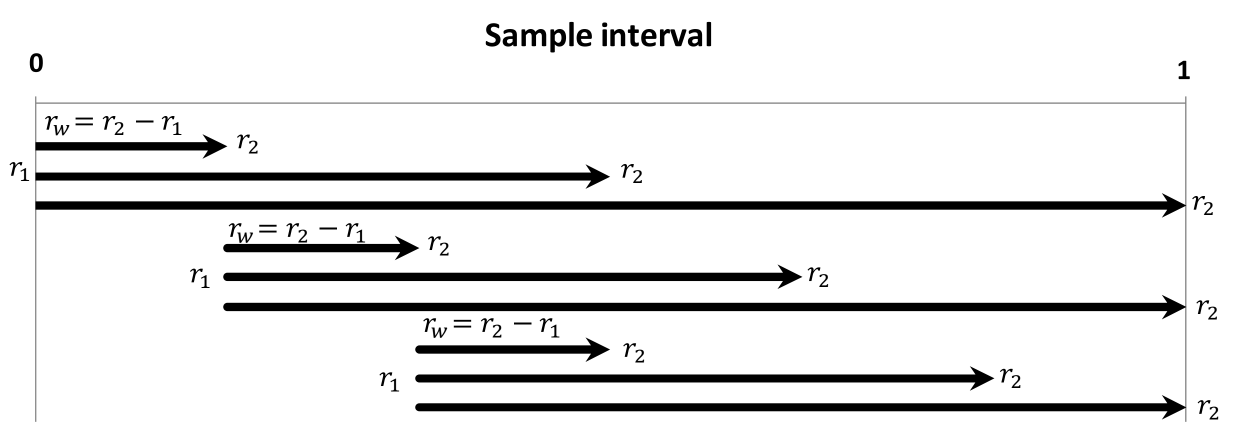
The SADF test is based on a proper recursion mechanism based of the ADF test statistics with an expanding window. The recursion mechanism goes as follows in this case:

Illustration of the SADF Procedure



The SADF test suffers from a loss of power in the presence of multiple periodically-collapsing occurrances of mildly explosive behavior. As a prefered alternative, Phillips *et al.* (2015) 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.* (2015). Formally, the GSADF test recursion can be illustrated as follows:

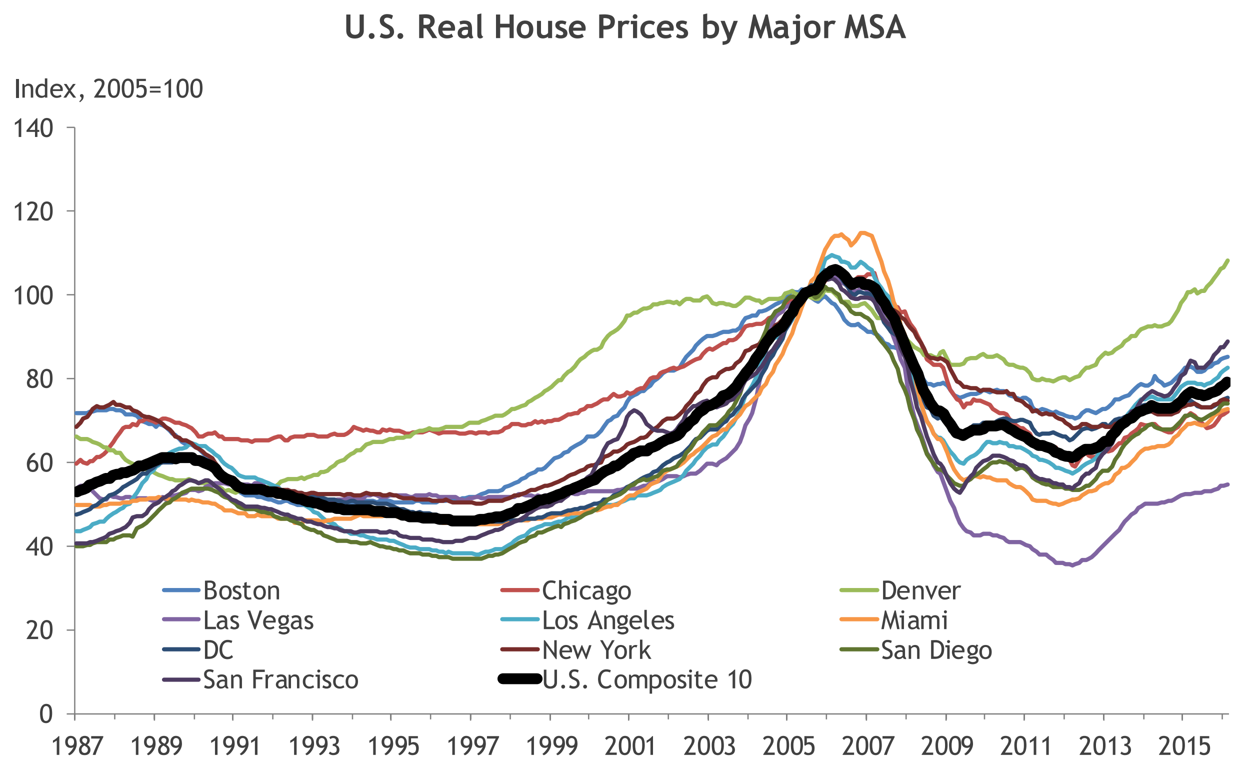
Illustration of the GSADF Procedure



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.

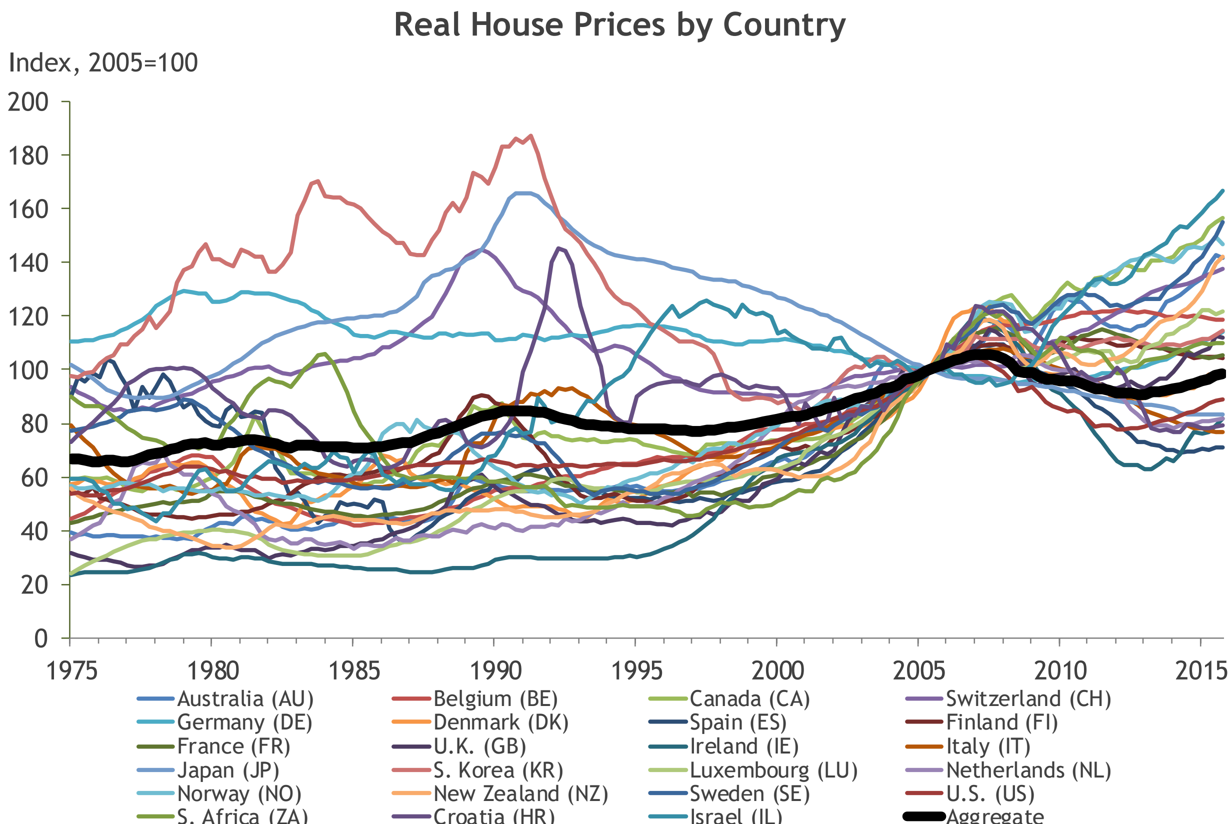
B. The Dataset of Real House Prices: Local and National Series

We examine the role of cross-sectional aggregation by conducting two large simulation experiments. The first experiment 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 data is illustrated graphically here:



Notes: Monthly house prices for the S&P/Case-Shiller 10-City Composite Home Price Index and its constituents, deflated by its corresponding local CPI. The sample covers the period from January 1987 to February 2016.
Sources: S&P Dow Jones Indices; Bureau of Labor Statistics.

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 data is illustrated graphically here:



Notes: Covers 23 countries, deflated with the country PCE deflator.
Source: Federal Reserve Bank of Dallas (Mack and Martínez-García (2011)).

C. Mapping the Power Gains of the GSADF Test Procedure Over the SADF Test Procedure

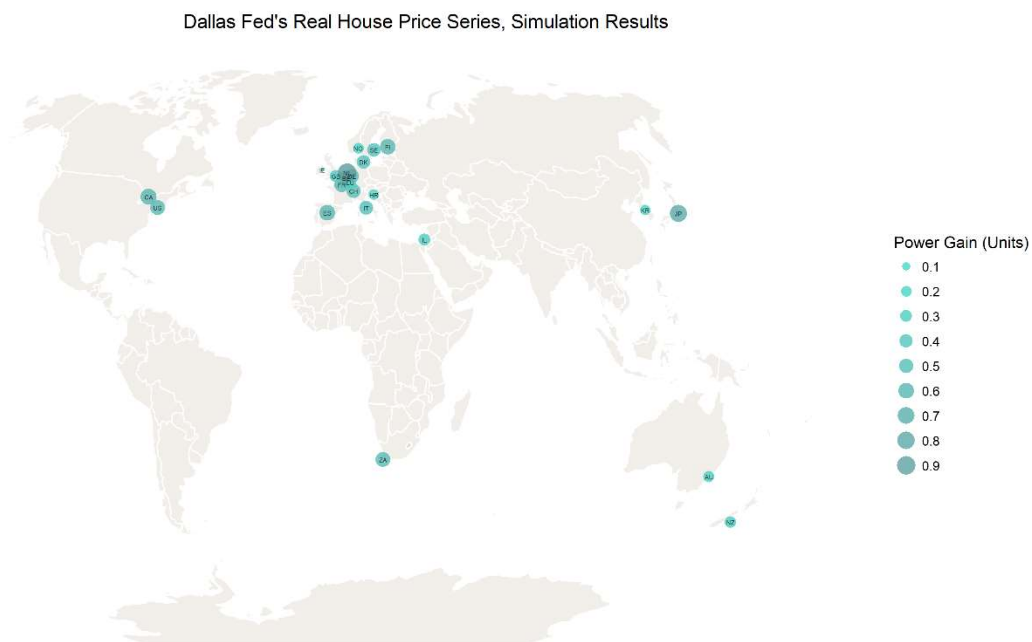
In accordance with the simulation results reported in Chart 1 for our first experiment based on real house prices for the major metropolitan statistical areas (MSAs) in the U.S., we summarize the power gains achieved by the GSADF test over the SADF test for each one of the constituent series in the following map:

S&P/Case-Shiller 10-City Composite Home Price Index, Simulation Results

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.

A striking contrast emerges from looking at Boston and Miami for which the power gain of the GSADF test is under 20 percent (0.2 in units) while the power gain exceeds 80 percent (0.8 in units) for Los Angeles (LA). 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 but with significant variation across MSAs.

In accordance with the simulation results reported in Chart 2 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 (Mack and Martínez-García, 2011), we summarize the power gains achieved by the GSADF test over the SADF test for each one of the constituent series in the following map:



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.

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 percent (0.2 in units) while the power gain exceeds 80 percent (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 with national indexes, but still find significant variation in the realized gains across locations.