Bubble Detection and Sector Trading in Real Time*

George Milunovich
Macquarie University

Shuping Shi
Macquarie University

David Tan
University of New South Wales

Abstract

We conduct a pseudo real-time analysis of the existence and severity of speculative bubbles in eleven US sectors over the period 1973 – 2015. Based on the bubble signals, a trading strategy is constructed which switches funds between the market index and those industry sectors that exhibit bubble dynamics. Our strategy generates the highest after-transaction-cost return and Sharpe ratio, and first-order stochastically dominates three other investments (including two alternative active strategies as well as the buy-and-hold investment in the market index). Subsample analysis and specification checks confirm the robustness of the findings.

Keywords: Speculative bubbles, price-earnings ratio, explosive dynamics, real-time trading, stochastic dominance

JEL classification: C12, C22, G01, G11

*George Milunovich, Department of Economics, Macquarie University; Email: george.milunovich@mq.edu.au. Shuping Shi, Department of Economics, Macquarie University; Email: shuping.shi@mq.edu.au. David Tan, School of Aviation, University of New South Wales; Email: david.tan@unsw.edu.au.

I. Introduction

In light of the recent financial crisis, much effort has been devoted to developing methods for detecting asset price bubbles. Although the derivations of alternative tests invoke somewhat different assumptions, the question of whether speculative bubbles are detectable in their inflationary phase is a key issue for all such procedures. Construction of new tests, advances in computational power and greater availability of data have led to improvements in the bubble detection methodology, and made such detection feasible in real time. Given these recent developments, market participants may wonder if it is possible to exploit bubble detection techniques for the purpose of active portfolio management. The issue of how an investor should act when she has detected a bubble in its expansionary stage becomes important. According to the efficient markets theory, rational investors are expected to short assets they know to be overpriced. However, some models suggest that investors may want to ride bubbles[1] for a period of time, before selling them (see Wurgler and Zhuravskaya (2002) and Abreu and Brunnermeier (2002, 2003), for instance). Whether such a strategy can consistently yield abnormal returns is an open empirical question which we address in this paper.

To begin with, we test for the presence of speculative bubbles in the US sector-level indices using a novel approach proposed by Phillips, Shi, and Yu (2015a,b, PSY hereafter). While most of the current literature on price bubbles relates to market-wide indices, questions regarding bubble formation at industry level concern both market participants and policy makers. For instance, it is unclear whether there is any variation in susceptibility to bubble dynamics across industry sectors, or what the frequency of bubble coincidences among different sectors may be. In this paper the analysis is conducted at the weekly frequency. This is to increase the chance of not entering a trading position too late in a bubble cycle on the one hand, and to avoid jumps present in daily data on the other (Eraker et al., 2003). One prominent feature of weekly financial data is heteroskedasticity or time-varying volatility (Engle (1982) and Bollerslev (1986)). This feature challenges the performance of the PSY test, which is designed primarily for monthly frequency data. Harvey et al. (2015) argue that the presence of heteroskedasticity increases the chance of drawing a false positive conclusion of bubble existence. Therefore, we implement the PSY bubble detection technique using a wild bootstrapping procedure, which produces heteroskedasticity consistent critical values as in Etienne et al. (2014).

Next, trading strategies are developed conditional on the bubble signals provided by the PSY test for sector-level indices. We construct two bubble indicators: (i) an indicator based on the original PSY test (PSY-BI), and (ii) a modified bubble indicator (MBI)

[1] A simple strategy would be to long the asset when the bubble is growing and to short it just prior to the collapse.

combining the original PSY bubble signal with a return sign signal. Despite the superior performance of the PSY test relative to other real-time bubble detection techniques, there is still a possibility of drawing false positive conclusions, as well as of experiencing delays in the estimation of bubble origination and termination dates, see e.g. Phillips et al. (2015a) and Phillips and Shi (2014). The modified indicator is expected to reduce the chance of entering a sector due to a false bubble detection, while improving the timeliness of exits from collapsing bubbles. We construct and rebalance equally-weighted bubble portfolios on the basis of such indicators. The performance of the timing strategies is benchmarked against the buy-and-hold strategy (BH) invested in a market index, and a strategy which is formed simply by relying on the direction of the most recent index movement – termed directional signal (DS) here. Profitability is assessed using a number of performance measures described below, and the strategies are ordered according to the stochastic dominance (SD) criterion. The SD framework allows us to rank the constructed portfolios without making explicit assumptions about the specification of the asset pricing model, investor utility functions, or the distribution of asset returns.

Although there are relatively few studies that directly evaluate the performance of bubble riding strategies, examples of profitable bubble trades go back to the 18th century. Temin and Voth (2004), for instance, provide an interesting account of the 1720–1721 trades made by Hoare's Bank in the South Sea Company bubble, which amounted to one of the most successful episodes in the history of speculation. More recent examples are studied by Brunnermeier and Nagel (2004), who examine stock holdings of hedge funds during the technology bubble of 1998–2000. They report that hedge funds managed to capture the upturn, and rode the bubble by having portfolios heavily skewed in favour of technology stocks. Interestingly, the funds were also able to time the collapse of the bubble, reducing their holdings about 6 months before the peak of the bubble. Focusing on the S&P 500 index over the 1946 – 2003 period, Brooks and Katsaris (2005) model asset price bubbles with a regime-switching model and examine its financial usefulness to generate trading rules. Their investor either buys (sells) the S&P 500 Composite Index when the probability of a rally (crash) is high, or invests in the 3-month U.S. Treasury Bill. Relative to alternative trading rules considered, Brooks and Katsaris' strategy produces higher Sharpe ratios and end-of-period wealth. Similarly, Guenster and Kole (2013) employ a regime-switching model to identify bubbles in the US industry portfolios. Their analysis suggests that investors who rebalance portfolios at a frequency of less than four months should optimally ride bubbles, whereas long-term investors with rebalancing horizons longer than six months should short asset bubbles.

Unlike the existing literature, our trading strategy is based on bubble indicators provided by the PSY test with a wild bootstrapping procedure for producing critical values. In essence, the PSY method dates inflationary stages of speculative bubbles to periods in time when the law of motion changes from martingale behaviour to explosive dynamics. We choose to employ the PSY approach over alternative procedures for a number of reasons. Most importantly, the PSY method is a real-time detection technique that mimics how a trader would access her bubble-timing information in practice. It has also been shown to exhibit a superior performance relative to several other real-time bubble detection techniques, including the recursive test of Phillips et al. (2011) and the CUSUM strategy of Homm and Breitung (2012), both asymptotically and in finite samples. Further, the PSY test is relatively easy to implement compared with tests where numerical optimization is required, such as the Markov-switching test of Hall et al. (1999) and the regime switching methods of van Norden (1996) and Brooks and Katsaris (2005). Nevertheless, our preferred trading strategy employs a modified bubble indicator which we construct by combining the inference provided by the PSY test with a one-day return sign indicator to generate trading signals. The direction of the most recent index movement appears to contribute information to the MBI that is not captured by the original PSY test, as discussed in more detail below.

We employ the bubble indicators to create and rebalance equally-weighted portfolios under real-world conditions. Specifically, we consider an investor who trades at a weekly frequency, observing the trading signal at the end of the trading day on Tuesdays and executing trades on Wednesdays. The investor holds the market portfolio, and buys (sells) those industry sectors which are found to enter the expansionary (deflationary) bubble stage. All positions are equally weighted and rebalanced as new information becomes available. A moderate round-trip transaction cost of 0.5% of the weekly real return is imposed. As mentioned, four trading strategies are considered: PSY-BI, MDI, BH and SD, and after transaction cost returns and final period wealth figures on initial $1 investments are computed. The last strategy is designed to gauge the profitability of the information content present in the sign of the most recent index return, in isolation of the PSY method.

After-transaction-cost performance is assessed on the basis of final period wealth, average returns, and Sharpe ratios. In addition, a ranking of trading strategies is provided according to the stochastic dominance (SD) criterion. The SD approach evaluates entire distributions of asset returns instead of performing simple comparisons of average values and variances. Unlike the mean-variance analysis, which assumes either that returns are normally distributed or that investors have quadratic utility functions, SD does not make such explicit assumptions. Nevertheless, the framework is consistent with expected utility maximizing behaviour under minimal assumptions about investor utility functions. For instance, non-satiation is the only investor preference supposed under first-order stochastic dominance. Since the SD approach is not dependent on a specific equilibrium model for asset prices, instances of stochastic dominance of one strategy over alternative investments are less likely to be caused by omissions of relevant risk factors from benchmark pricing models. In the literature, tests of stochastic dominance are found in a range of financial applications. Levy (1985), for instance, applies stochastic dominance to study option pricing, Seyhun (1993) investigates the small firm effect, Post (2003) studies the portfolio choice, Abhyankar et al. (2008) examine value versus growth firms, while Fong et al. (2005) test momentum strategies.

We implement stochastic dominance tests as generalizations of the Kolmogorov-Smirnov test for stochastic dominance, namely the procedures proposed in Linton et al. (2005), hereafter referred to as the LMW tests. LMW tests are consistent against all alternatives and apply to testing stochastic dominance at any arbitrary order in the K-variable ($K \geq 2$) case. This facilitates testing of stochastic dominance of the first- (SD1), second- (SD2) and third-order (SD3) across the pairs of investment strategies, as well as jointly. The tests also allow for serial dependence in observations, and for general dependence amongst tested variables. This is important for two reasons: (i) explosive dynamics in the price-earnings ratio result in autocorrelated returns series, and (ii) portfolios are constructed from the same underlying set of sector indices making them unlikely to be mutually independent. Lastly, the LMW methodology allows for tested series to be formed as residuals from conditional models, covering the case of portfolios constructed on the basis of the PSY test inference.

Our findings are as follows. Episodes of bubble behaviour are found in all sector portfolios, although their frequency varies notably across the indices. The timing of detected bubbles across sector indices coincides over periods of major market crashes such as 1987 and 2000, while the coincidences appear less evident during periods of normal activity. A trading strategy employing our modified bubble indicator significantly outperforms the investments based on the original PSY bubble indicator, the directional signal, and the buy-and-hold portfolio. Specifically, final period wealth (in real terms) generated by MBI is more than twice the wealth accumulated by the runner-up, while its Sharpe ratio is nearly 2 percent higher than that of the second highest Sharpe ratio. Pairwise and joint tests of first-, second-, and third-order stochastic dominance indicate that MBI ranks ahead of the other three strategies at the 5 percent level. We analyse the consistency of the outperformance of MBI relative to BH by considering rolling window investments of durations between 1 and 5 years. It appears that the MBI strategy outperforms (underperforms) between 35 (22) percent of the one-year subsamples, and 77 (21) percent in the case of the five-year horizons. Next, given that the constructed bubble indicators rely on statistical inference at the 99 percent confidence level, we are also interested in the sensitivity of the results to alternative levels of confidence. Thus we repeat the analysis of our preferred MBI strategy employing the 95 and 90 percent critical values in the PSY tests. When transaction costs are taken into account, the 99 percent confidence level tests appear to generate highest end-of-period wealth, real average annual returns and Sharpe ratios. The 99 percent confidence strategy also ranks best according to the SD tests. On the other hand, without transaction costs, the 90 percent confidence level outperforms according to all criteria, demonstrating the effect of trading frequency and transaction costs. Lastly, we consider two further modifications of the MBI strategy. First, the bubble indicator is augmented with a longer history of directional signals. Second, instead of retreating into the market index when no bubbles are found in sector indices, we invest in those industries which have not experienced a bubble episode for a certain period of time. The original MBI strategy beats these modifications according to all performance criteria, and is found to dominate them at the first-, and hence all orders of stochastic dominance. This demonstrates the robustness of the MBI bubble indicator across subsamples, as well as in relation to alternative confidence levels and to simple modifications of the trading strategy.

The rest of the paper is organized in the following order. In Section 2 we discuss empirical methods employed in the paper, including the PSY test for bubble detection, our trading strategy and a modification of the PSY procedure that it employs. Stochastic dominance tests used to assess portfolio performance are also presented here. In Section 3 we discuss the dataset, and provide empirical evidence on the detection of bubbles in the US sector-level indices. Section 4 assesses the performance of the proposed strategies, and Section 5 concludes.

II. Methodology: Bubble Detection, Trading Strategies and Stochastic Dominance

We describe the PSY test for detecting bubbles and discuss a wild bootstrap procedure that is used to obtain heteroskedasticity robust critical values. Following this, we propose several trading strategies, one of which exploits bubble signals identified by the PSY test, while another proposes a modification of the PSY method. Lastly, we discuss the notion of stochastic dominance and outline the SD tests that are employed to assess the effectiveness of the trading strategies.

A. Bubble Detection in Real Time

Diba and Grossman (1988) argue that in the presence of speculative bubbles prices are expected to exhibit explosive and periodically collapsing behaviour, as opposed to being martingales, which are observed during normal periods. The detection procedure of PSY is based on these two dynamic characterisations of asset prices. In particular, the question of interest is whether a particular observation comes from an explosive process (H_A) or from a normal – martingale – behaviour (H_0). The testing algorithm is based on a right-tailed unit root test proposed by Phillips et al. (2014). The regression model is

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{K} \gamma_i \Delta y_{t-i} + \varepsilon_t, \tag{1}$$

where y_t is the stock price at period t, K is the lag order (set to one in the application), and ε_t is the error term. The ADF statistic is defined as the t-ratio of the OLS estimate of β.

The PSY test requires conducting subsample regressions. Let τ_1 and τ_2 be the starting and ending points of a subsample regression with the corresponding ADF statistic denoted by $ADF_{\tau_2}^{\tau_1}$. The algorithm calculates the ADF statistic repeatedly on a sequence of backward expanding samples. Suppose τ is the observation of interest. The ending points of all samples are fixed on τ and the starting point of the samples varies from the first observation to $\tau - \tau_0$, where τ_0 is the minimum window size required to initiate a regression. The corresponding ADF statistic sequence is $\left\{ADF_{\tau_2}^{\tau_1}\right\}_{\tau_2=\tau}^{\tau_1 \in [1, \tau-\tau_0]}$. Inference of explosiveness for observation τ is based on the maximum value of the ADF sequence, denoted by $MADF_\tau$ and defined as

$$MADF_\tau = \max\left\{ADF_{\tau_2}^{\tau_1} : \tau_2 = \tau \text{ and } \tau_1 \in [1, \tau - \tau_0]\right\}.$$

For practical implementation, Phillips et al. (2015a) suggest setting τ_0 according the rule of $\tau_0 = (0.01 + 1.8/\sqrt{T}) \times T$ to reduce the probability of size distortion, especially in the presence of conditional heteroskedasticity. Nevertheless, Harvey et al. (2015) demonstrate by simulations that in the presence of non-stationary volatility, the size of the Phillips et al. (2011) procedure, which is a special case of the PSY procedure, is substantially above the nominal level, indicating a serious size distortion. A wild bootstrap procedure is shown to be asymptotically valid and is able to effectively control finite sample size under non-stationary volatility, and a similar conclusion is expected in our case.

While conditional heteroskedasticity is a widely recognized feature of many financial data series, non-stationary volatility, such as volatility shifts and trending volatility, is also not uncommon. Therefore, to reduce the chance of size distortion, we obtain critical values of the $MADF_\tau$ statistic using a wild bootstrapping procedure. The procedure is implemented as follows.

Step 1. Estimate the ADF model under the null hypothesis that $\beta = 1$ using the whole sample period

$$\Delta y_t = \alpha + \sum_{i=1}^{K} \psi_i \Delta y_{t-i} + \varepsilon_t,$$

where $t = K+2, ..., T$ and ε_t is the error term. We obtain the OLS parameter estimates $\hat{\alpha}$ and $\hat{\psi}_i$ and the $(T-K-1) \times 1$ dimension residuals denoted by e_t.

Step 2. Generate the bootstrap residuals e_t^b according to the device $e_t^b = \tilde{e}_t w_t$, where $\tilde{e}_t := e_j$, $\{w_t\}_{t=1}^{T-K-1}$ denotes an independent $N(0,1)$ scalar sequence, j is a random number generated from a uniform distribution running between $K+2$ and T, and $t = K+2, ..., T$. Conditional on e_j, e_t^b is independent over time with zero mean and variance e_j^2. The multiplicative factor w_t serves to replicate the pattern of heteroskedasticity present in the original shocks.

Step 3. The bootstrap sample is generated as follows. Let $\Delta y_t^b = \Delta y_t$ for $t = 2, ..., K+1$, which is obtained by

$$\Delta y_t^b = \hat{\alpha} + \sum_{i=1}^{K} \hat{\psi}_i \Delta y_{t-i}^b + e_{t-K}^b, \text{ for } t = K+2, ..., T.$$

We then calculate y_t^b as

$$y_t^b = y_1 + \sum_{j=1}^{t} \Delta y_j^b, \text{ for } t = 2, ..., T\,.$$

Step 4. Calculate the MADF statistics for the bootstrapped data series, $MADF^b$. We repeat the procedure 500 times and obtain a sequence of the bootstrapped test statistic $\left\{MADF^b\right\}_{b=1}^{500}$. The 90%, 95%, 99% bootstrapped critical value is calculated as the 90%, 95%, 99% percentiles of the statistic sequence.

B. Trading Strategies

In order to assess the practical significance of the signals produced by the PSY test, we develop a trading strategy that allocates funds across the US sector indices conditional on the computed bubble indicators. However, before we describe the strategy in more detail we discuss the issue of false bubble detection, which has the potential to adversely impact trading results. We also propose an approach that will reduce its occurrence.

As noted in Phillips et al. (2015a) and Phillips and Shi (2014), under certain conditions, the PSY test and similar procedures may lead to a false identification of bubble episodes on the one hand, and delays in detecting bubble termination dates on the other. For instance sudden breaks, such as jumps in the series under investigation, may trigger the PSY bubble indicator. This happens because of the changes in volatility which such shifts entail, and the fact that the MADF statistics are calculated from a model that assumes constant volatility within each subsample period. Although we guard against the effects of time-varying volatility by bootstrapping the PSY test, in order to reduce the possibility of trading on an erroneous signal, we also construct a *modified bubble indicator*. The proposed indicator combines the inference obtained from the PSY test with the direction of the most recent index movement.

In comparison to the PSY bubble indicator, which takes the value one if the bootstrapped test statistic $MADF_t$ is greater than the bootstrapped critical value scv_t

$$I_t = 1(MADF_t > scv_t), \tag{2}$$

and zero otherwise, the modified indicator I_t^* requires two conditions to be satisfied simultaneously:

$$I_t^* = 1(MADF_t > scv_t \ \& \ R_t > 0), \tag{3}$$

and zero otherwise. Thus, the modified indicator augments the inference provided by the PSY test by taking into account the sign of the most recent return R_t. In fact, as seen in (3), R_t is required to be positive in order for the inference made by the PSY test to be validated. Intuitively, by considering the direction of the most recent index movement, we hope to reduce the likelihood of producing a false signal initiating a bubble trade. Similarly, the indicator is quicker to reset back to zero, which is likely to result in more timely exists from collapsing bubbles.

We note that all indicators are calculated using weekly data observed at the end of trading day on Tuesdays. Given that t here denotes time measured in weeks, the indicators given in (2) and (3) are known to traders rebalancing their portfolios on Wednesdays. We consider four trading strategies, which are described next.

Strategy 1. Modified bubble indicator (MBI) strategy

This strategy entails buying those sectors for which the modified bubble indicator given in (3) switches from zero to one, and closing the positions when the MBI resets back to zero. If more than one sector exhibits explosive dynamics at the same time, funds are allocated in equal proportions across all such indices. If no bubble is identified, funds are shifted into the market index.

Strategy 2. PSY bubble indicator (PSY-BI) strategy

This strategy is similar to Strategy 1 but uses the original PSY bubble indicator described in (2), instead of the MBI indicator in (3). Specifically, investors apply equal weights to the sectors for which the PSY indicator takes the value of one, and close their positions as the indicators switch back to zero. If no bubbles are detected we hold the market portfolio.

Strategy 3. Directional signal (DS) strategy

The difference between strategies 1 and 2 is that while the PSY-BI strategy relies only on the PSY test to time bubble sectors, the MBI strategy also requires a validation of the PSY inference by the direction of the most recent index movement. Thus, one may wonder if it is the direction of the index movement itself that drives the profitability of the MBI strategy, rather than the joint condition described in (3). For this reason we decide to test a directional strategy that only depends on the directional indicator formulated as $I_t^D = 1(R_t > 0)$, and zero otherwise. Thus, the investor opens positions in sectors for which $I_t^D = 1$, and reverses the trades upon the return of the indicator to zero. If there is more than one sector for which the directional indicator is positive, we allocate funds in equal amounts to all such indices. Lastly, we hold the market index if $I_t^D = 0$ for all sectors.

Strategy 4. Buy-and-hold (BH) strategy

We compare the above three active strategies with the buy-and-hold benchmark denoted BH. This strategy is executed by entering an open position in the market portfolio at the beginning of the sample, and closing it at the end of the sample period.

C. Stochastic Dominance

While Section 1 highlights some of the advantages of the stochastic dominance approach over traditional performance measures, in this section we provide more specific definitions, discuss the literature, and present test statistics used in the subsequent analysis.

C.1. Definitions of Stochastic Dominance

Stochastic dominance provides a method to rank two investments, say X_ℓ and X_p, by considering the integrals of their cumulative distribution functions $F_\ell(z)$ and $F_p(z)$. Equivalently, denoting the class of all increasing von Neumann-Morgenstern type utility functions by $\mathcal{U}_1$, i.e. $u^{'} \geq 0$ for every $u \in \mathcal{U}_1$, we may define the notions of the first-, second-, and third-order stochastic dominance with reference to $\mathcal{U}_1$, and its subsets.

Formally X_ℓ first-order stochastic dominates (SD1) X_p if and only if either: (i) $F_\ell(z) \leq F_p(z)$ for all z, with strict inequality for some z; or (ii) $E[u(X_\ell)] \geq E[u(X_p)]$ for all $u \in \mathcal{U}_1$ and with strict inequality for some u. SD1 implies that the expected utility of holding investment X_ℓ is at least as great as that derived from holding X_p, and that for some utility functions in $\mathcal{U}_1$ it is greater. Thus, under the assumption that more wealth is better than less (increasing utility function), the ordering of investments is unambiguous for all expected utility maximizing investors. The notion of the second-order stochastic dominance focuses on a strictly concave subset of increasing utility functions $\mathcal{U}_2 \subset \mathcal{U}_1$, i.e. $u'' \leq 0$ for all $u \in \mathcal{U}_2$. More specifically, X_ℓ second-order stochastic dominates (SD2) X_p if and only if either: (i) $\int_{-\infty}^{z} F_\ell(t)dt \leq \int_{-\infty}^{z} F_p(t)dt$ for all z and with strict inequality for some z; or (ii) $E[u(X_\ell)] \geq E[u(X_p)]$ for all $u \in \mathcal{U}_2$, with strict inequality for some u. Under SD2 the area under F_ℓ is everywhere smaller than the corresponding area under F_p, or equivalently, the investors are required to be risk averse – they are described by monotonically increasing and concave utility functions. Finally, third order-stochastic dominance introduces a further requirement that investors have preference for investments which are positively skewed. This is accomplished by further restricting the set of utility functions to $\mathcal{U}_3 \subset \mathcal{U}_2$ in which $u''' \geq 0$ for all $u \in \mathcal{U}_3$. We say that X_ℓ third-order stochastic dominates (SD3) X_p if and only if either: (i) $\int_{-\infty}^{z}\int_{-\infty}^{k} F_\ell(t)dtdk \leq \int_{-\infty}^{z}\int_{-\infty}^{k} F_p(t)dtdk$ for all z with strict inequality for some z; or (ii) $E[u(X_\ell)] \geq E[u(X_p)]$ for all $u \in \mathcal{U}_3$, with strict inequality for some u. It is also of interest to note that SD(j) implies SD($j+1$) such that, for instance, SD(1) implies SD(2) and SD(3).

Following the literature, see e.g. Davidson and Duclos (2000), and Barrett and Donald (2003), we represent the orders of stochastic dominance using the operator $\mathcal{I}_k^{(s)}(.;F_k)$ which integrates the function F_k to the order $s-1$. For $F_k(z) = P(X_k \leq (z))$ we let

$$\mathcal{I}_k^{(1)}(z;F_k) = F_k(z);$$

$$\mathcal{I}_k^{(s)}(z;F_k) = \int_{-\infty}^{z} \mathcal{I}_k^{(s-1)}(t,F_k)dt \qquad for\ s \geq 2,$$

such that X_ℓ stochastically dominate X_p at the order s if $\mathcal{I}_\ell^{(s)}(z;F_\ell) \leq \mathcal{I}_p^{(s)}(z;F_p)$. Equivalently, using the difference operator

$$\mathcal{D}_{\ell p}^{(s)}(z) = \mathcal{I}_\ell^{(s)}(z;F_\ell) - \mathcal{I}_p^{(s)}(z;F_p) \qquad for\ s \geq 1,$$

X_ℓ is said to stochastically dominates X_p if $\mathcal{D}_{\ell p}^{(s)}(z) \leq 0$ for all z.

Defining the functional $d_1^{(s)} = \sup_{z\in\Lambda} \mathcal{D}_{\ell p}^{(s)}(z)$, where Λ represents the union of the supports of X_ℓ and X_p this is equivalently expressed by the following hypotheses:

$$H_0^{(s)} : d_1^{(s)} \leq 0 \quad \text{vs.} \quad H_A^{(s)} : d_1^{(s)} > 0. \tag{4}$$

While the above formulation provides *pairwise* stochastic dominance relationships, we may extend the analysis beyond the bivariate case by considering K investments denoted $X_1, X_2, \ldots, X_K$. The hypothesis that we wish to test in this *multivariate* framework is whether a tested variable X_ℓ stochastically dominates all other variables $\{X_k\}_{k=1;k\neq\ell}^{K}$ as expressed in the following hypotheses

$$H_0 : d_2^{(s)} \leq 0 \quad \text{vs.} \quad H_A : d_2^{(s)} > 0, \tag{5}$$

for $d_2^{(s)} = \max_{k:k\neq\ell}\sup_{z\in\Lambda}[\mathcal{D}_{\ell k}^{(s)}(z)]$. In this case the maximum in $d_2^{(s)}$ is taken over all $K-1$ pairs including X_ℓ.

C.2. Tests for Stochastic Dominance

Statistical tests for stochastic dominance replace the true, but unknown, cumulative distribution functions by their empirical analogues. The tests may be broadly classified as the tests which are consistent against all alternative hypotheses, and the tests which have power against a finite dimensional class of alternatives. The possibility for test inconsistency arises for two reasons: (i) stochastic dominance tests involve composite hypotheses on inequality constraints, and (ii) some tests conduct comparisons of distributions at a fixed number, rather than at all, points in the support of the distributions. As Barrett and Donald (2003) note, tests based on comparisons at all points in the variable range have the potential for being consistent for the full set of restrictions implied by stochastic dominance.

Amongst generalizations of the Kolmogorov-Smirnov test, which are based on comparisons at all points in the support, McFadden (1989) proposes tests of SD1 and SD2 under the assumptions of independence between tested variables and i.i.d. observations. His tests are implemented via a Monte Carlo-based computation of approximate critical values. Klecan et al. (1991) generalize the results of McFadden by relaxing the assumptions of i.i.d. observations. Barrett and Donald (2003) extend the analysis to consider tests for stochastic dominance of any prespecified order, and also allow for different sample sizes of the tested variables. They provide consistent multiplier and bootstrap methods which have an asymptotically exact size on the least favourable points in the null hypothesis. However, their approach assumes that the test variables are mutually independent, and the observations are i.i.d.

In this paper we compare investment strategies using a generalization of the Kolmogorov-Smirnov test provided in Linton et al. (2005, LMW). These tests are consistent against all alternatives and apply to testing stochastic dominance at any arbitrary order in the K-variable ($K \geq 2$) case. Further, the tests allow for both serial dependence in observations, general dependence amongst tested variables, and for tested series to be constructed as residuals from conditional models.

The test statistics take the following form

$$T_{T,m}^{(s)} = \sqrt{N}\hat{d}_m^{(s)} \qquad \text{for } m = 1,2 \quad \text{and } s = 1,2,3,$$

where $\hat{d}_m^{(s)}$ represents a centered empirical analogue of the $d_m^{(s)}$ functional defined in (4)-(5) such that setting $m = 1$ provides a bivariate test, while $m = 2$ corresponds to the multivariate SD test statistic. N is the sample size here. Varying s between 1 and 3 leads to the tests of SD1, SD2 and SD3, respectively.

Large values for the above test statistics indicate evidence against the null. However, the distributions of the test statistics depend on the unknown distributions of the data and cannot be tabulated, although LMW show that the statistics converge to functionals of a Gaussian process. Following LMW, we obtain p-values using two resampling procedures based on subsampling and block bootstrap. We implement two rather than one procedure because, as LMW note, the subsampling-based test may exhibit more power for some local alternatives near the boundary between the null and alternative, while the bootstrap approach is preferred against some alternatives which are far from the boundary. The Appendix outlines the algorithms used to implement these two methods.

III. Data and Empirical Bubble Episodes

A. Dataset and Explosive Dynamics

We study eleven US sector indices as constructed by Datastream International. They are oil and gas, basic materials, industrials, consumer goods, health care, consumer service, telecom, utilities, technology, financials, and diversified REITs. The sample period runs from 02/01/1973 to 12/05/2015 (containing 2211 weekly observations), except for the diversified REITs sector which becomes available from 27/01/1998 (903 observations).

Our dataset contains: (i) Tuesday price-earnings ratios of the eleven sectors for bubble detection; (ii) Wednesday price indices and dividend yields for the eleven sectors as well as the Datastream market index; and (iii) the consumer price index (all urban, all item, non-seasonally adjusted). The consumer price index (CPI) is downloaded at the monthly frequency and linearly interpolated into weekly data. The series in (ii) and (iii) are employed to implement the trading strategies, and to compute real returns.

To test for explosive bubble dynamics, we apply the PSY strategy to the Tuesday price-earnings ratios of the eleven sectors. We use the price-earnings (P/E) instead of more conventional price-dividend ratios as in PSY and Phillips et al. (2011) because dividends are dependent on the dividend payout ratio determined by the firms' boards of directors. The payout ratios are influenced by firm size, profitability, growth opportunities, but also signalling (Bar-Yosef and Huffman (1986) and Denis and Osobov (2008)). For instance, during the dotcom expansion period in the mid-1990s, many technology companies experienced rapid growth, however no dividends were paid. Thus, earnings provide a more reliable proxy for stock market fundamentals. The minimum window τ_0 is set according to the rule of $0.01 + 1.8/\sqrt{T}$, containing 63 observations for the diversified REITs sector and 106 observations for all other sectors. Therefore, the test statistic and critical value sequences are computed from 31/03/1999 for REITs and from 7/01/1975 for the remaining sectors. Critical values are obtained from the wild bootstrap procedure described in Section A. The bubble indicator takes the value of one if MADF is above the 99% critical value, and zero otherwise, as described in (2). The price-earnings ratios and the bubble indicators are displayed in Figure 1 for the oil and gas, basic material, industrials, consumer goods, health care, and consumer service sectors and in Figure 2 for the telecom, utilities, technology, financials, diversified REITs sectors. We make a number of noteworthy observations on the basis of the information provided in the figures.

First, the procedure identifies instances of bubble episodes across all industry sectors, although there appears to be noticeable variation in their frequency. This is confirmed by the figures presented in Table 1. For example, the utilities and technologies sectors exhibited 23 and 26 bubble episodes, respectively,[2] over the whole sample period, while oil and gas and industrials seem to be less prone to explosive dynamics with 5 and 9 detected episodes, respectively. Second, the comovement of the sector-based indicators suggests a high degree of covariation of bubbles across the sectors. For example, all sectors (except for industrials) exhibited bubble episodes during 1987,[3] while the technology, financials, health care, utilities, oil and gas, and basic materials experienced the 1998-2000 bubble. Third, there are periods during which the timeliness and duration of the bubble indicators varies across sectors. Utilities and financials, for instance, identify the bubble episode related to the crash of October 1987 a considerable amount of time before the crash, whereas the oil and gas bubble indicator is activated only months prior to the collapse.

It appears that some sectors also experience delays in identifying bubble collapse dates. The indicator for the technology sector reports a bubble termination date in September 2000, when on visual inspection, the P/E ratio peaks in March 2000. On the other hand, several sectors (industrials, health care, consumer services, utilities, and technology) were falsely identified as being in a bubble during the 2008 crisis due to sudden drops in their P/E ratios. This is potentially due to large changes in the volatility of their P/E ratios, and the fact that the MADF statistics are calculated from a model which assumes constant volatility within each subsample period. Thus, although the wild bootstrapping procedure accounts for time-varying volatility it would seem that it was unable to correct for this particular case of false identification.

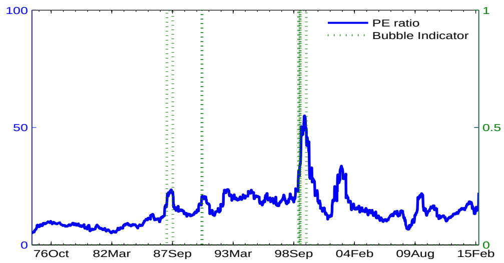
We conclude this section by providing some descriptive statistics about bubble durations in Table 1. As illustrated the number of bubble episodes ranges from 3 for diversified REITs to 26 for technology. The frequency of observations belonging to bubble regimes varies between 21 for telecom and 153 found in utilities. The median duration ranges from 1 to 22 weeks, with the maximum bubble duration of 64 weeks detected in consumer goods. Similarly, there is significant variation across the standard deviations computed

[2] A bubble episode is defined as a series of consecutive positive bubble signals, e.g. if the indicator switches from one to zero and then back to one we would record two bubble episodes.

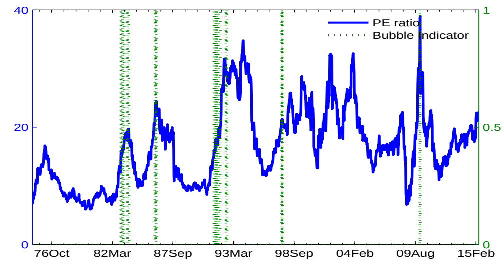
[3] We can not test for speculative bubbles in the diversified REITs sector during this period due to data unavailability.

Figure 1: The dynamics of the price-earning ratios and the bubble indicators (Part I)

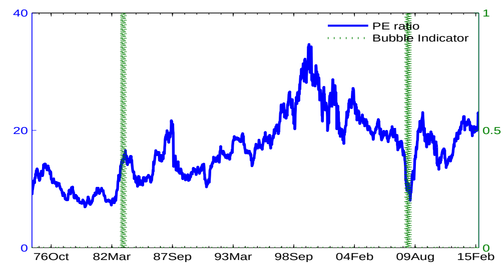
(a) Oil and gas



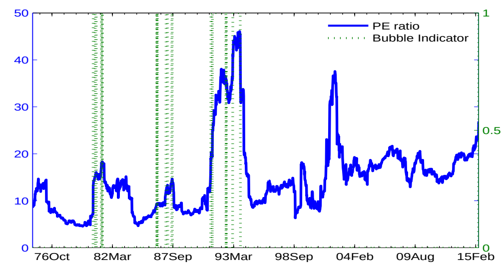
(b) Basic materials



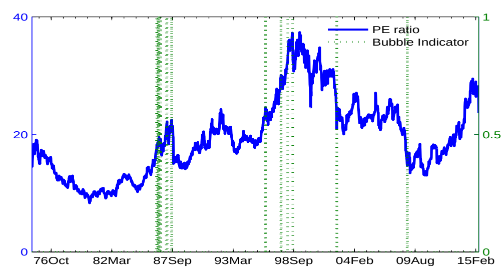
(c) Industrials



(d) Consumer goods



(e) Health care



(f) Consumer service

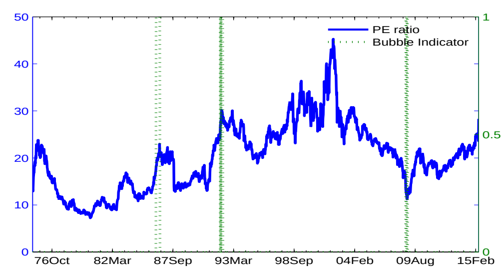
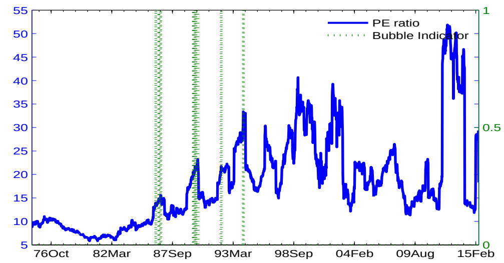
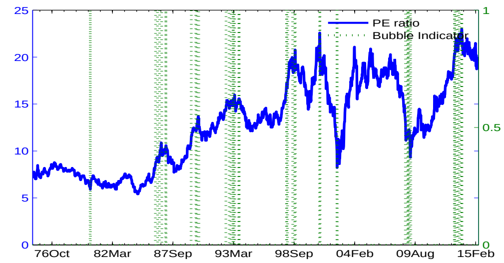


Figure 2: The dynamics of the price-earning ratios and the bubble indicators (Part II)

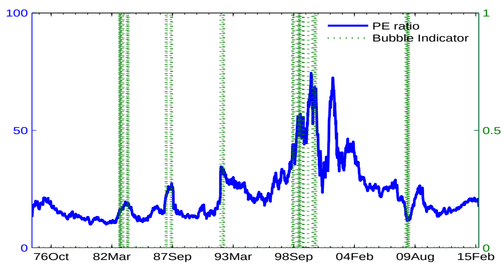
(a) Telecom



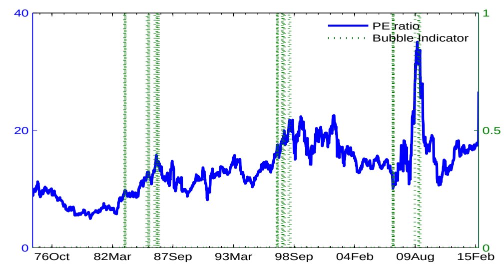
(b) Utilities



(c) Technology



(d) Financials



(e) Diversified REITs

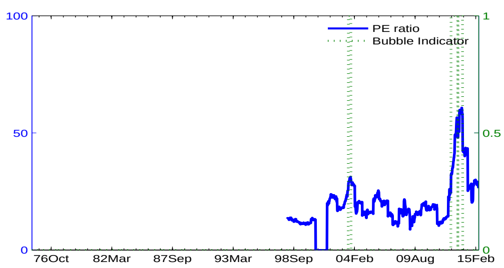


Table 1: Descriptive statistics of the bubble episodes

Industry Sector	# of bubble observations	# of bubble episodes	Duration of bubble episodes Median	Min.	Max.	Std.
Oil and gas	60	5	2	1	28	16.61
Basic materials	83	15	3	1	27	6.81
Industrials	25	9	2	1	5	1.72
Consumer goods	140	13	3	1	64	18.57
Health care	42	11	1	1	24	6.78
Consumer services.	36	6	4	1	18	6.42
Telecom	21	9	2	1	7	1.80
Utilities	153	23	4	1	24	6.54
Technology	135	26	2	1	20	5.82
Financials	63	16	1.5	1	17	4.49
Diversified REITS	67	3	22	13	32	9.50

Note: The bubble indicator is available from April 1999 for the Diversified REITs sector, and from January 1975 for all other sectors.

for bubble durations, with the telecom and consumer goods sectors exhibiting standard deviations of 1.8 and 18.57 weeks, respectively. This illustrates the disparities in the bubble-riding opportunities provided by different sectors.

IV. Trading Results and Performance Evaluation

Using the bubble signals computed from Tuesday closing prices we now employ Wednesday prices to implement the trading strategies described in Section 2.2. Real returns R_t are obtained as

$$R_t = (P_t + D_t)/P_{t-1} - \pi_t, \tag{6}$$

where P_t is the nominal price recorded at the end of the trading day on Wednesdays, D_t is the weekly dividend and $\pi_t = \Delta CPI_t/CPI_{t-1}$. To account for trading costs, we assume that a round trip of buying and selling incurs a trading cost of 0.5% of the real return, as in Brooks and Katasaris (2005).

A. Preliminary Analysis

Table 2 presents final wealth accumulated over the length of the sample period (07/01/1975 – 12/05/2015) on a one-dollar investment made in each of the four strategies discussed in Section 2.2. After transaction cost returns and Sharpe ratios are also provided.

Table 2: Performance Evaluation

Performance Measure	Investment Strategy			
	MBI	PSY-BI	DS	BH
Final Wealth ($)	68.37	17.98	0.00	24.01
Average Return (%)	11.00	7.39	-44.64	8.16
Sharpe Ratio (%)	8.78	5.95	-55.54	6.89

Notes: Average real returns (p.a.) are computed using weekly compounding. Trading costs are 0.5% of the real weekly return per buy-and-sell round trip.

As illustrated in the first column, the MBI strategy outperforms the other three investments by a large margin according to all three investment criteria. This strategy, which exploits the modified bubble indicator while keeping the funds in the market portfolio in the event of no bubbles, reports an after-transaction-cost return of 11.00 percent p.a. (in real terms). Thus, an investment of $1 made at the start of the trading period results in the final period wealth of $68.37. The second best result is generated by the buy-and-hold (BH) strategy that holds the market index for the duration of the investment period. This strategy returns 8.16 percent p.a. on average, and accumulates to $24.01 in 2015 from a $1 investment made in 1975. The performance of the BH strategy is followed by the return on the PSY-BI strategy that employs the original PSY bubble signal. A comparison of the figures presented in the first two columns of Table 2 illustrates the gains provided by augmenting the PSY indicator with the directional signal as in (3). Nevertheless, in the light of these results one may wonder if it is the directional signal itself that produces the superior performance of the MBI strategy. For this reason we also present the DS strategy which executes trades in accordance with the sign of the most recent return. As seen in Table 2 the DS strategy is ranked last, making an average loss of 44.64 percent p.a. and losing all of the investment by the end of the trading period. This result indicates that the effectiveness of the MDI strategy is not solely dependent on the directional signal, but that the condition in (3) regarding the sign of the return provides valuable trading information only in conjunction with the PSY bubble index. Lastly, the ranking of the four strategies remains the same if we consider risk-adjusted returns, as measured by Sharpe ratios in Table 2.

Next, in Table 3 we examine the characteristics of each strategy in more detail. We

Table 3: Trading Strategy Characteristics

Investment Strategy	# of Trading Periods (/2106)	Av. Weekly Returns (Std.) % Trading Period	Av. Weekly Returns (Std.) % Non-Trading Period	Sharpe Ratio % Trading Period	Sharpe Ratio % Non-Trading Period
MBI	363	1.51 (2.34)	-0.04 (0.02)	62.90	-2.39
PSY-BI	532	0.39 (3.05)	0.09 (2.25)	11.16	3.66
DS	1900	-0.76 (1.74)	-4.35 (2.40)	-44.43	-183.13
BH	2	N/A	N/A	N/A	N/A

Notes: Average real returns (p.a.) are computed using weekly compounding. Trading costs are 0.5% of the real weekly return per buy-and-sell round trip.

present the number of trading periods for each strategy, as well as average weekly returns, standard deviations and Sharpe ratios computed over trading and non-trading periods. Comparing the strategies presented across the first two rows we observe the difference in the number of trading periods between the MBI and PSY-BI strategies. By augmenting the PSY-BI with a directional signal we avoid opening a position when the returns are negative, but exit whenever the bubble signal turns to zero or the return becomes negative. Therefore, the number of trading periods is reduced from 532 to 363. This more selective bubble trading rule provides a higher average return/Sharpe ratio. As presented in Table 3, the average weekly return (Sharpe ratio) computed over bubble trading periods rises from 0.39% (11.16%) for the PSY strategy to 1.51% (62.90%) with the MBI strategy. DS is the most traded strategy with 1900 trading periods. This high trading frequency incurs large transaction costs and thereby results in negative after-transaction-cost returns and Sharpe ratios over both bubble trading and non-bubble trading periods.

B. Stochastic Dominance Test Results

The analysis presented above evaluates trading performance on the basis of accumulated wealth, average returns and Sharpe ratios. In this section we extend the study by re-examining the strategies according to the stochastic dominance criterion. We first present pairwise comparisons of the four strategies in Table 4.

With respect to the MDI strategy, the results of SD tests are consistent with the findings obtained on the basis of average returns and Sharpe ratios presented in Table 2. Looking across the top panel of Table 4 we observe the p-values for the tests of stochastic dominance of MDI over each of the other three strategies. The p-values computed via block bootstrap (KS1) and subsampling (KS2) methods are in accord across all cells of the top three rows. Namely, we are unable to reject the null hypotheses that the MDI strategy stochastically dominates the BH, DS and PSY-BI investments at any conventional level of significance. Moreover, the findings apply equally to the first-, second- and third-order stochastic dominance criteria.

Table 4: Pairwise Comparisons (p-values for SD Tests)

Null	SD1		SD2		SD3	
Hypothesis	KS1	KS2	KS1	KS2	KS1	KS2
MDI$\succeq$BH	0.396	0.205	0.342	0.266	0.456	0.378
MDI$\succeq$DS	1.000	1.000	0.978	1.000	0.946	1.000
MDI$\succeq$PSY-BI	0.369	0.283	0.806	0.674	0.768	0.674
PSY-BI$\succeq$BH	0.008	0.000	0.000	0.000	0.018	0.000
PSY-BI$\succeq$DS	0.998	1.000	0.696	0.821	0.914	1.000
PSY-BI$\succeq$MDI	0.000	0.000	0.000	0.000	0.000	0.000
DS$\succeq$BH	0.000	0.000	0.000	0.000	0.000	0.000
DS$\succeq$PSY-BI	0.000	0.000	0.000	0.000	0.000	0.000
DS$\succeq$MDI	0.000	0.000	0.000	0.000	0.000	0.000
BH$\succeq$DS	1.000	1.000	0.977	1.000	0.949	1.000
BH$\succeq$PSY-BI	0.000	0.000	0.874	0.955	0.788	0.718
BH$\succeq$MDI	0.001	0.000	0.001	0.000	0.022	0.011

Notes: A$\succeq$B indicates the hypothesis that investment strategy A stochastically dominates investment strategy B as given in (4). SD1, SD2, SD3 denote stochastic dominance of order 1, 2, and 3, respectively. KS1 test is implemented via block bootstrap; KS2 test is based on subsampling.

Looking over the second panel of Table 4 we note that the PSY-BI strategy stochastically dominates the DS strategy at all three orders. However, the null of stochastic dominance of PSY-BI over the BH and MDI is rejected at the 5 percent level by both tests, and for all orders of SD. Next, the p-values for the null that the DS strategy dominates each of the remaining three strategies are zero (to three decimal places) providing strong evidence against the null. This is in line with the results presented in Table 2, where the DS strategy also generates the worst performance. Finally, considering the first row of the last panel of Table 4 we observe that the buy-and-hold investment stochastically dominates the DS strategy according to SD1, SD2 and SD3. On the other hand, we reject (at the 5 percent level) the null that the BH dominates the MDI strategy for all orders of stochastic dominance. An interesting case is the relationship between the BH and PSY-BI where we reject the null that BH dominates PSY-BI under SD1, but are unable to reject it for SD2 or SD3. The results for SD2 and SD3 are consistent with the findings presented in panel two, while the finding for SD1 provides inconclusive inference. Overall, the results presented in Table 4 are largely consistent across all panels, and are strongly in favour of the MDI strategy for all three orders of SD.

While the above analysis compares pairs of trading strategies on the basis of SD1, SD2 and SD3, we now wish to consider all four strategies jointly. Specifically, we are interested in whether each of the four strategies stochastically dominates all the other investments as postulated in (5). Table 5 provides p-values for these tests.

Table 5: Testing the null hypothesis that the tested strategy stochastically dominates all other strategies (p-values)

Tested	SD1		SD2		SD3	
Strategy	KS1	KS2	KS1	KS2	KS1	KS2
MDI	0.376	0.205	0.354	0.266	0.459	0.378
PSY-BI	0.008	0.000	0.000	0.000	0.000	0.000
DS	0.000	0.000	0.000	0.000	0.000	0.000
BH	0.000	0.000	0.002	0.000	0.023	0.105

Notes: SD1, SD2, SD3 denote stochastic dominance of order 1, 2, and 3, respectively. KS1 test is implemented via block bootstrap; KS2 test is based on subsampling.

As illustrated in Table 5 there is strong evidence in favour of MDI as it stochastically dominates the remaining three investments at all three orders of SD considered here. Specifically, the p-values presented across the first row indicate that the null hypothesis of MDI dominating the other three strategies is accepted at any conventional level of significance, according to both KS1 and KS2 tests. This result confirms our findings obtained from pairwise tests presented in Table 4. Next, we observe small p-values across the second and third rows of Table 5, indicating that we reject the null hypotheses of PSY-BI and DS each dominating the remaining strategies at the 5 percent level. Finally, we consider p-values for the buy-and-hold strategy presented in the last row. The null of BH dominating the remaining strategies at SD1 and SD2 is rejected strongly by both tests. However, the conclusion is not reached uniformly for SD3. The KS1 test p-value computed on the basis of a block bootstrap procedure rejects the null at the 5 percent level. On the other hand, the KS2 test which relies on a subsampling method produces a p-value of 0.105, leading us to conclude (at the 5 percent level) that BH dominates the other three strategies according to SD3. Nevertheless, on the balance of evidence presented in Tables 4 and 5 we conclude that the MDI strategy stochastically dominates the other three prospects, both individually and jointly.

V. Robustness and Sensitivity Analysis

In this section we consider a number of alternative scenarios to evaluate the robustness of the results reported above. First, we evaluate sub-sample performance of our strategies. This is followed by an investigation of profitability at alternative confidence levels, as well as further modifications of the PSY bubble indicator.

A. Investing over Sub-sample Periods

In the previous section we compared the performances of the four alternative strategies over the entire 07/01/1975 – 12/05/2015 sample period. Although this may provide an adequate evaluation window for institutional investors, smaller private investors are likely to be interested in learning about the investment returns over periods shorter than the 40 year horizon considered above. In this subsection we assess the relative performances over a range of shorter sub-samples by varying the investment window length between 1 and 5 years, and rolling the investment period by 1 week at a time. This section focuses on the modified bubble trading strategy (MBI) and the buy-and-hold investment in the market index (BH), as these two strategies respectively ranked first and second in the analysis presented above.

Our setup here is such that investors may enter the market at any point in time, and exit k periods after with $k = 52, 52 \times 2, \cdots, 52 \times 5$ weeks. Returns, which are computed at the end of each rolling window, will inform us about the degree of consistency in the performance of the MBI strategy over different holding periods, as well as its sensitivity to any pivotal sub-periods.

Let $\Delta\tilde{R}^{P}_{t,k}$ denote the difference between the average annualized real returns of the MBI and BH portfolios for the period running from $t-k$ to t, namely

$$\Delta\tilde{R}^{P}_{t,k} = \tilde{R}^{MBI}_{t,k} - \tilde{R}^{BH}_{t,k},$$

where $\tilde{R}^{MBI}_{t,k}$ and $\tilde{R}^{BH}_{t,k}$ denote the average annual real returns on the MBI and BH portfolios, for the period from $t-k$ to t, with k being the rolling window size. Table 6 reports the fraction of time over which the MBI strategy outperforms ($\Delta\tilde{R}^{P}_{t,k} > 0$), underperforms ($\Delta\tilde{R}^{P}_{t,k} < 0$) and performs equally as well as the BH investment ($\Delta\tilde{R}^{P}_{t,k} = 0$). We report the average annualised return difference of these two strategies, along with their t-statistics in brackets[4].

Across all rolling windows, the percentage of times that MBI outperforms BH is substantially higher than the frequency at which it underperforms. For instance, setting

[4] The average difference in returns and robust standard error (and hence t-statistics) are obtained by running an OLS regression of $\Delta\tilde{R}^{P}_{t,j}$ on a constant with Newey-West standard errors.

Table 6: Rolling Window Performance of the MBI strategy relative to BH

Window	Outperformance		Underperformance		Equal Perfomance
Size	Frequency	Av. Return Diff.	Frequency	Av. Return Diff.	Frequency
1 year	34.94%	12.46% (5.14)	22.22%	-3.21% (-8.85)	36.84%
2 years	52.42%	7.27% (5.81)	29.16%	-1.86% (-9.11)	18.42%
3 years	64.99%	5.28% (6.59)	26.09%	-0.02% (-9.51)	8.92%
4 years	73.88%	4.35% (7.26)	21.91%	-1.54% (-9.30)	4.21%
5 years	77.42%	4.02% (7.90)	21.06%	-1.34% (-10.38)	1.52%

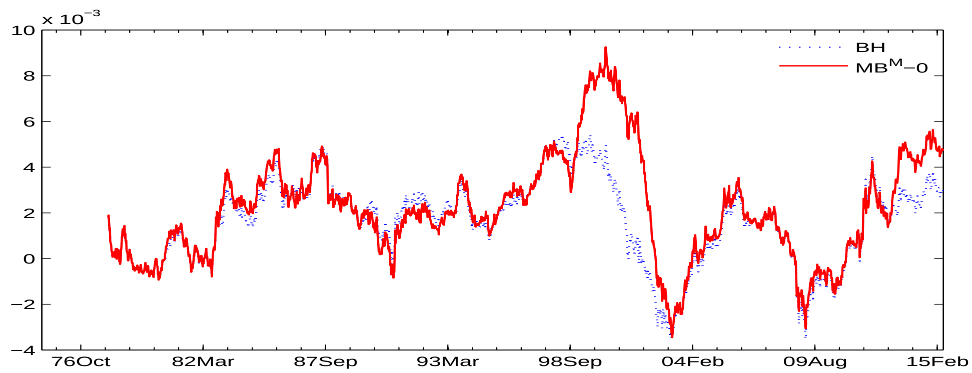
Note: HAC robust t-statistics are reported in brackets.

$k = 52$, which corresponds to the one year investment horizon, we observe that MBI outperforms BH in 34.94 percent of all one-year investment horizons contained in the entire sample period. In comparison MBI underperforms BH 22.22 percent of the time according to the average return criterion. The magnitude of the outperformance is 12.46 percent p.a., while the underperformance amounts to about -3.21 percent p.a. In order to evaluate the statistical significance of these figures we also report heteroskedasticity and autocorrelation (HAC) robust t-ratios in the brackets next to the return differences. It is clear that both the outperformance and underperformance figures are statistically significant at any conventional level of statistical significance.

Lastly, the frequency of equal performance periods is computed as the fraction of the total number of k horizon investment periods not accounted for by the instances of overperformance and underperformance. In the case when $k = 52$ we see that about 36.84 percent of the time the MBI and BH strategies provide the same average annualized return. This figure is largely due to the fraction of one-year investment periods over which the MBI and BH strategies are equal, i.e. when all MBI indicators are zero and the MBI strategy allocates funds into the market index. As the length of the rolling window increases from 1 to 5 years, the frequency of equal performance periods decreases while the number of times the MBI outperforms increases. The improvement in outperformance occurs because MBI is afforded more time to detect bubbles, and hence for the funds to be invested into growing bubble sectors. The magnitude of average annualised return differences reduces as the size of the rolling window is increased. This is due to the fact that there are more non-bubble trading periods in the MBI strategy during longer investment horizons, and hence, the difference in returns between the MBI and BH strategies narrows.

In order to shed more light on the numbers reported in Table 6, we provide a time series plot of the rolling window investment returns (in real terms p.a.) for the MBI and BH strategies for $k = 3 \times 52$ weeks. Graphs for the remaining rolling window lengths exhibit similar patterns and are available from the authors upon request. A noteworthy

Figure 3: The rolling window average annual real returns over a three-year window



observation is the asymmetric pattern in the outperformance and underperformance of the MBI versus the BH strategy. Namely, when MBI outperforms BH it does so with a noticeably greater magnitude than when it underperforms. This is most clearly illustrated during the period of the dot-com bubble, starting around September 1998, as well as over the three-year period at the end of the sample. Over these subperiods, MBI provides impressive outperformance results. On the other hand, the underperformance of MBI over the period 1991 – 1993 is small by comparison. Following the burst of the tech bubble in the early 2000 we observe large losses in both strategies. By construction, the MBI strategy retreats into the market index when the bubble indicators are off. Therefore, during the period of the dot-com bubble crash both portfolios contained predominantly the market index, which accounts for similar downward trajectories. It is also of interest to note that since 2013, the MBI strategy is mainly invested in the two sectors (namely utilities and diversified REITs) – see Figure 1 and 2.

B. Trading at Alternative Confidence Levels

Previous analysis employed the PSY test to detect bubble formation dates with 99 percent confidence. In this section we investigate the impact of alternative confidence levels on the profitability of the MBI strategy, and report our findings in Table 7. For the sake of comparison the table also provides figures with and without transaction costs, as well as the tests of stochastic dominance.

First, we note the inverse relationship between the confidence level and the number of trading periods. As the confidence decreases from 99 percent to 95 and 90 percent, the number of trading periods increases from 363 to 473 and 650, respectively. However, the impact of these changes is very different for the outcomes computed with and without transaction costs.

Considering the case when the performance is measured after subtracting transaction costs, we observe that the final wealth, annual return and Sharpe ratios all monotonically decrease as the number of active trading periods increases[5]. Nevertheless even with 90 percent confidence – the lowest confidence level considered here – the MBI strategy still outperforms the BH investment, whose return is reported in Table 2. The impact of transaction costs is demonstrated by comparing the first three and the last three columns of Table 7. In contrast to the previous case, all three performance criteria increase with lower confidence levels when transaction costs are ignored. The most striking difference is observed in the final wealth figures whose magnitudes clearly diverge across the two sets of results. For instance, the ratio between final wealth numbers for the scenarios without and with transaction costs increases in the sequence $\{10, 30, 160\}$ as the confidence level decreases over the $\{99\%, 95\%, 90\%\}$ range. This occurs due to more timely detections of major as well as smaller bubbles, which are identified at lower confidence levels. Nevertheless, as the after-transaction-cost figures indicate, these benefits are quickly eroded by large transaction costs associated with the trading of such smaller bubbles.

Table 7: Performance evaluation and SD tests for alternative confidence levels (p-values)

Performance		With Transaction Costs			Without Transaction Costs		
Measure		MBI-90%	MBI-95%	MBI-99%	MBI-90%	MBI-95%	MBI-99%
Final Wealth ($)		28.56	39.39	68.37	4586	1234	706.31
Annual Return (%)		8.69	9.49	11.00	23.14	19.21	17.58
Sharpe Ratio (%)		7.04	7.68	8.78	17.12	14.50	13.38
# of Trading Periods		650	473	363	650	473	363
p-values for SD tests of H_0: tested strategy stochastically dominates the other two strategies							
	SD1	0.036	0.069	0.703	0.872	0.000	0.000
KS1	SD2	0.007	0.000	0.863	0.814	0.000	0.000
	SD3	0.002	0.001	0.934	0.777	0.000	0.000
	SD1	0.000	0.000	0.645	0.497	0.000	0.000
KS2	SD2	0.035	0.000	1.000	0.289	0.000	0.000
	SD3	0.000	0.000	1.000	0.289	0.000	0.000

Notes: After-transaction-cost returns are compounded weekly and reported in real terms per annum. SD1, SD2, SD3 denote stochastic dominance of order 1, 2, and 3, respectively. KS1 test is implemented via block bootstrap; KS2 test is based on subsampling.

Turning to the figures presented in the bottom panel of Table 7 we observe p-values computed for the null hypothesis that each tested strategy stochastically dominates the

[5]Recall that MBI invests funds in the market index when no bubbles are detected in any of the sector indices.

remaining two strategies as specified in (5). Thus, large p-values provide evidence in support of stochastic dominance of the tested strategy. The tests are computed for two sets of returns, one which takes transaction costs into account and one which does not. Each set consists of the MBI strategies generated on the basis of the 99, 95 and 90 percent confidence levels. We first note that the KS1 (block bootstrap) and KS2 (subsampling) tests produce largely consistent results, at the 5 percent significance level. It is also the case that the p-values are in line across the tests for the first-, second-, and third-order stochastic dominance.

Considering the first three columns it is clear that the MBI returns computed on the basis of the 99 percent confidence stochastically dominate the returns calculated with 95 and 90 percent confidence bubble tests, at the 5 percent significance level. The only exception is the KS1 test of SD1 computed for MBI-95%, which results in the p-value of 0.069 and thus accepts the null that this strategy dominates the 99 and 90 percent strategies, at the 5 percent significance level. However, given that the null hypotheses of SD2 and SD3 are rejected for this instance, while in theory SD1 implies SD2 and SD3, we attribute this finding to chance rather than regularity. This interpretation is further supported by the evidence provided by the KS2 test for SD1, which clearly rejects the null hypothesis.

In the case when transaction costs are not accounted for, we clearly reject the null hypotheses of stochastic dominances of MBI-99% and MBI-95% for SD1, SD2 and SD3. On the other hand, the evidence strongly suggests that MBI-90% dominates the other two investments at all three orders of SD. These findings corroborate the conclusions reached on the basis of Sharpe ratios, and illustrate the effect of transaction costs on the profitability of bubble-timing strategies. It appears that all investors with increasing utility functions would prefer investments made with 99 percent confidence when faced with transaction costs, while their preference would shift towards the 90 percent confidence level in the case of no transaction fees.

C. Further Modifications of the MBI Strategy

As a sensitivity check to alternative modifications of the PSY-BI strategy we investigate two further specifications. The first modification of the MBI strategy augments the PSY indicator with longer information sets used by the directional signal when opening and closing a position. In comparison to (3), the new bubble signal is set to one when the PSY indicator is one, *and* all of the past j periods experience positive returns. On the other hand, the bubble indicator switches from one to zero if either the PSY indicator changes to zero, *or* the past j^* period returns are all negative. This indicator is more concisely described as follows:

$$I_t^+ = \begin{cases} 1(BSADF_t > scv \ \& \ R_t > 0, ..., R_{t-j} > 0 | I_{t-1}^+ = 0) \\ 1 - 1(BSADF_t < scv \text{ or } R_t < 0, ..., R_{t-j^*} < 0 | I_{t-1}^+ = 1), \end{cases}$$

where R_t represents the nominal weekly return on the period t, and we consider $j = 0, 1, 2$, and $j^* = 0, 1, 2$. This strategy is denoted MBI(j, j^*) in Table 8 below.

First of all, we note that the MBI(1,0) strategy outperforms the remaining MBI(j, j^*) modifications according to the three performance criteria considered here. It is closely followed by the MBI(0,1) scheme, which generates a slightly lower average return and Sharpe ratio. Nevertheless, the original MBI strategy or MBI(0,0), reported in the last row, still ranks in first place and performs better than its modifications. This is best illustrated by its final period wealth of \$68.37, which is almost double the size of the largest final wealth figure provided by modified strategies. The ranking of the remaining MBI(j, j^*) strategies is not uniformly ordered by the three criteria. Some strategies which rate high on the basis of final wealth and average return, fall in the ranking order once risk is taken into account by Sharpe ratios. For instance, MBI(0,1) ranks higher than MBI(2,0) in terms of average return, although these two strategies rank the same according to the Sharpe ratio. Similarly, the rankings of MBI(0,2) and MBI(2,2) are reversed when considering Sharpe ratios relative to average returns.

Table 8: Performance evaluation and SD tests for MBI(j, j^*) modification with a null hypothesis that tested strategy stochastically dominates all other strategies (p-values).

Investment Strategy	Final Wealth (\$)	Average Return (%)	Sharpe Ratio (%)	KS1 SD1	KS1 SD2	KS1 SD3	KS2 SD1	KS2 SD2	KS2 SD3
MBI(0,1)	32.96	9.01	7.23	0.012	0.000	0.000	0.000	0.000	0.000
MBI(0,2)	27.90	8.57	6.87	0.010	0.000	0.000	0.000	0.000	0.000
MBI(1,0)	37.41	9.35	7.66	0.179	0.004	0.018	0.038	0.000	0.077
MBI(1,1)	22.15	7.95	6.51	0.001	0.000	0.000	0.000	0.000	0.000
MBI(1,2)	19.01	7.54	6.15	0.002	0.000	0.000	0.000	0.000	0.000
MBI(2,0)	29.38	8.70	7.23	0.017	0.000	0.015	0.000	0.000	0.000
MBI(2,1)	19.72	7.64	6.37	0.028	0.000	0.000	0.000	0.000	0.000
MBI(2,2)	27.22	8.50	6.99	0.000	0.000	0.000	0.000	0.000	0.000
MBI	68.37	11.00	8.78	0.197	0.648	0.687	0.059	0.679	0.663

Notes: After-transaction-cost returns are compounded weekly, and are reported in real terms per annum. SD1, SD2, SD3 denote stochastic dominance of order 1, 2, and 3, respectively. KS1 test is implemented via block bootstrap; KS2 test is based on subsampling.

The last six columns of Table 8 present p-values computed for the null hypothesis that each of the tested strategies stochastically dominates the remaining eight investments, as specified in (5). KS1 (block bootstrap) and KS2 (subsampling) tests are computed for the set of nine strategies listed in the table, and like the results obtained for SD tests reported previously, these procedures lead to largely consistent results at the 5 percent significance level. In addition, the p-values are in accord across the tests for the first-, second-, and third-order stochastic dominance.

There are only two strategies for which the null hypothesis of stochastic dominance over the remaining investments is accepted at the five percent level. These are the original MBI investment and its MBI(1,0) modification. Considering the last row of Table 8 we observe relatively large p-values across all SD tests. Namely, the null of SD1, SD2 and SD3 is accepted for this strategy by both KS1 and KS2 tests. However, the evidence for SD of MBI(1,0) is on softer grounds as KS1 and KS2 tests provide somewhat conflicting results. While the KS1 procedure suggests that MBI(1,0) stochastically dominates the remaining investments at the first order, KS2 provides evidence of only SD3 for this strategy. Given that SD1 implies SD2, the result provided by the KS1 test in this instance, may be questioned since this test strongly rejects SD2 and SD3 while at the same time accepts SD1.

Overall, it seems that the informational content of directional changes observed in the most recent past, as reflected in the basic MBI strategy, produces the best correction to the PSY bubble indicator. This, in turn, results in the most profitable strategy. On the other hand, delayed directional signals present in MBI(j, j^*), for some $j > 0$ or $j^* > 0$ diminish investment performance.

The second modification to the basic MBI scheme invests in sector indices during non-bubble periods, instead of holding the market index like the original MBI strategy. Funds are not, however, allocated across all non-bubble sectors as this would incur significant transaction costs and ultimately amount to investing in the entire market. Instead we discriminate between sectors on the basis of duration d since the last bubble episode. Thus, in the absence of bubbles in any of the sector indices the new strategy invests, with equal weights, in those sectors that have not experienced a bubble episode for at least d weeks. Besides reducing the transaction costs, this strategy attempts to enter sectors that are relatively insusceptible to the dynamics of the most recently collapsed bubbles. We form and assess the performance of the portfolios for which $d = 26, ..., 156$, i.e. the portfolios which have not experienced bubble episodes for periods ranging from six months to three years. In order to further limit the amount of transaction costs, we do not rebalance these non-bubble sector portfolios until the MBI indicator detects a bubble forming in one or more sectors. This strategy is labelled as MBI-NB(d) in Table 9 below.

It is clear that the performance of the MBI-NB(d) strategies improves with the duration d that the non-bubble sectors have spent outside of bubble episodes. As d varies from six months (26 weeks) to three years (156 weeks), the final wealth on a \$1 investment made at the beginning of the sample period increases from \$12.73 to \$30.46. A similar

Table 9: Performance evaluation and SD tests for MBI-NB(d) modification

Investment Strategy	Final Wealth ($)	Average Return (%)	Sharpe Ratio (%)	KS1 SD1	KS1 SD2	KS1 SD3	KS2 SD1	KS2 SD2	KS2 SD3
MBI-NB(26)	12.73	6.48	5.33	0.000	0.000	0.000	0.000	0.000	0.000
MBI-NB(52)	15.28	6.96	5.67	0.000	0.000	0.000	0.000	0.000	0.000
MBI-NB(78)	21.79	7.91	6.36	0.000	0.000	0.000	0.000	0.000	0.000
MBI-NB(104)	26.16	8.39	6.76	0.000	0.000	0.000	0.000	0.000	0.000
MBI-NB(130)	28.21	8.60	6.90	0.000	0.000	0.000	0.000	0.000	0.000
MBI-NB(156)	30.46	8.80	7.05	0.000	0.000	0.000	0.000	0.000	0.000
MBI	68.37	11.00	8.78	1.000	0.919	0.841	1.000	0.862	0.849

Notes: After-transaction-cost returns are compounded weekly, and are reported in real terms per annum. p-values are provided for SD tests of the null hypothesis that a tested strategy stochastically dominates all other strategies. SD1, SD2, SD3 denote stochastic dominance of order 1, 2, and 3, respectively. KS1 test is implemented via block bootstrap; KS2 test is based on subsampling.

pattern is observed in the Sharpe ratios, which increase with d. Nevertheless, as is the case with the previously examined MBI(j, j^*) strategy, the current MBI-NB(d) modification still underperforms the basic MBI strategy according to all three criteria. Similarly, the large p-values provided by the SD tests for the case of the original MDI strategy fail to reject the null of MBI dominating the remaining six investments at the first-, second-, and third-order SD. On the other hand, the same null hypothesis is rejected for all MBI-NB(d) modifications at any conventional level of significance. Therefore, the SD tests presented across the last six columns of Table 9 are all in accord, and provide strong evidence in favour of the original MBI scheme. It appears that, in the absence of bubbles, keeping funds in the market index rather than the non-bubble sector indices generates a strategy that is preferred by any investor with increasing utility function.

VI. Conclusion

This paper applies the bubble detection approach of Phillips et al. (2015a,b) to the eleven US industrial sectors for the sample period of 1973 – 2015 at the weekly frequency. A wild bootstrapping procedure is employed to tackle the potential heteroskedastic feature of the weekly stock prices. The testing results provide important insights with regards to (i) the existence and duration of speculative bubbles and (ii) the possibility to construct portfolios invested in sector indices that ride bubbles.

We report evidence of bubble dynamics across all sectors of the economy. The two major episodes of speculative behaviour occurred in 1987 and 1998 – 2000. The 1987 bubble episode started from the financials and utilities sectors and spread to all other sectors except industrials. Similarly, the dot-com bubble episode originated in the technology and financials sectors and affected the health care, utilities, oil and gas, and basic materials sectors.

To evaluate the financial usefulness of the detected bubble signals, we propose two trading strategies based on, respectively, the bubble signals provided by the original PSY test, and a modified bubble indicator. The modified bubble signal combines the PSY bubble signal with the direction of the most recent movement of the index. Performance of the trading strategies is evaluated by final period wealth, average after-transaction-cost returns and Sharpe ratios, and benchmarked against a buy-and-hold trading strategy and a directional signal based on the most recent index movement. The relative performance of the trading strategies are further ranked according to the stochastic dominance criteria, which are implemented as the tests proposed in Linton et al. (2005).

All performance measures and tests provide evidence of superior performance of the MBI strategy, which is based on the modified bubble signal. Specifically, with $1 initial investment at the beginning of the sample period, one would expect a final wealth (after transaction cost) of $68.37 in 40 years, implying an (average) 11 percent annual return. This is compared to a final wealth of $24 and an average annual return of 8.16% generated by the buy-and-hold strategy. Subsample analysis and specification checks confirm the robustness of our findings. Though we uncover alternate bubble-trading strategies that outperform the buy-and-hold benchmark, the MBI strategy dominates in terms of real returns and risk-adjusted returns, as well as according to the stochastic dominance criterion.

Of the few studies that do implement real-time trading strategies, such as Brooks and Katsaris (2005), any substantial profits gained are largely eroded once trading costs are accounted for. Our findings suggests that there exist trading strategies that can exploit the early detection of equity bubbles using recent technology in the literature, and produce substantial profits for investors even in the presence of transaction costs. Since "riding bubbles" is not a form of arbitrage, the existence of pricing bubbles is expected to persist. As long as there are traders willing to ride bubbles, even the correct identification of bubbles will not result in an immediate market correction.

References

Abhyankar, A., K.-Y. Ho, and H. Zhao (2008). Value versus growth: stochastic dominance criteria. *Quantitative Finance 8*(7), 693–704.

Abreu, D. and M. K. Brunnermeier (2002). Synchronization risk and delayed arbitrage. *Journal of Financial Economics 66*(2), 341–360.

Abreu, D. and M. K. Brunnermeier (2003). Bubbles and crashes. *Econometrica 71*(1), 173–204.

Bar-Yosef, S. and L. Huffman (1986). The information content of dividends: a signalling approach. *Journal of Financial and Quantitative Analysis 21*(01), 47–58.

Barrett, G. F. and S. G. Donald (2003). Consistent tests for stochastic dominance. *Econometrica 71*(1), 71–104.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Brooks, C. and A. Katsaris (2005). A three-regime model of speculative behaviour: Modelling the evolution of the s&p 500 composite index*. *The Economic Journal 115*(505), 767–797.

Brunnermeier, M. and S. Nagel (2004). Hedge funds and the technology bubble. *The Journal of Finance 59*(5), 2013–2040.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 1171–1179.

Chernozhukov, V. and I. Fernández-Val (2005). Subsampling inference on quantile regression processes. *Sankhyā: The Indian Journal of Statistics*, 253–276.

Davidson, R. and J.-Y. Duclos (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica 68*(6), 1435–1464.

Denis, D. J. and I. Osobov (2008). Why do firms pay dividends? international evidence on the determinants of dividend policy. *Journal of Financial economics 89*(1), 62–82.

Diba, B. T. and H. I. Grossman (1988). Explosive rational bubbles in stock prices? *The American Economic Review 78*(3), 520–530.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society 50*(4), 987–1007.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*(3), 1269–1300.

Etienne, X. L., S. H. Irwin, and P. Garcia (2014). Bubbles in food commodity markets: Four decades of evidence. *Journal of International Money and Finance 42*, 129–155.

Fong, W. M., W. K. Wong, and H. H. Lean (2005). International momentum strategies: a stochastic dominance approach. *Journal of Financial Markets 8*(1), 89–109.

Guenster, N. and E. Kole (2013). Bubbles and long-term investors.

Hall, S. G., Z. Psaradakis, M. Sola, et al. (1999). Detecting periodically collapsing bubbles: a markov-switching unit root test. *Journal of Applied Econometrics 14*(2), 143–154.

Harvey, D. I., S. J. Leybourne, R. Sollis, and A. R. Taylor (2015). Tests for explosive financial bubbles in the presence of non-stationary volatility. *Journal of Empirical Finance*.

Homm, U. and J. Breitung (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Klecan, L., R. McFadden, and D. McFadden (1991). A robust test for stochastic dominance. *Unpublished paper, MIT*.

Levy, H. (1985). Upper and lower bounds of put and call option value: Stochastic dominance approach. *The Journal of Finance 40*(4), 1197–1217.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies 72*(3), 735–765.

McFadden, D. (1989). Testing for stochastic dominance. In *Studies in the Economics of Uncertainty*, pp. 113–134. Springer.

Phillips, P. C. and S. Shi (2014). Financial bubble implosion. *Available at SSRN 2487601*.

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

Phillips, P. C., S. Shi, and J. Yu (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International Economic Review*.

Phillips, P. C., S. Shi, and J. Yu (2015b). Testing for multiple bubbles: Limit theory of real time detectors. *International Economic Review*.

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

Politis, D., J. P. Romano, and M. Wolf (1999). Weak convergence of dependent empirical measures with application to subsampling in function spaces. *Journal of statistical planning and inference 79*(2), 179–190.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical association 89*(428), 1303–1313.

Post, T. (2003). Empirical tests for stochastic dominance efficiency. *The Journal of Finance 58*(5), 1905–1932.

Seyhun, H. N. (1993). Can omitted risk factors explain the january effect? a stochastic dominance approach. *Journal of Financial and Quantitative Analysis 28*(02), 195–212.

van Norden, S. (1996). Regime switching as a test for exchange rate bubbles. *Journal of Applied Econometrics 11*(3), 219–251.

Wurgler, J. and E. Zhuravskaya (2002). Does arbitrage flatten demand curves for stocks? *The Journal of Business 75*(4), 583–608.

VII. Appendix: Stochastic Dominance Test Statistics and p-values

Given that we wish to investigate stochastic dominance relationships between return series of portfolios whose prices exhibit explosive dynamics at times, we expect the return series to be correlated over time. Therefore, we apply a block bootstrap procedure which takes into account time-series dependence in the data. Our procedure is implemented as a non-overlapping block bootstrap, see e.g. Carlstein (1986), with a suitably centered test statistic. We describe this procedure in more detail as follows.

Step 1. Compute the test statistic $T_{N,m}^{(s)}$ for $m = 0, 1, 2$ using the full sample $\mathcal{Z}_N = \{Z_i = (X_{1i}, X_{2i}, \ldots, X_{Ki})' : i = 1, 2, \ldots, N\}$.

Step 2. Create $B = \frac{N}{L}$ blocks $\mathcal{Z}_{N,L,b} = \{Z_{bL+1}, Z_{bL+2}, \ldots, Z_{bL+L}\}$ of length L for $b = 0, 1, \ldots, B-1$ and $0 \leq \gamma \leq 1$ such that $b \to \infty$ and $\frac{b}{N} \to 0$ as $N \to \infty$.

Step 3. Generate the bootstrap sample $\mathcal{Z}_N^* = \{Z_i^* : i = 1, 2, \ldots, N\}$ obtained by sampling B blocks randomly with replacement from the B non-overlapping blocks described above.

Step 4. Compute the recentered test statistic $T_{N,m}^{(s)*} = \sqrt{N}\hat{d}_m^{(s)*}$ for $m = 1, 2$ using the bootstrap sample $\mathcal{Z}_N^*$ where

1. $\hat{d}_1^{(s)*} = \sup_{z\in\Lambda}[(\hat{\mathcal{I}}_k^{(s)*}(z; \hat{F}_k^*) - \hat{\mathcal{I}}_k^{(s)}(z; \hat{F}_k)) - (\hat{\mathcal{I}}_l^{(s)*}(z; \hat{F}_l^*) - \hat{\mathcal{I}}_l^{(s)}(z; \hat{F}_l))]$
2. $\hat{d}_2^{(s)*} = \max_{l:l\neq k}\sup_{z\in\Lambda}[(\hat{\mathcal{I}}_k^{(s)*}(z; \hat{F}_k^*) - \hat{\mathcal{I}}_k^{(s)}(z; \hat{F}_k)) - (\hat{\mathcal{I}}_l^{(s)*}(z; \hat{F}_l^*) - \hat{\mathcal{I}}_l^{(s)}(z; \hat{F}_l))]$.

where $\hat{\mathcal{I}}_k^{(s)*}(z; .)$ and $\hat{\mathcal{I}}_k^{(s)}(z; .)$ are the bootstrap and the full sample empirical operators.

Step 5. Let R be the number of bootstrap repetitions. Obtain an approximate asymptotic p-value as $p_{L,i}^{(s)*} = \frac{1}{R}\sum_{i=1}^{R} 1(T_{N,m}^{(s)} > T_{N,m}^{(s)*})$ for $m = 0, 1, 2$.

In contrast to the bootstrap method, the subsampling approach takes chunks of data as subsamples from which to compute the p-values. It is introduced in Politis and Romano (1994) and reviewed is provided in Politis et al. (1999). We implement a centered version of the method which is proposed in Chernozhukov and Fernández-Val (2005), and which makes the subsampling approach robust to different subsample sizes. The subsampling is described as follows.

Step 1. Compute the test statistic $T_{N,m}^{(s)}$ for $m = 0, 1, 2$ using the full sample $\mathcal{Z}_N = \{Z_i = (X_{1i}, X_{2i}, \ldots, X_{Ki})' : i = 1, 2, \ldots, N\}$.

Step 2. Create subsamples $\mathcal{Z}_{N,S,i} = \{Z_i, Z_{i+1}, \ldots, Z_{i+S-1}\}$ for $i = 1, 2, \ldots, N-S+1$ of size S such that $S \to \infty$ and $\frac{S}{N} \to 0$ as $N \to \infty$.

Step 3. Compute centered test statistics $T_{N,S,i,m}^{(s)}$ for $m = 0, 1, 2$ and $i = 1, 2, \ldots, N-S+1$ using the subsamples $\mathcal{Z}_{N,S,i}$ and analogues of $\hat{d}_1^{(s)*}$ and $\hat{d}_2^{(s)*}$ described in the bootstrap procedure.

Step 4. Obtain an approximate asymptotic p-value as $p_{b,i}^{(s)} = \frac{1}{N-S+1}\sum_{i=1}^{N-S+1} 1(T_{N,m}^{(s)} > T_{N,S,i,m}^{(s)})$ for $m = 1, 2$, and $s = 1, 2, 3$.