PRELIMINARY: PLEASE DO NOT QUOTE WITHOUT PERMISSION

The recent behavior of commodity prices: fundamentals, speculative bubbles and relation to the global economic environment

Isabel Figuerola-Ferretti [a], Christopher L. Gilbert [b], and J. Roderick McCrorie [c]

[a]*Departamento de Economía de la Empressa, Universidad Carlos III de Madrid, Calle Madrid 126, 28093 Getafe (Madrid), Spain*
[b]*Dipartimento di Economia, Università degli Studi di Trento, via Inama 5, 38122 Trento, Italy*
[c]*School of Economics and Finance, University of St Andrews, St Salvator's College, St Andrews, KY16 9AL, Scotland, UK*

17 May 2012

ABSTRACT

This paper tests for the existence of speculative bubbles in five key commodities – crude oil, gold, silver, aluminum and copper – over the last decade or so, focusing on possible bubble characteristics around the time of the financial crisis. We utilize the new methodology of Phillips, Wu and Yu (2011, International Economic Review 52, 210-226), and Phillips and Yu (2011, Quantitative Economics 2, 455-491) that allows for date-stamping the origination and collapse of bubbles, but in a form where critical values are made robust to allow for possibly different data spans and sampling frequencies. We find evidence of bubble behavior in the copper, gold and silver markets in the first half of 2006. Results are less conclusive for the aluminum market. We fail to find convincing evidence for a 2007-08 crude oil bubble.

1. Introduction

Recent, sharp increases in commodity prices, including increases in the prices of crude oil, industrial metals and precious metals, have led to suggestions that price rises may have been driven by non-fundamentally-based speculation. Phillips and Yu (PY, 2011) have pointed to some commodity prices having exhibited bubble-type behavior – see also Gilbert (2010b), and Shi and Avora (2012). The Commodity Futures Trading Commission (CFTC), which is the U.S. futures market regulator, defines a speculative bubble as[1]

> ". . . a rapid run-up in prices caused by excessive buying that is unrelated to any of the basic, underlying factors affecting the supply or demand for a commodity or other asset. Speculative bubbles are usually associated with a "bandwagon" effect in which speculators rush to buy the commodity (in the case of futures, "to take positions") before the price trend

[1] http://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/glossary_s

ends, and an even greater rush to sell the commodity (unwind positions) when prices reverse"

While there is awareness among market participants that institutional investors including hedge funds and commodity index traders play an important role in determining commodity futures prices, it is not at all clear whether their activities lack a fundamental basis in supply and demand, or whether they are instead simply part of a mechanism whereby information on market fundamentals becomes impounded in competitively determined prices. There is currently little consensus on this matter, either among practitioners or academics. In the former group, Masters (2008, 2010) and Soros (2008) have argued that speculative positions taken by an influx of new traders into oil and other commodity futures markets, viewing commodities purely as an asset class, were essential in explaining the run-up in prices seen in the last decade. This process has been described as the *increased financialization* of commodities markets (see, e.g., Mayer, 2010; Irwin and Sanders, 2011; Tang and Xiong, 2011; Fattouh et al. 2012). That such financialization has changed the composition of such markets is undisputed; for example, Tang and Xiong quote a CFTC staff report[2] that estimated the total value of various commodity index-related instruments purchased by institutional investors to have increased from around $15 billion in 2003 to over $200 billion in mid-2008. Drawing on the same report, Gilbert (2010b) notes that this translates into index fund share having grown, as measured in June 2008, to over a one-quarter of total market open interest in crude oil futures, over two-fifths in copper, and over a fifth in gold and silver. Apart from a sharp decline during the recent slowdown, the rise in index investment has continued unabated, with current values easily surpassing previous highs. Much more controversial is whether increased financialization has been the *major driver* of commodity futures prices over the last decade, whether it has created "excessive speculation" manifesting itself as bubbles, or whether the large positions taken by such investors have led to self-fulfilling movements in prices that represent "market manipulation". Certainly, such views have been invoked to underpin a number of recent bills in the U.S. Congress that seek to limit index fund speculation in commodity futures; for example the *2010 Dodd-Frank Wall Street Reform and Consumer Protection Act* granted the CFTC board the power to set aggregate speculative position limits on futures and swap positions in all (non-exempt) U.S. physical commodity markets. The academic support for a key role having been played by increased financialization in commodity futures markets is mixed. Some support has been

[2] CFTC, 2008. Staff report on commodity swap dealers & index traders with commission recommendations. (http://cftc.gov/stellent/groups/public/@newsroom/documents/file/cftcstaffreportsonswapdealers09.pdf)

provided by Gilbert (2010a,b), Silvenoinen and Thorp (2010), Mayer (2011), and Tang and Xiong (2011). However, Sanders and Irwin (2010, 2011), Stoll and Whaley (2010, 2011), and Fattouh et al. (2012) find the evidence much less compelling.

In parallel with approaches based on financialization, there has been a developing literature that emphasizes fundamental supply and demand factors in a macroeconomic context, setting commodity price movements explicitly within the global economic environment. Hamilton (2009) provides an explanation of the 2007-08 oil price spike in terms of the growth in world real gross domestic product (GDP), an increase in consumption notably from China and other developing nations, and a stagnation of global oil production between 2005 and 2007; the fall in mid-2008 being accounted for by the global economic downturn. Harris et al. (2009) take a similar view but also relate the oil price to U.S. monetary policy, and envisage there being commonality in its impact on all commodities. Using the structural approach of Kilian (2009) that offers a basis for disentangling the effects of oil price fluctuations, Kilian and Murphy (2011) find that, while speculation could have played a role in earlier oil price shocks, the run-up in the price of oil between 2003 and mid-2008 was not explained by unexpectedly diminishing oil supplies or speculative demand but by the fluctuations in the flow demand for oil driven by the global business cycle. Caballero et al. (2008a,b) proposed a general equilibrium model to enucleate the recent financial crisis, linking global financial asset scarcity, global imbalances, the rise in U.S. real estate prices, the subsequent subprime crisis and spike in commodity prices. Essential to the model is a pattern of bubble creation followed by collapse that migrates from sector to sector. It predicts that, after the beginning of the U.S. subprime crisis and the ensuing global liquidity and credit crunch (Brunnermeier, 2009), there was a period where investors in a “flight-to-quality” reaction looked for alternative stores of value in some commodity markets, creating bubbles there that were later destroyed as economic growth slowed. Using powerful tests they developed, PY claim to have corroborated this migrating bubble and crash hypothesis. Using the same data, Figuerola-Ferretti, Gilbert and McCrorie (FGM, 2012) offer more qualified support when considering tests using the same methodology that were made robust in relation to changes to data span and sampling frequency.

The purpose of this paper is to use robust statistical tests to analyze the time series properties of five key commodities over the last decade or so, focusing on possible bubble characteristics around the time of the financial crisis. Our primary aim is to produce an evidence base that we can use to help interpret recent commodity price behavior in the context of some of the competing theories and explanations discussed above. Our analysis

will be of potential relevance to policymakers precisely because these competing theories have different policy implications. If increased financialization really has led to a speculative bubble that interferes with the price discovery and risk-shifting functioning of futures markets, say by inducing suppliers to over-produce and/or reducing the effectiveness of hedging, the case for legislation to limit speculation would be strengthened, even taking account of the costs imposed on some market participants. On the other hand, if the run-up seen in commodity prices over the last decade primarily reflected supply and demand fundamentals, limits on speculation would be less justified and might even harm the efficient functioning of markets.

We consider five commodities – crude oil, gold, silver, copper and aluminum – as being indicative of the energy and metals markets,[3] and consider a data span of twelve years, beginning just after the so-called "Internet Bubble" in the NASDAQ index in the late 1990's. We choose this starting point because it marks the beginning of the Caballero et al. (2008a) account of the recent financial crisis. The commodities can be divided into three distinct types: oil, precious metals and industrial metals. Gold is a precious metal widely acknowledged to be a "store of value", especially in times of crisis; its price increased nearly fourfold over the 12-year period analyzed. The run-up in prices took place especially after mid-2003, showing a small spike in mid-2008, and reaching a high of $1920 in September 2011. Silver, sometimes known as "poor man's gold", offers a useful comparison with gold because, although traditionally weaker as a store of value, is still acknowledged to be an investment asset. Its price over the sample period increased nearly fivefold and, like gold, showed a run-up especially from mid-2003; its run-up in 2006, however, was more marked, and its spike in 2008 more pronounced. Copper and aluminum are both industrially consumed raw materials, with the crucial difference that China is a major importer of copper, but does not import aluminum. Over the sample period, the price of copper increased more than fourfold and the price of aluminum more than doubled. Indeed copper, which averaged under $1,700/ton over 2000-03, was trading at over $3,000/ton by the end of 2004, and over $8,000/ton in May 2006 and again in 2008. Aluminum showed a similar pattern, peaking in 2006 and again in 2008, but unlike copper, its post-2008 prices have still to reach new highs. Last but not least, crude oil is both an industrial commodity used as an input into other activities, and is directly consumed, as petroleum, by final consumers. Its price has risen more than six-fold over the sample period. Specifically, crude oil prices, which had averaged

around $19/barrel in the 1990's, rose to over $30/barrel by the end of 2003, over $50/barrel in 2005 and peaked at over $140/barrel in 2008.

While we concentrate on testing for possible bubble characteristics, we also note two important identification issues that limit our ability to attribute the source of a bubble to individual characteristics that underlie a particular series. Certainly, bubbles can result from the type of speculative behavior discussed above where the basis of pricing departs from fundamental supply and demand considerations – this is a bubble of type embodied in the CFTC definition given above. It is also possible, however, that a bubble in one of the fundamental factors driving the price can manifest itself in the observed series. For example, a crude oil bubble might transmit itself through transport and other costs to generate bubbles in the prices of other commodities. Another consideration is that rational bubbles can emerge purely as endogenous responses to the fundamentals that drive asset prices even in the simplest of underlying structural models (see Branch and Evans, 2011). For these reasons, evidence derived solely from price changes can never be sufficient to allow us to infer that a particular price pattern, such as a bubble, is based on non-fundamental factors. This does not, however, imply that our approach is without meaning; its power derives from the evidence base that we can create from being able to date-stamp the origination and collapse of bubbles in a way that is robust to data span and sampling frequency. This allows us to offer strong conclusions on what theories and explanations are corroborated by or rejected by the data. For example, we can rigorously test whether the 2007-08 oil price spike was a bubble; whether the commodities showed differences in their time series characteristics that would offer evidence against the primacy of an explanation based on increased financialization; and whether the data are consistent with a fundamentals-based explanation, say involving the demand for raw materials from China and the developing world. We are also able to test the predictions of the Caballero et al. (2008a,b) sequential model they proposed to explain the financial crisis.

In order to test for possible bubble behavior, we apply the methodology proposed by PWY and PY that allows for date-stamping the origination and collapse of bubbles, but in a form where critical values are made robust to allow for possibly different data spans and sampling frequencies. The method is statistically rigorous, being based on earlier work by Phillips and Magdalinos (PM, 2007a,b) which offered a solution to the longstanding problem

[3] In a companion paper currently in progress, we are considering commodity price behavior in agricultural markets over the same period.

of constructing appropriate distribution theory in the context of explosive time series;[4] however in recent work, FGM have found that the tests are sometimes not robust to the different data spans and sampling frequencies we expect to encounter with financial data. Accordingly, we will base our empirical analysis on the method that we used to robustify the PWY/PY methodology, which is essentially provides a more careful calibration of critical values.

The plan of the paper is as follows. In Section 2, we briefly outline the basis of the PWY-PY testing methodology and summarize our approach to providing robust critical values for data with respect to different data spans and sampling frequencies, even for a given time series. Section 3 describes the data sets used and presents our empirical findings for crude oil, gold, silver, copper and aluminum price data. Section 4 provides an interpretation of this evidence. Section 5 concludes. Appendix A provides the tables and graphs that underpin the main text, and Appendix B contains some details on the calibration of critical values and on the power of the tests.

2. Testing for financial bubbles: the PWY-PY recursive method and the construction of robust critical values

In this section, we first offer essential background on testing for bubbles, outline the PWY-PY testing methodology and then describe FGM's modification to make the method robust to different data spans and sampling frequencies. Finally, we discuss how to construct robust critical values.

2.1. Essential background on testing for bubbles

Most tests for bubbles are based on the equation[5]

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f} \right)^i E_t(D_{t+i} + U_{t+i}) + B_t, \tag{1}$$

[4] The essential problem addressed by PM, extant since the celebrated paper by Anderson (1959), is that in the standard test for random walk behavior against autoregressive alternatives, no invariance principle applies under explosivity, meaning that any limiting distribution would depend on the actual (unknown) distribution of the disturbances. PM provided an invariance principle for a class of process, called a mildly explosive process (defined here in Section 2), that is wide enough to have empirical relevance. This theory helped pave the way for the tests for bubbles recently constructed by PWY and PY.

[5] See Gürkaynak (2008) and Phillips, Wu and Yu (2011) for up-to-date bibliographies on empirical tests for bubbles. Branch and Evans (2011) provide a reasonably complete bibliography for the rational bubble literature.

where P_t is the after-dividend price of an asset (i.e. the stock price), D_t is the payoff received from the asset (i.e. the dividend), U_t is the unobservable fundamental and r_f is the risk-free interest rate. $P_t - B_t$ is the market fundamental, and B_t defines the bubble component, which is assumed to satisfy the (explosive) property

$$E_t(B_{t+1}) = (1 + r_f)B_t. \tag{2}$$

In the absence of bubbles $(B_t = 0)$, the degree of non-stationarity of the asset price is controlled by the nature of the dividend series and the unobservable fundamentals, and so if, say, D_t is an I(1) process and U_t is either an I(1) or I(0) process, then the asset price is at most an I(1) process, although if U_t exhibits more than I(1) non-stationarity, this will manifest itself in the stock price. Given (2), if the U_t are at most I(1) processes, then explosive behavior in the asset price might appear to provide corroboratory evidence of bubble behavior.

The problem, however, is subtler if the bubbles are periodically collapsing. Evans (1991) proposed a model where (2) continues to hold, but bubbles below a certain threshold grow at a constant lower rate than they do above the threshold where there is an assumed non-zero probability of collapse to a non-zero value before the process resumes its path as before. In a simulation experiment, Evans showed that under this regime, the conventional co-integration test of Diba and Grossman (1988), which relies on a right-tailed unit root test, was not powerful in detecting explosive bubbles. Intuitively, when a time series is viewed *across such episodes*, the observed path seen as a whole can appear like an I(1) process or even a stationary process, suggesting that standard tests for explosivity based on the whole sample could have low power. This intuition provides the motivation to consider tests that are recursive and indeed PWY established in the context of Evans's model that such methods could offer the basis of a powerful test for periodically collapsing bubbles. While PWY in fact provided three tests for bubble behavior, we are primarily interested in their third test, subsequently modified by PY, that involves sequential testing based on forward recursive regressions and allows for date-stamping the origination and collapse of a bubble.

2.2. *The PWY-PY recursive test*: *a statistical method to estimate the origination and collapse of bubbles*

The PY regression model is the following autoregressive (AR) model that is estimated on the basis of a sample of size *n* by ordinary least squares (OLS):

$$x_t = \mu + \delta x_{t-1} + \varepsilon_t \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2) \qquad (t = 1, \ldots, n) \tag{3}$$

where we assume that x_0 is a fixed initial value, and μ, δ and σ $(0 < \sigma < \infty)$ are unknown parameters.[6] The null hypothesis $H_0 : \delta = 1$ is tested against a right-sided alternative that entails (possibly localized) mild explosivity, where the autoregressive parameter is a function of sample size such that $\delta_n = 1 + c/k_n$, c is a constant, and $(k_n)_{n \in N}$ is a deterministic sequence such that $k_n \to \infty$ and $k_n/n \to 0$. This set-up entails a smoother passage away from the null than is implied by the local-to-unity model introduced by Phillips (1987) and Chan and Wei (1987): see PM (2007a, b) for specific details. PY consider the following data generation process (DGP) under the alternative hypothesis[7]

$$\begin{aligned} x_t &= x_{t-1} 1(t < \tau_e \text{ or } t > \tau_f) + \delta_n x_{t-1} 1(\tau_e \le t \le \tau_f) \\ &\quad + \left(\sum_{k=\tau+1}^{t} \varepsilon_k + X^*_{\tau_f}\right) 1\{t > \tau_f\} + \varepsilon_t 1\{t \le \tau_f\} \qquad (t = 1, \ldots, n) \end{aligned} \tag{4}$$

$$\delta_n = 1 + c/n^{\alpha}, \quad c > 0, \quad \alpha \in (0, 1). \tag{5}$$

Under the conditions on c and α, the autoregressive parameter δ_n is greater than unity, and the model has what PM (2007a) call a *mildly-integrated root* (on the explosive side of unity). The system exhibits unit root behavior until τ_e, mildly explosive behavior between τ_e and τ_f, and then resumes its unit root behavior at τ_f, with the process moving to a different level, $X^*_{\tau_f}$, of re-initialization upon the collapse of the bubble. Generally, provided the tests have sufficient power when directed against (4) and (5), tests that reject the null hypothesis of (global) martingale behavior can be interpreted as providing corroboratory evidence for this alternative.

The regression in the first recursion uses $\tau_0 = [n r_0]$ observations for some initially chosen fraction r_0 of the total sample, where [.] denotes the integer part of its argument. Subsequent regressions use this originating data set supplemented successively, giving a sample size of $\tau = [nr]$ for each sample fraction r $(r_0 \le r \le 1)$. The corresponding coefficient test and Dickey-Fuller *t*-statistic, denoted by DF_r^{δ} and DF_r^{t}, are given by

[6] PM more generally allow the initial value to be drawn from a distribution in such a way that the process is initialized at some $x_0 = o_p(\sqrt{k_n})$ independent of the σ-algebra generated by the disturbances $\varepsilon_1, \ldots, \varepsilon_n$. The independent, identically distributed (i.i.d.) assumption can also be relaxed, say to allow for errors that are martingale differences.

[7] The process is formulated with a *zero intercept* since a non-zero intercept would produce a dominating deterministic component of an empirically unrealistic explosive form. See pp. 5-6 of Phillips and Yu (2009).

$$DF_r^{\delta} = \tau(\hat{\delta}_\tau - 1), \qquad DF_r^{t} = \left(\frac{\sum_{j=1}^{\tau} \tilde{X}_{j-1}^2}{\hat{\sigma}_\tau^2} \right)^{1/2} (\hat{\delta}_\tau - 1), \tag{6}$$

where $\hat{\delta}_\tau$ is the OLS estimate of δ based on the first $\tau = [nr]$ observations, $\hat{\sigma}_\tau^2 = \tau^{-1} \sum_{j=1}^{\tau} \left(\tilde{X}_j - \hat{\delta}_n \tilde{X}_{j-1} \right)^2$ is the estimate of σ^2, and $\tilde{X}_{j-1} = X_{j-1} - \tau^{-1} \sum_{j=1}^{\tau} X_{j-1}$. DF_1^{δ} and DF_1^{t} are the usual full-sample test statistics.

To test explicitly for the origination and collapse of a bubble, PWY and PY use forward recursive regressions with the estimate $\hat{\tau}_e = [n\hat{r}_e]$, where

$$\hat{r}_e = \inf_{s \geq r_0} \{ s : DF_s^{\delta} > cv_{\beta_n}^{\delta} \}, \text{ or } \quad \hat{r}_e = \inf_{s \geq r_0} \{ s : DF_s^{t} > cv_{\beta_n}^{df} \}, \tag{7}$$

where $cv_{\beta_n}^{\delta}$ and $cv_{\beta_n}^{df}$ are *prescribed deterministic approximations* (in a sense to be made clear in Section 2.4 below) to the right-sided $100\beta_n\%$ critical values of the finite-sample distributions of the DF_r^{δ} and DF_r^{t} statistics based on $\tau_s = [ns]$ observations, and β_n is the size of the (one-sided) test. Conditional on a bubble origination date $\hat{r}_e$ having been found, the collapse is dated by $\hat{\tau}_f = [n\hat{r}_f]$, where

$$\hat{r}_f = \inf_{s \geq \hat{r}_e + \frac{\log(n)}{n}} \{ s : DF_s^{\delta} > cv_{\beta_n}^{\delta} \}, \text{ or } \quad \hat{r}_f = \inf_{s \geq \hat{r}_e + \frac{\log(n)}{n}} \{ s : DF_s^{\delta} > cv_{\beta_n}^{df} \}, \tag{8}$$

depending on the statistic used. This definition imposes a rule that ensures the duration of the bubble is non-negligible – the collapse date $\hat{r}_f$ involves a search over $\tau_s \geq [n\hat{r}_e + \log(n)]$ – meaning that episodes of smaller order than $\log(n)$ are not considered significant in the dating algorithm.[8] Phillips and Yu (2009) show under general regularity conditions that, under (4), (5) and (8), DF_r^{δ} and DF_r^{t} consistently estimate r_e and r_f, the probability limits of $\hat{r}_e$ and $\hat{r}_f$ respectively.

PY modify the PWY test to address a weakness that it depends on using the first observation of the full sample as the initial observation in each recursive regression. The estimates, $\hat{\tau}_e$ and $\hat{\tau}_f$, of the bubble origination and collapse dates are made robust first by constructing an initial base sample $\{X_{\tau_e - n_m + 1}, \ldots, X_{\tau_e}\}$ of observations, where n_m determines the initial proportion of the sample before $\hat{\tau}_e$ they use, which they assess using

[8] PY incorporate an additional tuning parameter in their minimum bubble-duration condition that they note can be chosen, in principle, on the basis of sampling frequency. They did not, however, use it in their subsequent empirical work, and for this reason we have chosen here to suppress it. See PY, pp. 467-468, for further discussion.

the Schwartz (1978) BIC criterion for unit root or explosive characteristics. The BIC value of the unit root model is

$$\ln\left(\sum_{t=\tau_e-n_{\min}-n_k}^{\tau_e}(\Delta X_t-\overline{X})^2\right)-\ln(n_k+n_{\min})+\frac{\ln(n_k+n_{\min})}{n_k+n_{\min}}, \tag{9}$$

and the BIC value of the autoregression is

$$\ln\left(\sum_{t=\tau_e-n_{\min}-n_k}^{\tau_e}(X_t-\hat{\mu}-\hat{\delta}X_{t-1})^2\right)-\ln(n_k+n_{\min})+\frac{2\ln(n_k+n_{\min})}{n_k+n_{\min}}, \tag{10}$$

where $X=\frac{1}{n_{\min}+n_k}\sum_{t=\tau_e-n_{\min}-n_k}^{\tau_e}X_t$, and $\hat{\delta}$ and $\hat{\mu}$ are the OLS estimators of δ and μ from the autoregressive model (3). If the BIC of the unit root model is smaller than the BIC of the AR(1) model whose point estimate of the autoregressive parameter is greater than unity, then the initial condition is reset to $\tau_e-n_{\min}+1$ and the PWY test is re-performed with this initialization. Otherwise the initial base sample is extended by an additional observation from the past, and the BIC procedure is repeated, and again and again provided the BIC does not select an explosive AR process over a unit root process. The initialization is then taken to be either the point where the BIC first selects an explosive AR process over a unit root process for the sample being considered, or t = 1 if an explosive AR process is never selected. The PY test is then re-performed using this initialization, giving robust estimates $\hat{\tau}_e(\hat{\tau}_0)$ and $\hat{\tau}_f(\hat{\tau}_0)$ that possibly differ from $\hat{\tau}_e$ and $\hat{\tau}_f$. While PY did not demonstrate the effectiveness of the proposed initialization method, we have found it to be instrumental in providing robustness to the empirical work that follows.

2.3. *Making the PWY-PY test robust to different data spans and sampling frequencies*

One essential difficulty with the PWY-PY methodology is that the models (3)-(5) are based on fixed discrete-time AR specifications which are not time-invariant, in consequence of AR models not being self-similar models. This means that if, say, daily data were truly generated from a given AR specification, weekly and monthly data corresponding to the same time series would not satisfy the same model.[9] Using the *same* AR model for weekly and monthly data would therefore entail using a *misspecified* model. FGM show that the lack of time invariance of the model matters when using the PY test, and some of the apparently sharper conclusions obtained by PY in their empirical work become less secure, and are

sometimes even overturned, when the tests are run with the critical values given by PY using different data spans and sampling frequencies of the same time series.[10]

FGM proposed as a remedy to the lack of time invariance an approach analogous to one Phillips (1987a, b) and Perron (1989) took in the conventional unit root testing framework, to embed the discrete-time specifications in a continuous-time model whose solution generates equidistant data that *exactly satisfy* a discrete-time model of the form of (3). The approach, first solved and enacted by Phillips (1972) and described in detail by Bergstrom (1984), allows inference in AR models to take place on the same basis under all sampling schemes. Given a number of observations n, we consider different spans of the data S, and parameterize the sampling frequency by h, with $n = S/h$. FGM show that the appropriate continuous time model that embeds (3) is the Vasicek (1977) model with unknown mean:

$$d\,x(t) = \gamma[x(t) - \mu]\,dt + \sigma\,dW(t), \quad (\gamma > 0), \; x(0) = x_0 \quad (t > 0); \tag{11}$$

where $x(t)$ is a stochastic process of finite variance, μ and γ are parameters with the condition on γ ensuring the process is explosive, and $W(t)$ is standard Brownian motion. For a given sampling frequency h, (11) has an *exact* discrete-time analog:

$$x_{th} = \mu(1-\delta_h(\gamma)) + \delta_h(\gamma)\,x_{(t-1)h} + \sigma\sqrt{\frac{\exp(2\gamma h)-1}{2\gamma}}\,\varepsilon_t, \quad x_{0h} = x_0, \quad (t = 1, 2, \ldots, n) \tag{12}$$

where $\delta_h(\gamma) = \exp(\gamma h)$ and ε_t is $NID(0, 1)$.[11]

A test of the null hypothesis $\gamma = 0$ in (11) is clearly equivalent to a test that $\delta_h = 1$ in (12), and in this case the variance reduces to $\sigma\sqrt{h}$. The regression model (12) has the same form as (3) except that, crucially, it now depends on sampling frequency.

Instead of assessing power using PY's DGP (4) and (5), we now use

$$\begin{aligned} x_{th} &= x_{th-1}\,1(th < \tau_e \text{ or } th > \tau_f) + \delta^*_{n,h}\,x_{th-1}\,1(\tau_e \le th \le \tau_f) \\ &\quad + \left(\sum_{k=\tau+1}^{th} \varepsilon_k + X^*_{\tau_f}\right) 1\{th > \tau_f\} + \varepsilon_t\,1\{th \le \tau_f\} \quad (t = 1, \ldots, n) \end{aligned} \tag{13}$$

$$\delta^*_{n,h} = \exp[(c/k_n)\times h], \quad c > 0, \quad \alpha \in (0, 1). \tag{14}$$

---

[9] The reason is that an unobserved or missing variable problem is induced: see, e.g., Perron (1989), Marcellino (1999) or McCrorie and Chambers (2006). Wei (2006) provides a textbook treatment.

[10] The immediate conclusion to be drawn would be that the tuning parameter PY included to make possible provision for different sampling frequencies should be used to help ensure the robustness of estimates. That said, using such a parameter would bring a whole extra dimension to the problem of statistical inference, given that the size and power properties of the test statistics would depend on it.

[11] We continue to assume a fixed initial value. FGM discuss the possibility that the initial value is drawn from a distribution, and important differences then emerge when comparing (12) with the model used by PM (2007a).

For the purpose of calibrating critical values, daily data can be generated using (12), which are then skip-sampled to generate weekly and monthly data of the same series. Note that skip-sampling daily data generated using (3) would not generate appropriate weekly or monthly data because the model specification itself would only be correct for daily sampling.

2.4. Constructing critical values using the PY and FGM approaches

One of the innovations in the PWY and PY papers, potentially applicable beyond the current setting, is that deterministic functions were used to prescribe critical values that proxy the true critical values pertaining to the finite-sample distributions relevant for each recursive application of the tests. Unusual as this may seem, the efficacy of this approach depends, in the usual way, on the resultant size and power properties of the tests: the better the choice of deterministic function, meaning (whether by accident or design) the closer the function to the true finite-sample distribution of the test statistic, the better will be the statistical test. The potential problem with the approach relates, again, to the extent to which the lack of time invariance of PWY's and PY's model specification requires the deterministic functions to be varied across different data spans and sampling frequencies, because the *same* functions cannot be equally valid *across* data spans and sampling frequencies. This is true even for the *same* commodity series sampled in different ways. PY have demonstrated an awareness of this issue by allowing their minimum bubble-duration condition to depend on a tuning parameter. That said, in the actual implementation of the tests in both papers, the critical value functions used were only loosely calibrated on the basis of results in the Appendix of Phillips and Yu (2009), and no consideration was given as to whether the results would change were different data spans and sampling frequencies to be used given these critical value functions. Also, in PY the same critical value functions were used for the three main series and other secondary series in spite of their having different data spans and sampling frequencies. In PWY and Phillips and Yu (2009, 2010) the critical value functions are of the form

$$cv_{\beta_n}^{\delta} = \kappa_{\delta} * \ln\ln^2[nr]; \text{ and } cv_{\beta_n}^{df} = \kappa_{df} * \ln\ln^2[nr], \tag{15}$$

where κ_{δ} and κ_{df} are constants; and in PY they are of the from

$$cv_{\beta_n}^{\delta} = -0.44 + \ln([nr]/C \text{ and } cv_{\beta_n}^{df} = -0.08 + \ln([nr]/C, \tag{16}$$

where –0.44 and –0.08 are the 5% critical values for the full-sample statistics, DF_1^{δ} and DF_1^{t}, respectively, and C is a constant. FGM offer a systematic discussion of the robustness of the PWY and PY tests, and show that using (3)–(5) and assuming $\kappa_{\delta}, \kappa_{df}$ and C are constants can

materially affect inference across different data spans and sampling frequencies. In response, they allow for a different choice of multiplier for each sampling frequency that is calibrated against the data. Here, we follow the same approach: we generate artificial daily data using (12) under the null, skip-sample to derive artificial weekly and monthly data, and then construct robust critical values for the 12-year data span in hand, for daily, weekly and monthly sampling frequencies. The details of this procedure are discussed in Appendix B.

3. Data and Empirical Findings

The primary data constitute five commodity price series: daily front month futures New York Mercantile Exchange (NYMEX) gold prices, daily front month silver futures NYMEX prices,[12] three month copper and aluminum London Metal Exchange (LME) prices and daily crude oil spot prices downloaded from the Energy Information Administration (EIA) web site. Weekly and daily prices for crude oil were also downloaded from the EIA whereas weekly and monthly prices for the remaining commodities are constructed by skip sampling the daily samples. For crude oil, where PY deflate prices by US end period stocks of crude oil,[13] we consider both the deflated and undeflated series at the weekly and monthly frequency but, due to lack of data availability, only the undeflated series at the daily frequency.

The sample period ranges from January 1999 to December 2010, which covers the decade-long bull-run in commodities, with the start date being chosen around the end of the so-called Internet bubble in the NASDAQ index that marks the beginning of the predictive period of CFG's sequential model.[14] We apply the PWY/PY recursive test methodology using the time-invariant regression model (12), using our calibrated critical value functions for daily, weekly and monthly data frequencies using our specific data span of 12 years (as reported Appendix B).[15] We consider rejections of the null hypothesis in the direction of the DGP given by (13) and (14) based on estimation using the time-invariant regression model. Note that, as is intrinsic to the PWY/PY methodology, the DGP is not a special case of the regression model.

[12] NYMEX future prices were downloaded from Datastream. Rolling is performed the first day of the month.
[13] Source: U.S. Energy Information Administration.
[14] See p. 7 of Caballero et al. (2008a).
[15] Although we use a time-invariant model, we do need to calibrate different critical value functions for different sampling frequencies because fixing the data span means that different sampling frequencies entail different sample sizes, and so different implied finite-sample distributions.

The minimum bubble-duration condition is integral to the empirical implementation of the PWY/PY procedure. We consider a fixed span of 12 years, with daily, weekly and monthly data. Taking a week as comprising 5 trading days, a month 20 trading days, and a year 240 trading days, and applying their log-rule, gives a minimum duration condition of ln(240*12) for daily data, which rounds to eight consecutive days, ln(48*12) for weekly data, which rounds to six consecutive weeks, and ln(12*12) for monthly data, which rounds to five consecutive months.

Tables 1-5 in Appendix A report our empirical findings for crude oil, copper, gold, silver and aluminum based on the time-invariant model. We refer to the procedure that does not check for robustness with respect to the initial values of our series as *basic*, and the procedure that does using PY's BIC-based procedure, as *robust*. For each commodity, we graph the DF_r^t statistics and associated critical values from the daily sample, using both the initial sample start date and that generated by the robust BIC-based procedure.

3.1. Crude oil

Looking first at crude oil, the basic procedure identifies a mildly explosive region in or within the four months April to July 2008, exactly when oil prices peaked. However, this region is too short to meet the minimum duration criterion for a monthly sample. Deflation of the price by stocks alters the estimated bubble start and end dates at the weekly frequency but does not affect the qualitative conclusions. Once we move to the robust procedure, the bubbles seen at the daily and weekly frequencies all evaporate – either because the test statistic falls short of the critical value or because any putative bubble that is found is too short.

The source of these results is readily seen from the DF_r^t statistic using daily data, where the BIC algorithm revises the sample start date from 5 January 1999 to 8 June 2007. However, no bubble is identified using this sample start date – the maximum DF_r^t statistic of 0.56 occurs on 21 May 2008 but this is short of the critical value, relative to the new sample start date, of 0.99. Using the entire sample, the test statistic lies above the critical value function for eight days in May 2008, just long enough to satisfy the minimum bubble length condition of six days – see Figure 1. However, using the BIC-revised sample, the statistic does not approach the (now lower) critical value function. The graphs (not shown) for the DF_r^δ statistic and using weekly data are similar. The monthly data undeflated for stocks do

continue to show a bubble, now robustly identified, but only for the three months May – July 2008.

Our conclusion is that the evidence for a crude oil bubble in 2008 is weak: if there was a bubble, it had only a relatively short duration. This result, taken at face value, suggests the claims of both PY and Gilbert (2010b) of there being a bubble in crude oil in 2008, made on the basis of critical values not calibrated for data span or sampling frequency, are seen to have been premature.

3.2. Non-ferrous metals

Tables 2 and 3 repeat the analysis for the two base metals, aluminum and copper. The basic procedure using daily or weekly data suggests a short aluminum bubble at the start of 2006. This episode is too short to be caught using monthly data. When the robust procedure is used, no bubble is discerned, the evidence of mild explosivity being too short to call a bubble except when using the DF_r^{δ} statistic and the PY-type critical value function. This is illustrated in Figure 2. As in the case of crude oil, the evidence for an aluminum bubble appears weak.

The analysis for copper is more complicated, Gilbert (2010b) having suggested there could be two bubbles, in 2004 and in 2006. At the daily and weekly data frequencies, the basic procedure stops at the 2004 episode, although this fails the minimum duration condition at the monthly frequency. However, the identification of the 2004 episode as a bubble is not robust to the BIC procedure in relation to the starting date. Once the 2004 hurdle has been overcome, the robust procedure identifies the 2006 bubble, although this result depends on the choice of critical value function at the monthly frequency.

3.3. Precious Metals

Using the basic procedure, gold and silver both show evidence of possible bubbles – for gold in 2006 and 2008 while for silver in each of 2004, 2006, 2008 and 2010. Strictly, our DGP of (13) and (14), under which we computed test power, envisages there being only one bubble. To interpret a rejection of the null as providing corroboration for multiple bubbles therefore requires robustness in this direction, which has not been established herein. Homm and Breitung (2012) have suggested that PWY-type tests could indeed be robust to multiple bubbles, although in addressing this issue directly Phillips, Shi and Yu (2012) show that there are potential problems if the duration of the first bubble is longer than the second. As in the previous three cases, the procedure robust to the initial start date offers fewer instances of

bubbles being discerned. In gold, a short bubble (9 or 14 working days, depending on the test) is identified in May 2006 at the daily frequency but this fails to meet the minimum duration requirement at the weekly and monthly frequencies. In silver, a bubble is also confirmed for the spring and early summer (April to June) of 2006, this time at the weekly as well as the daily sampling frequency. However, this bubble is also too short to be detected at the monthly frequency.

3.4. Summary

Overall, the evidence is consistent with 2006 bubbles in copper (where the results are the most clear) and silver and gold. There is much weaker evidence for a bubble in aluminum prices around the same time. The evidence for a bubble in crude oil prices, where the relevant period is the summer of 2008, is even weaker. We have noted there were a number of cases where possible bubbles were rejected through the failure of the minimum bubble-duration condition, in spite of there being some consecutive periods of seemingly mildly explosive behavior. We discuss the efficacy of this condition in the next section.

4. Implications

The results discussed in Section 3 have both methodological and substantive implications. We consider these in turn.

4.1. Methodological issues

In our empirical analysis, we have followed Phillips and Yu (2009), PWY and PY and used their minimum bubble-duration condition where, thinking in terms of test consistency, there is asymptotically a large amount $O(n)$ of data separating bubble origination and collapse. Set in this context, the minimum condition ensures smaller $o(\ln n)$ separations are not considered significant for the dating algorithm. Because the proposed test in essence repeatedly applies a conventional right-sided unit root test to increasing fractions of the total sample, there is the potential for one or two consecutive random draws from the finite sample null distribution, if realised from the right-tail of the distribution, to mislead the investigator. Requiring a certain number of consecutive periods before a bubble is declared offers a way of controlling the Type I error of the test, and Phillips and Yu (2009) show that the given condition does so in a way that ensures the test remains consistent. Here, we explicitly calibrate critical values that embody their minimum bubble-duration condition, for various significance levels on the basis of our chosen time span, for daily, weekly and monthly

sampling frequencies. We note that using higher frequency data provides, in principle, more identifying information about bubbles: bubbles of very short duration, say of just a few days or weeks, could be too short to be able to be picked up using lower-frequency data. On the other hand, the condition relies on our observing a number of *consecutive* periods of possible mild explosivity in order that a bubble be actually declared. It is possible, therefore, that there are two consecutive episodes of mildly explosive behavior that are "close together" but are nonetheless not each of quite of long enough duration to count as a bubble. In such a context, it is just possible that a bubble will be discerned using lower-frequency data, say weekly data, and not be seen in higher-frequency daily data. All in all, we believe it is sensible therefore to report results using all three sampling frequencies. Finally, we note an implication of using the minimum bubble-duration condition as PY did across all sampling frequencies: in practice this will mean the condition entails using a smaller number of time periods, but with a longer time having elapsed, when sampling at lower frequencies than at higher frequencies. We envisage there being scope for using the provision PY made for an extra parameter in their method (the tuning parameter) to put the treatment of daily, weekly and monthly data on a truly consistent basis, and will examine this in future work. In terms of the current paper, we note that our test results are consistent with there having been bubbles identified on the basis of daily data, that were not identified at the weekly frequency (gold) or the monthly frequency (gold, silver, and, using the PWY-type critical value function, copper).

4.2. Substantive Issues

We now discuss the implications of our empirical results taken at face value. These results offer two simple (but contrasting) conclusions:

a) The evidence for a 2008 bubble in crude oil prices is weak.
b) There is substantial evidence for a bubble affecting both non-ferrous and precious metals prices in the first half of 2006. The evidence is clearest for copper and silver and weaker for gold and aluminum.

On the first, it is important to bear in mind the context where in the raw data there is a *known* run-up in the oil price that occurred between 2003 and mid-2008, and a *known* spike in the price in 2007-08. The crucial contribution of our work, recognizing that the size and power properties of a statistical test depend on the time span and sampling frequency of the data, is in the robust calibration of critical values for the span and sampling frequencies in hand. This helps us provide a proper statistical conclusion that cannot be achieved by casual inspection of the data, which might see the data too closely through the lens of 2007-08. Our

two basic tests – the tests that do not potentially correct for initial value impact – suggest there were bubbles of very limited duration in 2008, but our BIC-adjusted tests offers no evidence of a bubble at the 5% significance level. It is just possible that the BIC-adjustment had the effect of imposing a false structural break in the form of a level shift on our specification, that was decisive in causing our null hypothesis not to be rejected, and so we prefer on the basis of our tests to say that the evidence for the 2007-08 oil price spike being a bubble is weak. Certainly, it is not so strong as to suggest there was undue speculation in the oil price that caused Gheit (2008) to assert that:-

> "I firmly believe the current oil price of $135 per barrel is inflated. I believe, based on supply and demand fundamentals, crude oil prices should not be above $60 per barrel . . . I cannot think of any reason that explains the run-up in crude oil process, beside excessive speculation." (as sourced by Irwin, Sanders and Merrin, 2011)

Hamilton (2009) provides a contrary view, offering a fundamentals-based view of the oil price spike based on supply and demand considerations. Another perspective on our result can be given in the context of Kilian and Murphy's (2011) work, where the emphasis was on explaining the run-up in the oil price from 2003 to mid-2008 rather than just the spike of 2007-08. Indeed, taking account of the six-fold price rise over the sampling period, or even just the rise from 2003, rather than just focussing on the spike in 2007-08, and noting the rise in the price thereafter, we see a departure from trend during the spike not being at all obvious. The absence of a crude oil bubble offers evidence against the Caballero et al. (2008a,b) migrating bubble hypothesis, where the promulgated view is that a supposed bubble in the U.S. real estate prices migrates to the crude oil and other commodities markets.

Our conclusion that there is evidence for a 2006 metals bubble is also problematic for the Caballero et al. hypothesis because it arises *too early to be consistent with their story*. This is because the U.S. subprime crisis manifested itself later in 2006 and through early 2007.[16] One interpretation that might be offered to support their hypothesis would be that investors started anticipating the collapse of the U.S. real estate market and began moving into the metals market earlier than Caballero et al. predicted. But if that were the case, one would expect investors also to have moved then into the more prominent crude oil market. A second possibility is that at the beginning of the U.S. subprime crisis, metals including gold and silver were seen as being overvalued, in spite of their historical roles as stores of value, in

[16] In their empirical analysis, PY date the U.S. subprime event as beginning in late 2006, and the $5bn bailout of the small German bank IKB in the last week of July 2007 as being the *public onset* of the subprime crisis. Lehman Brothers failed on 15 September 2008.

consequence of a bubble that had been created earlier for other reasons. If so, it is difficult to imagine that such behavior is consistent with the primacy of increased financialization view in explaining commodity price movements, because this would require evidence of increased co-movement between the energy and metals markets from around 2003.

Our view, therefore, is that it is preferable to consider the 2006 bubbles as a *general metals-market phenomenon*. The commonality of the movements across metals can be explained either in terms of common fundamental driving factors, or because investors and speculators took views on metals as a class, or because of contagion across the group. These explanations are not mutually exclusive. One important common driving factor could be the expansion of Chinese industrial demand against a background of low levels of mine and smelter investment in the 1990's and early 2000's in conjunction with long investment lead times (see Radetzki, 2006). This factor also explains the relatively subdued performance of aluminum compared with copper. It is unclear, however, that this explanation is easily applicable to gold, which may have made its own running as a store of value as the financial crisis developed.

An alternative possibility is that the commonality of these movements in the metals market arises out of the financial market, including "non-fundamental" financial market factors such as the growth of index-related commodity investment and trend-following speculative behavior, as discussed previously. It could be argued that, by amplifying chance price movements, investors might collectively generate the very trends that, at the individual level, they are seeking to follow (see, e.g. De Long et al. 1990).[17] As we have noted, behavior of this sort has the potential to result in bubbles. What is clear, however, that if bubbles were indeed created through this source, it is difficult to argue that they were so excessive as to justify regulation in financial markets.

5. Conclusion

The recursive methodology recently proposed by PWY and PY for testing for bubbles is rapidly establishing itself, having brought a much-needed evidential basis for testing for explosivity and bubbles and, at the same time, being straightforward to implement. FGM have shown there is a potential drawback to the test as initially constructed, owing to its sometimes lack of robustness to changes in data span and sampling frequency. We have addressed this by proposing an approach based on using a time-invariant form of PY's model

specification. This enabled us to calibrate critical value functions against given data spans and sampling frequencies in a way that robustifies the PWY and PY tests and retains all of the advantages of their statistical methodology. Our conclusions based on using these robust critical values are that while there was a general increase in commodity futures prices over 2004 to 2008, the commodities did not move in unison, with metals showing greatest run-up in 2006 and oil in 2008. We did not find strong evidence that the oil price spike in 2007-08 was a bubble but we found substantial evidence for bubble behavior in the metals market affecting both non-ferrous and precious metals at the point of greatest run-up in 2006. Indeed, we have suggested the data exhibit features that suggest the run-up was a general metal markets phenomenon around this time, consistent with the point of strongest demand for industrially-consumed raw materials from China and other developing countries being just before the onset of the U.S. subprime crisis, and the subsequent collapse arising from global repercussions for growth following the subprime crisis. The data do not support the Caballero et al. (2008a,b) hypothesis firstly because the evidence for a bubble in crude oil in 2008 is very weak; and secondly because the greatest run-up in metals prices came towards the end of, but not after, the run-up in U.S. real estate prices that preceded the subprime crisis. Also, our results do not support there having been co-movement of prices between the energy and metals markets. While this observation does not rule out financialization being an important determinant of commodity futures prices generally, it is unable fully to account for and explain the variability seen across commodities, suggesting other factors are essential.

There are, of course, limitations in terms of what can be inferred from a pure time series approach and in future work, we should like formally to test fundamentals-based explanations against increased financialization using a structural model, with particular focus on possible common bubbles in the metals market in 2006. The framework proposed by Kilian (2009) could offer possibilities, but will require the development of new statistical theory to encompass aspects of mild-explosivity and bubbles, and in particular the date-stamping of the origination and collapse of bubbles, as offered in the current paper. While conclusions about the commodities literature can depend upon the lens through which one looks, say microeconomic or macroeconomic, or even economist versus financial analyst, we do feel our empirical results have provided an evidence base that allows for some interpretation and discrimination among competing hypotheses of commodity futures price behavior.

[17] Other studies have failed to find evidence of any impact of financial actors on commodity futures prices. See Sanders and Irwin (2010, 2011) and Stoll and Whalley (2010, 2011).

Acknowledgements

We have benefited greatly from comments from seminar participants at the Federal Reserve and CFTC (Washington DC), the Department of Business Administration, Universidad Carlos III de Madrid, the Centro de Investigación Económica, ITAM, and the University of St Andrews. Figuerola-Ferretti thanks the Spanish Ministry of Education and Science for support under grants SEJ2010-0047-001 and SEJ2011-0031-001.

References

Anderson, T.W., 1959. Asymptotic distributions of estimates of parameters of stochastic difference equations. Annals of Mathematical Statistics 30, 676-687.

Bergstrom, A.R. "Continuous Time Stochastic Models and Issues of Aggregation over Time." In Griliches, Z., Intriligator, M.D. (Eds.) Handbook of Econometrics, volume 2, North-Holland, Amsterdam, pp.1145-1212.

Branch, W.A., Evans, G.W., 2011. Learning about risk and return: a simple model of bubbles and crashes. American Economic Journal: Macroeconomics 3, 159-191.

Brunnermeier, M.K., 2009. Deciphering the 2007-08 liquidity and credit crunch. Journal of Economic Perspectives 23, 77-100.

Buyuksahin, B., Robe, M.A., 2011. Does "paper oil" matter? Energy markets' financialization and equity-commodity co-movements. Working Paper, U.S. Commodity Futures Trading Commission.

Caballero, R.J., Fahri, E., Gourinchas, P.-O, 2008a. Financial crash, commodity prices and global imbalances. Brookings Papers on Economic Activity, Fall, 1-55.

Caballero, R.J., Fahri, E., Gourinchas, P.-O, 2008b. An equilibrium model of 'global imbalances' and low interest rates. American Economic Review 92, 358-393.

Chan, N. H., Wei, C.Z. 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050-1063.

De Long, J.B., Shleifer, A., Summers, L.H., Waldman, R.J., 1990. Positive feedback investment strategies and destabilizing rational expectations. Journal of Finance 45, 379-395.

Diba, B., Grossman, H., 1988. Explosive rational bubbles in stock prices. American Economic Review 78, 520-530.

Evans, G.W., 1991. Pitfalls in testing for explosive bubbles in asset prices. American Economic Review 81, 922-930.

Fabozzi, F.J., Füss, R. and Kaiser, D.G. (2008) The Handbook of Commodity Investing. Wiley, Hoboken (NJ).

Figuerola-Ferretti, I., Gilbert, C.L., McCrorie, J.R., 2012. Recursive tests for explosive behavior and bubbles: extending the Phillips-Wu-Yu methodology to varying data spans and sampling intervals, preprint. **[FGM]**

Fattouh, B., Killian, L., Mahadeva, L., 2012. The role of speculation in oil markets: what have we learned so far? Centre for Economic Policy Research Discussion Paper 8916.

Gheit, F., 2008. Testimony before the Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, U.S. House of Representatives, Washington (DC).

Gilbert, C.L., 2010a. How to understand high food prices. Journal of Agricultural Economics 61, 398-425.

Gilbert, C.L, 2010b. Speculative influence on commodity prices 2006-08. Discussion Paper 197, United Nations Conference on Trade and Development (UNCTAD), Geneva.

Gürkaynak, R.S., 2008. Econometric tests of asset price bubbles: taking stock. Journal of Economic Surveys 22, 166-186.

Hamilton, J.D., 2009. Causes and consequences of the oil shock of 2007-08. Brookings Papers on Economic Activity, Spring, 215-259.

Harris, E.S., Kasman, B.C., Shapiro, M.D., West, K.D., 2009. Oil and the macroeconomy: Lessons for monetary policy. Proceedings of the U.S. Monetary Policy Forum. Available at:- http://research.chicagobooth.edu/igm/events/conferences/2009-usmonetaryforum.aspx

Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: a comparison of alternative methods. Journal of Financial Econometrics, forthcoming.

Hull, J.C., 2006. Options, Futures and Other Derivatives, sixth edition. Prentice Hall, Upper Saddle River (NJ).

Irwin, S.H., Sanders, D.R., 2011. Index funds, financialization and commodity futures markets. Applied Economic Perspectives and Policy 33, 1-31.

Irwin, S.H., Sanders, D.R., 2012. Testing the Masters hypothesis in commodity futures markets. Energy Economics 34, 256-269.

Kilian, L., 2009. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. American Economic Review 99, 1053-1069.

Kilian, L., Murphy, D., 2011. The role of inventories and speculative trading in the global market for crude oil. Working Paper, University of Michigan.

Marcellino, M., 1999. Some consequences of temporal aggregation in empirical analysis. Journal of Business and Economic Statistics 17, 129-136.

Masters, M.W., 2008 Testimony before the U.S. Senate Committee of Homeland Security and Government Affairs, Washington (DC), May 20.

Masters, M.W., 2010. Testimony before the Commodity Futures Trading Commission, 25 March.

Mayer, J, 2010. The financialization of commodity markets and commodity price volatility. In Dullien, S., Kotte, D.J., Marquez, A. Priewe, J. (Eds.) The Financial and Economics Crisis of 2008-2009 and Developing Countries. UNCTAD, New York and Geneva.

Mayer, J., 2011. Financialized commodity markets: the role of information and policy issues. Économie Appliquée 44, 5-34.

McCrorie, J. R., Chambers, M.J., 2006. Granger causality and the sampling of economic processes. Journal of Econometrics 132, 311-326.

Perron, P., 1989. Testing for a random walk: a simulation experiment of power when the sampling interval is varied. In Raj, B. (Ed.) Advances in Econometrics and Modeling, Kluwer, Dordrecht, pp. 47-68.

Phillips, P.C.B., 1972. The structural estimation of a stochastic differential equation system. Econometrica 40, 1021-1041.

Phillips, P.C.B., 1987a. Time series regression with a unit root. Econometrica 55, 277-301.

Phillips, P.C.B., 1987b. Towards a unified asymptotic theory for autoregression. Biometrika 74, 277-301.

Phillips, P.C.B., Magdalinos, T., 2007a. Limit theory for moderate deviations from unity. Journal of Econometrics 136, 115-130.

Phillips, P.C.B., and T. Magdalinos, 2007b. Limit theory for moderate deviations from unity under weak dependence. In Phillips, G.D.A., Tzavalis, E. (Eds.) The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis, Cambridge University Press, Cambridge (UK), pp.123-162.

Phillips, P.C.B., Shi, S.-P., Yu, J., 2012. Testing for multiple bubbles. Cowles Foundation Discussion Paper No. 1843, Yale University.

Phillips, P.C.B., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? International Economic Review 52, 210-226. **[PWY]**

Phillips, P.C.B., Yu, J., 2009. Limit theory for dating the origination and collapse of mildly explosive periods in time series data CoFiE Working Paper No. 5, Singapore Management University.

Phillips, P.C.B., Yu, J., 2010. Dating the timeline of financial bubbles during the subprime crisis. Cowles Foundation Discussion Paper 1770, Yale University.

Phillips, P.C.B., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. Quantitative Economics 2, 455-491. **[PY]**

Radetzki, M., 2006. The anatomy of three commodity booms. Resources Policy 31, 56-64.

Sanders, D. R., Irwin, S.H., 2011. New Evidence on the impact of index funds in U.S. grain futures markets. Canadian Journal of Agricultural Economics 59, 519–532.

Silvenoinen, A., Thorp, S., 2010. Financialization, crisis and commodity correlation dynamics. Research Paper Series 267, Quantitative Finance Research Centre, University of Technology, Sydney.

Singleton, K.J., 2011. Investor flows and the 2008 boom/bust in oil prices. Working Paper, Stanford University. Available at: http://www.stanford.edu/~kenneths/OilPub.pdf

Shi, S., Arora,V., 2012. An application of a model of speculative behaviour to oil prices. Economics Letters 115, 469-472.

Soros, G., 2008. Testimony before the U.S. Senate Commerce Committee Oversight Hearing on FTC Advanced Rulemaking on Oil Market Manipulation, Washington D.C., 4 June 2008.

Stoll, H.R., Whaley, R.E., 2010. Commodity index investing and commodity futures prices. Journal of Applied Finance 20, 7-46.

Stoll, H.R., Whaley, R.E., 2011. Commodity index investing: speculation or diversification? Journal of Alternative Investments 14, 50-60.

Tang, K., Xiong, W., 2011, Index investment and financialization of commodities. Working Paper, Princeton University. (Revised version of NBER Working Paper 16385, 2010.) Available at:- http://www.princeton.edu/~wxiong/papers/commodity.pdf

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177-186.

Wei, W.W.S., 2006. Time Series Analysis. Pearson, Upper Saddle River (NJ).

Appendix A: tables and graphs

Table 1: Crude Oil Bubble Results

	Critical value		Daily		Weekly		Monthly		Weekly (stock-deflated)		Monthly (stock-deflated)	
			DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}
Simple	PWY	Start	06/05/08	16/05/08	18/04/08	09/05/08	too	too	09/05/08	09/05/08	too	too
		End	02/06/08	02/06/08	18/07/08	11/07/08	short	short	11/07/08	11/07/08	short	short
	PY	Start	07/05/08	16/05/08	09/05/08	09/05/08	04/08	04/08	09/05/08	06/06/08	too	too
		End	02/06/08	02/06/08	11/07/08	11/07/08	07/08	07/08	11/07/08	18/07/08	short	short
Robust	PWY	Start	no	no	too	no	too	too	no	too	too	too
		End	bubble	bubble	short	bubble	short	short	bubble	short	short	short
	PY	Start	no	no	no	no	too	too	too	no	too	too
		End	bubble	bubble	bubble	bubble	short	short	short	bubble	short	short

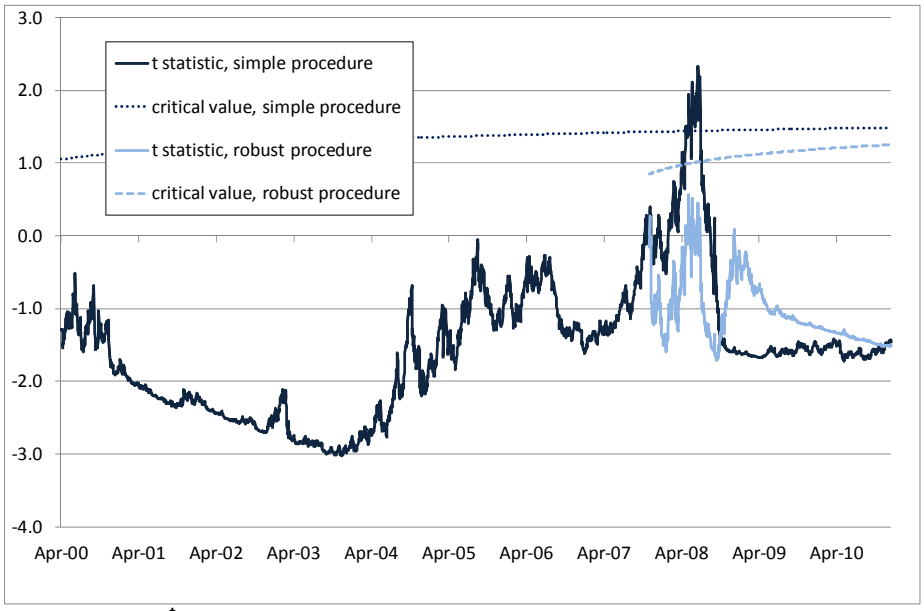

Figure 1: DF^{t}, crude oil, daily data

Figure 1 charts the DF^{t} statistic and the associated PY critical value for the initial start date (dark) of 5 January 1999 and the robust start date of 8 June 2007 as revised by the BIC procedure. Using the robust sample start date, no bubble is identified.

Table 2: Aluminum Bubble Results								
Critical value			Daily		Weekly		Monthly	
			DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}
Simple	PWY	Start	26/01/06	28/04/06	12/04/06	19/04/06	too	too
		End	10/02/06	18/05/06	31/05/06	24/05/06	short	short
	PY	Start	26/01/06	28/04/06	12/04/06	too	too	too
		End	10/02/06	18/05/06	31/05/06	short	short	short
Robust	PWY	Start	too	too	too	too	too	too
		End	short	short	short	short	short	short
	PY	Start	31/01/06	too	28/12/05	too	too	too
		End	09/02/06	short	08/02/06	short	short	short

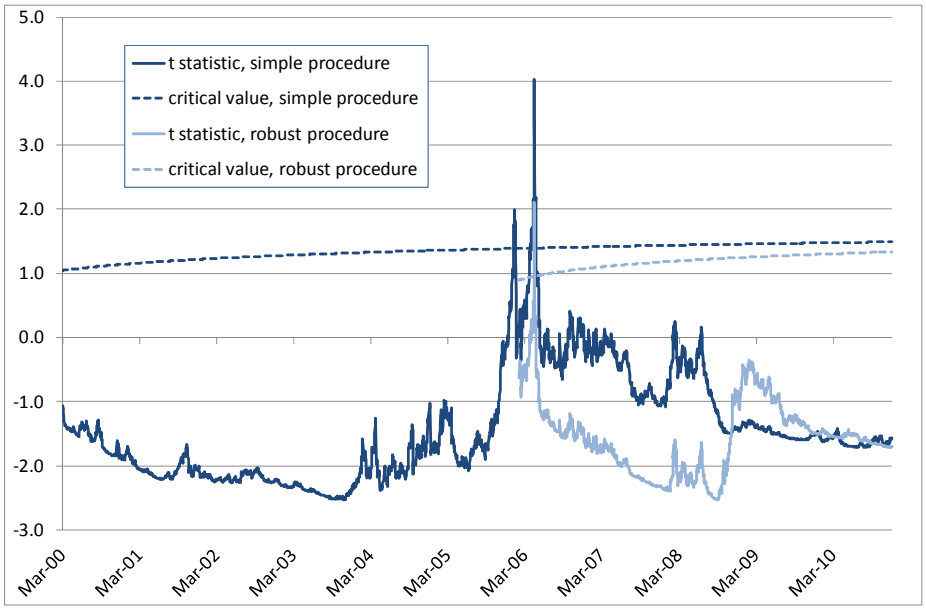

Figure 2: DF^{t} daily data, aluminum

Figure 2 charts the DF^{t} statistic and the associated PY critical value for the initial start date (dark) of 5 January 1999 and the robust start date of 4 August 2004 as revised by the BIC procedure. Using the robust sample start date, the length of the putative 2004 bubble, previously estimated as having lasted for 14 working days using the simple procedure, is seen as having lasted for only three days using the robust procedure. It is therefore disqualified by the minimum duration requirement of 7 days.

Table 3: Copper Bubble Results								
Critical value			Daily		Weekly		Monthly	
			DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}
Simple	PWY	Start	02/02/04	12/02/04	04/02/04	04/02/04	11/05	10/05
		End	22/02/04	20/02/04	21/02/04	21/02/04	10/06	10/06
	PY	Start	02/02/04	12/02/04	04/02/04	11/02/04	07/05	07/05
		End	22/02/04	20/02/04	21/02/04	21/02/04	10/06	11/06
Robust	PWY	Start	13/04/06	10/04/06	12/04/06	05/04/06	too	too
		End	08/06/06	09/06/06	07/06/06	07/06/06	short	short
	PY	Start	12/04/06	10/04/06	12/04/06	05/04/06	03/06	11/05
		End	08/06/06	09/06/06	07/06/06	07/06/06	09/06	10/06

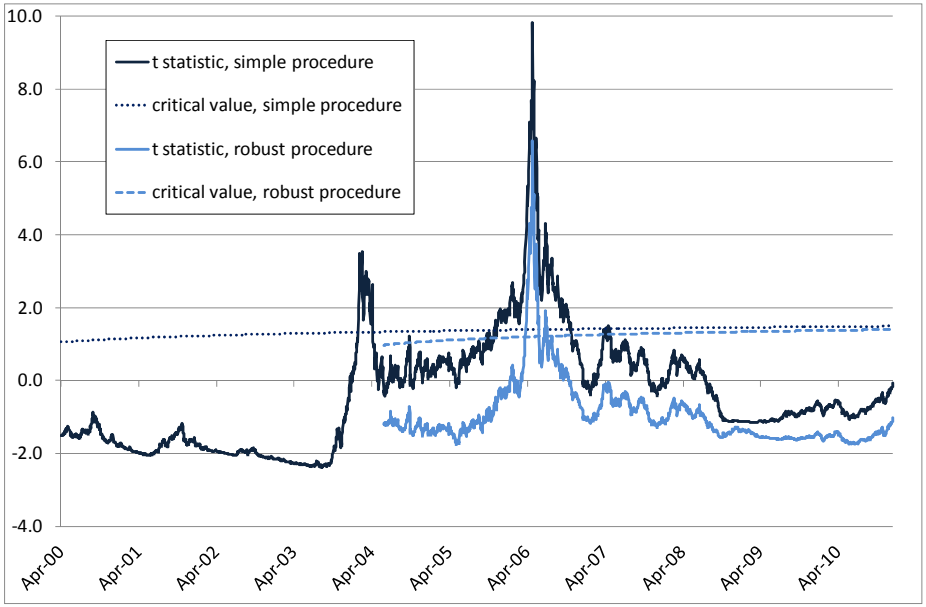

Figure 3: DF^{t}, copper, daily data

Copper appears to exhibit two possible bubble episodes, the first in 2004 and the second in 2006. The simple procedure catches the short 2004 bubble at both the daily and weekly sampling frequency, but discards this episode as too short at the monthly frequency. The robust procedure identifies the longer 2006 bubble. At the daily and monthly sampling frequency, this is dated as running from April to June 2006. The monthly estimates are sensitive to the choice of critical value function – the supposed bubble is too short to qualify if the PWY function is used but lasts for twelve months on the DF^{t} statistic iff the PY function is used.

Table 4: Gold Bubble Results								
Critical value			**Daily**		**Weekly**		**Monthly**	
			DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}
Simple	PWY	Start	03/04/06	03/04/06	29/03/06	29/03/06	01/06	09/07
		End	07/06/06	07/06/06	07/06/06	07/06/06	05/06	07/08
	PY	Start	03/04/06	03/04/06	29/03/06	05/04/06	01/06	01/06
		End	07/06/06	07/06/06	07/06/06	07/06/06	08/06	08/06
Robust	PWY	Start	too	01/05/06	too	too	too	too
		End	short	12/0506	short	short	short	short
	PY	Start	02/05/06	28/04/06	too	too	too	too
		End	12/05/06	17/05/06	short	short	short	short

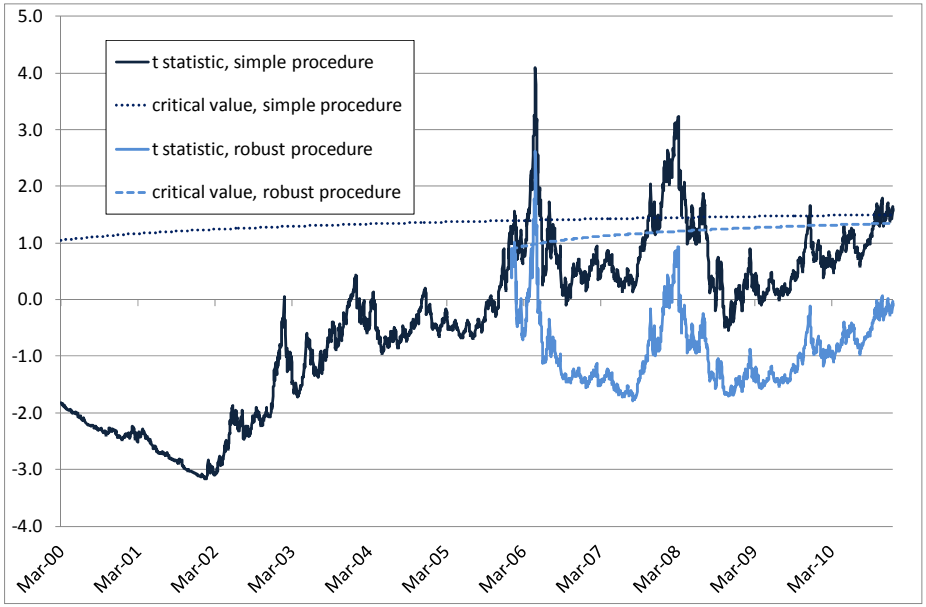

Figure 4: DF^{t}, gold, daily data

The simple estimates suggest that, like copper, gold may exhibit two possible bubble episodes, the first in 2006 and the second in 2008. The simple procedure identifies the 2006 bubble except is the case of the DF_r^t, where the claimed bubble is too short, allowing identification of the supposed 2008 bubble. The robust procedure confirms a short bubble in May 2006 on daily data in three cases out of four, but the bubble is now too short to be detected at the weekly or monthly frequency. The supposed 2008 bubble turns out not to be robust relative to the revised 2004-05 sample start dates.

Table 5: Silver Bubble Results								
Critical value			Daily		Weekly		Monthly	
			DF^{δ}	DF^{t}	DF^{δ}	DF^{t}	DF^{δ}	DF^{t}
Simple	PWY	Start	19/03/04	31/03/04	04/02/04	04/02/04	too	too
		End	12/04/04	12/04/04	21/04/04	21/04/04	short	short
	PY	Start	19/03/04	31/03/04	04/02/04	11/02/04	too	too
		End	12/04/04	12/04/04	21/04/04	21/04/04	short	short
Robust	PWY	Start	06/04/06	10/04/06	12/04/06	05/04/06	too	too
		End	19/04/06	20/04/06	07/06/06	07/06/06	short	short
	PY	Start	03/04/06	06/04/06	12/04/06	05/04/06	too	too
		End	19/04/06	19/04/06	07/06/06	07/06/06	short	short

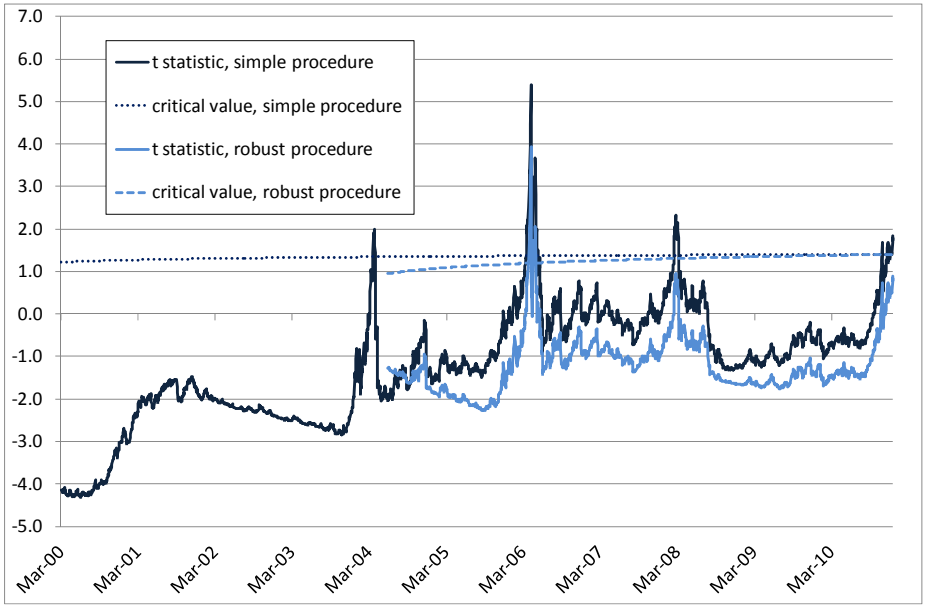

Figure 5: DF^{t}, silver, daily data

The simple estimates suggest four possible bubble episodes, in 2004, 2006, 2008 and again in 2010. The simple procedure identifies the 2004 bubble at the daily and weekly frequency, but all the possible bubbles turn fail to satisfy the minimum duration condition at the monthly frequency. The robust procedure discards the 2004 bubble but confirms a bubble in the first half of 2006 which again fails to meet the minimum duration criterion at the monthly data frequency.

Appendix B: Calibration and power

We explicitly calibrate critical values for the data span in hand, generating artificial daily data using (12) under the null hypothesis.[18] Skip-sampling these daily data generates proper weekly data and monthly data, as (12) is by construction a time-invariant model. We take S = 12, with sampling frequencies h = 1/240 (daily data), 1/48 (weekly data) and 1/12 (monthly data).

The PWY-type critical value functions are used by PWY and Phillips and Yu (2009, 2011), and are of the form

$cv_{\beta_n}^{\delta} = \kappa_{\delta} * \ln\ln^2[nr]$; and $cv_{\beta_n}^{df} = \kappa_{df} * \ln\ln^2[nr]$.

The PY-type critical value functions used by PY are of the form

$cv_{\beta_n}^{\delta} = DF_{1;\beta_n}^{\delta} + \ln([nr]/C$ and $cv_{\beta_n}^{df} = DF_{1;\chi_n}^{t} + \ln([nr]/C$,

where κ_{δ}, κ_{df} and C are given constants, and $DF_{1;\beta_n}^{\delta}$ and $DF_{1;\chi_n}^{t}$ are the β_n% right-sided critical values for the full-sample statistics, DF_1^{δ} and DF_1^{t}, respectively.

Our approach to calibration is to allow the multipliers above to vary, and on the basis of 10,000 simulations of precisely 12 years of generated daily data (corresponding to exactly the span used in the empirical analysis), select values of the multipliers on the basis of a grid search that delivers critical values corresponding to the right-tailed test sizes of 1%, 2.5%, 5% and 10%, approximated to 3 decimal places. Such calibration is therefore undertaken on the basis of the *finite-sample* distributions of the test statistics, and automatically embodies the minimum bubble-duration condition and our required initial choice of fraction r_0 of the total sample on which we base our first estimate using the recursive method. We set r_0 = 0.1, i.e. at 10% of the total sample, to ensure our application of the recursive method begins well before the major run-up in commodity prices seen from 2003.

[Table 6 about here]

[18] Under the null, the constant in the regression model (12) is automatically eliminated, unlike in the standard discrete-time case used by PY, where strictly the test is of the joint hypothesis of zero drift and a unit AR parameter.

Table 6: Calibrated critical values under the FGM methodology for a 12 year span: the *DFt* statistic and PY-type critical value functions

	Daily	Weekly	Monthly	C
5%	5.22	4.583	4.4805	-0.07
2.50%	5.396	4.764	4.408	0.23
1%	5.68	5.117	5.112	-0.07
10%	4.946	5.116	4.092	-0.44

For the power analysis, we now generate data under the DGP (13) and (14). This requires us to specify, in addition to what was needed above, a bubble start date, a bubble end date (and so, implicitly, a bubble duration) and various values of the autoregressive parameter $\delta^*_{n,h}$ under the alternative. This allows us to trace out a power function for each significance level. We consider two possible bubbles under the alternative: one notionally beginning in January 2006 and ending in December 2008, and one corresponding to the bubble PY found in crude oil. In each case we consider four values of the AR parameter under the alternative. We report the same statistics that Phillips and Yu (2009) reported to underpin both PWY and PY.

[Table 7 about here]

Our results provide justification for interpreting rejections of the null hypothesis as corroboration of bubble behavior, and we have additionally given some insight into how power varies with the choice of significance level.