

Using Market Expectations to Test for Speculative Bubbles in the Crude Oil Market*

Efthymios G. Pavlidis

Economics Department, Lancaster University Management School, LA1 4YX, UK
e.pavlidis@lancaster.ac.uk

Ivan Paya[†]

Economics Department, Lancaster University Management School, LA1 4YX, UK
i.paya@lancaster.ac.uk

David A. Peel

Economics Department, Lancaster University Management School, LA1 4YX, UK
d.peel@lancaster.ac.uk

Keywords: crude oil market, speculative bubbles, survey expectations, financial market expectations, recursive unit root tests, predictive regression

JEL codes: C53, G14

*For comments and suggestions, we are grateful to two anonymous referees, Jörg Breitung, Lutz Kilian, Bent Nielsen, participants of the 10th International Conference on Computational Financial Econometrics, the Lancaster Workshop on Macroeconomic and Financial Time Series Analysis, the Applied & Econometrics Workshop at the University of Alicante, and the 26th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics. This study brought together existing data obtained subject to licence restrictions from a number of different sources. Details on how these data were obtained are available at <http://dx.doi.org/10.17635/lancaster/researchdata/215>. We thank Alisa Yusupova for research assistance.

[†]Corresponding author.

Abstract

The wide fluctuations of oil prices from 2003 to 2008 have attracted the interest of academics and policy makers. A popular view is that these fluctuations were caused by speculative bubbles due to the increased financialization of oil futures markets. This hypothesis, however, is difficult to examine since the fundamental price of oil is unobservable and, therefore, econometric evidence in favour of bubbles may actually be due to misspecified market fundamentals. In this paper, we extend two recently proposed methodologies for bubble detection that alleviate this problem by using market expectations of future prices. Both methodologies provide no evidence of speculative bubbles.

1 INTRODUCTION

Oil price shocks may be associated with significant changes in the overall economic performance of countries (Blanchard and Galí, 2010; Baumeister and Kilian, 2016c; Kilian, 2008a, 2014; Hamilton, 2009, 2013). As such, they have been subject to detailed scrutiny for more than three decades (for comprehensive reviews, see Hamilton, 2013, and Baumeister and Kilian, 2016a). Traditionally, research about the key drivers of oil price fluctuations has been dominated by arguments about fundamental supply and demand shocks (Hamilton, 2003; Kilian, 2008b, 2009; Barsky and Kilian, 2002). It is only recently that the role of speculation has been put forward as a serious alternative (see, e.g., Hamilton, 2009; Kilian and Murphy, 2014; Knittel and Pindyck, 2016).

One of the reasons for this recent emphasis on speculation is the eruption of index trading and financialization of oil markets from 2004, and an almost simultaneous surge in the price of oil. The focus on speculative pressures was further fueled by press coverage of the testimonies of Michael Masters before the U.S. Senate (Masters, 2008, 2010). The possibility that financial speculation plays a central role in the determination of oil prices naturally raises the issue of whether bubbles are a feature of the dynamics of oil prices. This view has found

seeming support by studies providing evidence of bubbles in oil prices in the 2000s (see, e.g., Lammerding et al., 2013; Phillips and Yu, 2011; Tsvetanov et al., 2016).

Studies based on structural models, on the other hand, provide strong evidence that the surge in the real price of oil was driven primarily by shifts in the demand for oil associated with global business cycle fluctuations and that speculation did not have an important impact on the spot price (for a summary of this literature, see Fattouh et al., 2013). For example, Kilian and Murphy (2014) develop a structural vector autoregressive model of the global market for crude oil that allows for shocks to the speculative demand for oil. They find that expectations of agents about future demand and supply conditions not already captured by flow demand and flow supply shocks accounted for less than ten percent of the long-run variation in the real price of oil. The conclusion that speculative demand did not play a significant role in the determination of the real price of oil since 2003 is robust to the use of alternative proxies for inventory data and a wide range of values of price elasticities of oil demand (Kilian and Lee, 2014).

In this paper, we contribute to this debate by employing prices and market expectations to analyse the presence of speculative bubbles in the crude oil market. In particular, we extend two recent methodologies, proposed by Pavlidis et al. (2017), to test for periodically collapsing bubbles. The key idea behind these methodologies is simple: In the presence of an ongoing speculative bubble, future spot and expected prices will diverge because market participants rationally attach a non-zero probability to the bubble bursting when forming expectations. Under general conditions, the deviation of the future spot price from the expected price will be a function of the bubble process (and, as such, explosive) but it will not depend on market fundamentals. Consequently, one can simply test for bubbles, first, by running recursive right-tailed unit root tests on the difference between future spot and expected prices and, second, by sequentially testing the unbiasedness hypothesis in the oil market.

Various unit root tests have been proposed in the literature to test for explosive dy-

namics in time series (see, e.g., Homm and Breitung, 2012; Gürkaynak, 2008). In principle, any of these could be used to investigate the existence of bubbles. We choose to work with the Generalized Supremum Augmented Dickey Fuller (GSADF) of Phillips et al. (2015a,b). Simulation evidence in Phillips et al. and Pavlidis et al. (2017) suggests that the GSADF displays good size properties and is superior to its rivals in detecting multiple bubble episodes. Furthermore, this test is accompanied by a date-stamping strategy that permits the identification of the exact periods of market exuberance. With regard to testing the unbiasedness hypothesis, we adopt the rolling-window approach of Pavlidis et al. (2017). This approach consists of sequentially estimating predictive regressions and drawing statistical inference using the IVX instrumentation method of Phillips and Magdalinos (2009), Phillips and Lee (2013), and Kostakis et al. (2015). The IVX method is particularly attractive in this setting because it allows robust chi-square inference for a wide range of autoregressive processes, from stationary to mildly explosive.

The main advantage of the above methodologies is that, by exploiting the information incorporated in market expectations, they do not require the specification of market fundamentals and, thus, ameliorate the well-known joint hypothesis problem. Both methodologies require, instead, a proper measure of market expectations.¹ Pavlidis et al. (2017) use derivative prices to proxy for expectations in the foreign exchange and stock markets. However, derivative prices are contaminated by a risk premium which, if large and persistent, may confound the analysis. In this paper, we employ two alternative measures of expectations that do not suffer from this shortcoming. The first measure is obtained as in Baumeister and Kilian (2016b) by adjusting futures prices by an estimate of the risk premium based on the term structure model of Hamilton and Wu (2014). The second is directly obtained from survey data. By applying the GSADF and IVX tests to price and expectation data on the crude oil market from 1990 to 2013, we find that the two expectation measures yield qualitatively similar results. Overall, there is no evidence in favour of speculative bubbles. This is despite the fact that oil prices and expectations individually exhibit periods of explosive

dynamics. Thus, our findings are consistent with the literature that supports the view that oil price movements are driven by changes in fundamental factors.

The rest of the paper is structured as follows. Section 2 describes the theoretical framework. Section 3 discusses the evidence of risk premia in the oil futures market and the implications for tests based on derivative prices. Section 4 deals with the measurement of market expectations. Section 5 provides a brief outline of the econometric methods employed in the paper. Section 6 presents the empirical application, and the final section concludes.

2 THEORETICAL FRAMEWORK

We adopt a general rational expectations asset pricing framework in which the spot price of oil, S_t , consists of a fundamental component, X_t , and a periodically-collapsing speculative bubble, B_t ,

$$S_t = X_t + B_t. \tag{1}$$

Previous studies that deal with the presence of speculative bubbles in oil markets typically make use of Pindyck's (1993) present value model of rational commodity pricing. In this model, similarly to the present value model for stocks and bonds, the fundamental price of a commodity, X_t , is given by the sum of current and expected discounted future benefits from physically holding the commodity. This stream of benefits, termed the *convenience yield*, includes the operational flexibility of refiners that hold inventories to smooth production in the presence of unexpected changes in demand, avoid stockouts, and accommodate the scheduling of production and sales (Alquist and Kilian, 2010; Alquist et al., 2014; Pindyck, 2001). Putting B_t aside, the present value model predicts that movements in oil prices can be attributed to changes in the latent convenience yield, which in turn reflect changes in current and expected future market conditions. As such, the model can be considered a reduced-form model of dynamic supply and demand. In many empirical applications, the marginal convenience yield is assumed to be integrated of order one, $I(1)$, so that, in the

absence of bubbles, prices and fundamentals are co-integrated (see, e.g., Pindyck, 1993). For our analysis, a specific model for market fundamentals is not required. We make a much less restrictive assumption, that the forecast errors made by rational agents for future market fundamentals are non-explosive. For illustration purposes, we let X_t follow an autoregressive model

$$X_t = \phi X_{t-1} + \theta_t, \quad (2)$$

where ϕ is unrestricted to take values in the I(0), I(1), and explosive regions, and θ_t is a white noise process.

With regard to the second component of the spot price, the speculative bubble B_t satisfies the rational expectations condition of Diba and Grossman (1988)

$$E_t(B_{t+1}) = (1 + r)B_t, \quad (3)$$

where $r > 0$ is a structural parameter that depends on the model describing the economy. A direct implication of this condition and equation (1) is that, if bubbles are present in oil markets, then the spot price will display explosive dynamics. Under risk neutrality, the price of a futures contract at time t with maturity n -periods ahead, defined as $F_t^n \equiv E_t(S_{t+n})$, will also be explosive since, from equations (2) and (3), the expected price is a linear function of the bubble process

$$E_t(S_{t+n}) = E_t(B_{t+n}) + E_t(X_{t+n}) = (1 + r)^n B_t + \phi^n X_t. \quad (4)$$

On this basis, researchers have tested for bubbles by applying right-tailed unit root tests to data for S_t and F_t^n (Gilbert, 2010; Tsvetanov et al., 2016). One drawback of such an approach is that it relies on the implicit assumption that the fundamental component is itself non-explosive. Relaxing this assumption by allowing ϕ to exceed unity implies that the spot rate may inherit explosive dynamics from fundamentals rather than bubbles, which

renders the results of unit root tests inconclusive. Another drawback is that such tests are prone to generating spurious rejections during temporary, but persistent oil demand booms.

An alternative approach that has been employed in the oil literature consists of obtaining an estimate \hat{X}_t for the unobserved market-fundamentals component and, conditional on this estimate, test whether the deviation process from fundamentals $S_t - \hat{X}_t$ is non-stationary (Lammerding et al., 2013; Brooks et al., 2015; Shi and Arora, 2012). This approach is also problematic in that it heavily relies on the researcher using the correct economic model for the oil market. As argued by Flood and Garber (1980), among others, any omitted variable or measurement errors in fundamentals can bias bubble detection tests in favour of rejection of the no-bubble null hypothesis. To put it differently, econometric tests for bubble detection are subject to the well-known joint-hypothesis problem: Rejection of the null hypothesis may be due to misspecified fundamentals, speculative bubbles, or both (Hamilton and Whiteman, 1985; Gürkaynak, 2008).

To mitigate this problem, Pavlidis et al. (2017) propose two novel methodologies that exploit the fact that, in finite samples, periodically-collapsing bubbles cause future spot prices and market expectations -as approximated by futures prices- to diverge. These methods are general enough to cover the most popular types of bubble processes, such as the types proposed by Blanchard (1979) and Evans (1991). To illustrate the basic idea most simply, we let B_t follow the rational bubble process proposed by Blanchard (1979). This process has two states that occur with probabilities π and $1 - \pi$. In the first state, the bubble grows at the rate $(1 + r)/\pi - 1$, while in the second it collapses to a white noise

$$B_{t+1} = \begin{cases} \frac{(1+r)}{\pi} B_t + \epsilon_{t+1}, & \text{with prob. } \pi \\ \epsilon_{t+1}, & \text{with prob. } 1 - \pi. \end{cases} \quad (5)$$

It is easy to verify that this bubble process satisfies the rational expectation condition (3).

Under the scenario that a bubble is erupting between time periods t and $t + n$, the actual

spot rate at $t + n$ is given by

$$S_{t+n} = \phi^n X_t + \left(\frac{1+r}{\pi}\right)^n B_t + \epsilon_{t+n}^*, \quad (6)$$

where ϵ_{t+n}^* consists of two stationary moving average processes, $\sum_{i=1}^n \phi^{n-i} \theta_{t+i}$ and $\sum_{i=1}^n \left(\frac{1+r}{\pi}\right)^{n-i} \epsilon_{t+i}$. Subtracting the expectation formed at time t for the oil price at $t + n$ from the actual spot price at time $t + n$ yields

$$S_{t+n} - E_t(S_{t+n}) = (1+r)^n \left(\frac{1}{\pi^n} - 1\right) B_t + \epsilon_{t+n}^*. \quad (7)$$

The above expression reveals that the future spot price and the expected price will diverge as the bubble process explodes. This wedge between the two prices arises because, when forming expectations, agents rationally attach a positive weight to the event of the bubble bursting. Consequently, their forecast for the growth of the bubble component falls short of the actual realization. The fact that (7) does not depend on market fundamentals but it depends on bubbles implies that explosive dynamics in $S_{t+n} - E_t(S_{t+n})$ can be solely attributed to the latter. Thus, contrary to S_{t+n} and $E_t(S_{t+n})$, right-tailed unit root test results on $S_{t+n} - E_t(S_{t+n})$ can provide conclusive evidence in favour of speculative dynamics.

Pavlidis et al. (2017) focus on the time series properties of the difference between future spot prices and market expectations as proxied by futures prices. However, using the same rationale as above, it is easy to show that the presence of periodically collapsing bubbles will also create a wedge between expectations for S_t made at different time points in the past. From equation (4), we can obtain the following expression for the difference in expectations

of the spot price at time t made n_1 and n_2 periods prior to t , with $n_1 < n_2$

$$\begin{aligned}
E_{t-n_1}(S_t) - E_{t-n_2}(S_t) &= (\phi^{n_1} X_{t-n_1} + (1+r)^{n_1} B_{t-n_1}) - (\phi^{n_2} X_{t-n_2} + (1+r)^{n_2} B_{t-n_2}) \\
&= \sum_{i=0}^{n_2-n_1-1} \phi^{n_1+i} \theta_{t-n_1-i} + (1+r)^{n_2} \left(\frac{1}{\pi^{n_2-n_1}} - 1 \right) B_{t-n_2} + \\
&\quad + \sum_{i=0}^{n_2-n_1-1} \left(\frac{(1+r)^{n_1+i}}{\pi^i} \right) \epsilon_{t-n_1-i}.
\end{aligned} \tag{8}$$

In the absence of a bubble in the price series, the revision of expectations will simply be a stationary moving average of the shocks to fundamentals between periods $t - n_1$ and $t - n_2$. While, if bubbles are present, it will be explosive. On this basis, we suggest running recursive right-tailed unit root tests on $E_{t-n_1}(S_t) - E_{t-n_2}(S_t)$ as a complement to the tests on the difference between the future spot price and its expectation described above. This can provide reassuring evidence about the presence, or absence, of speculative bubbles.

Apart from unit root tests, periodically-collapsing bubbles also have important implications for predictive regressions. Consider the following Fama-type regression

$$S_{t+n} - E_t(S_{t+n}) = \alpha_n + \beta_n (E(S_{t+n}) - S_t) + u_{t+n}, \tag{9}$$

where α_n and β_n are regression coefficients, and u_{t+n} is an error term. Regression (9) and variants of it have been widely used to test the efficiency of asset markets under quadratic loss. In the absence of speculative bubbles, under rational expectations and risk neutrality, the Efficient Market Hypothesis (EMH) postulates that the expected spot price is an unbiased predictor of the actual spot price so that the slope coefficient in (9) is zero. Pavlidis et al. (2017) show that this prediction fails in the presence of an ongoing bubble. By letting fundamentals follow a random walk, they obtain the following expression for the plim of the slope coefficient²

$$\text{plim } \hat{\beta}_n = \frac{(1+r)^n \left(\frac{1}{\pi^n} - 1 \right)}{(1+r)^n - 1} > 0 = \beta^{\text{EMH}}. \tag{10}$$

Since speculative dynamics drive the slope coefficient in the Fama-type regression above the EMH value, researchers can investigate the presence of bubbles by sequentially testing the null hypothesis $H_0 : \beta_n = 0$ against the alternative $H_1 : \beta_n > 0$.

3 RISK PREMIA IN THE OIL FUTURES MARKET

The methodologies of Pavlidis et al. (2017) are based on the assumption of risk neutrality, a direct implication of which is that expected and futures oil prices are indistinguishable. However, under the more realistic scenario of risk-averse behaviour in commodity markets, hedgers will be willing to pay a ‘risk premium’ in the futures market to relieve themselves of commodity price risk, while arbitrageurs will enter the futures market only if they have the expectation of collecting a premium in the form of positive expected returns (Keynes, 1930; Johnson, 1960; De Roon et al., 2000). Recent research provides evidence of significant time-variation of risk premia in the crude oil market. For instance, Hamilton and Wu (2014) document a considerable change in the pricing of risk from 2005, coinciding with an increase in the volume of trading in futures oil contracts. Baumeister and Kilian (2016b) report estimates for a battery of alternative risk-premium models and are the first to quantify risk premia in dollars. The authors show that risk premia are economically significant and exhibit substantial fluctuations over time. The presence of time-varying risk premia in oil markets is also supported by the large number of studies that find predictability in oil futures returns by using various financial and macroeconomic variables (De Roon et al., 2000; Pagano and Pisani, 2009; Etula, 2013).

In light of these results, futures prices constitute a poor measure of market expectations and their use for detecting bubbles may confound econometric tests. By letting ρ_t^n denote the horizon-specific risk premium, the definition of the futures oil price changes to

$$F_t^n \equiv E_t(S_{t+n}) + \rho_t^n \tag{11}$$

and the difference between the future spot and futures prices becomes

$$S_{t+n} - F_t^n = (1 + r)^n \left(\frac{1}{\pi^n} - 1 \right) B_t + \epsilon_{t+n}^* - \rho_t^n. \quad (12)$$

Clearly, in this framework, the time series properties of $S_{t+n} - F_t^n$ depend on ρ_t^n . From both a theoretical and an empirical perspective, it seems unlikely that ρ_t^n displays explosive dynamics and, thus, it may lead to spurious inference in favour of bubbles. But, if large and persistent, the risk premium may mask the bubble process in small samples, lowering the power of unit root tests. That is, a risk premium may lead to higher frequencies of Type II errors. The risk premium can also impact on EMH tests. Following Fama (1984), several studies have illustrated that, if risk premia and fundamentals are correlated, then the presence of a risk premium will cause the slope coefficient in the Fama regression to deviate from the EMH value (e.g., Obstfeld et al., 1996). To avoid complications associated with the use of futures data, we utilize two alternative measures of crude oil price expectations: a *financial market* expectation measure proposed by Baumeister and Kilian (2016b) and a *survey* expectation measure based on data from *Consensus Economics Inc.*

4 MEASURES OF OIL PRICE EXPECTATIONS

Oil price expectations, besides their use in forward-looking economic models, play a central role in generating economic projections, in understanding the extent to which energy markets anticipate changes in economic policy, and in the determination of the value and size of strategic petroleum reserves (Baumeister and Kilian, 2016b; Baumeister et al., 2017; Newell and Prest, 2017). Given their importance and wide use, it is no surprise that different expectation measures have been proposed in the literature. The question of which expectation measure is the *best* is addressed in a recent study by Baumeister and Kilian (2016b).

Baumeister and Kilian (2016b) argue that, in the absence of a fully articulated economic decision problem, the conventional metric in evaluating the accuracy of oil price predictions

is the Mean Square Prediction Error (MSPE), $E(S_{t+n} - \tilde{S}_{t+n})^2$. As is well-known, under a quadratic loss function, the optimal prediction for the future spot price, \tilde{S}_{t+n} , is given by the conditional mean of the price process, $E_t(S_{t+n})$ (Granger, 1969).³ It follows directly from the definition of the futures price (11) that, if a risk premium in the oil futures market exists, then the price of a futures contract will constitute a suboptimal predictor of the future spot price. Furthermore, expectation measures that are based on inaccurate or biased estimates of the risk premium would also be suboptimal. Thus, different expectation measures can be ranked according to their MSPE relative to the MSPE of the futures price.

Baumeister and Kilian (2016b) use the MSPE criterion to identify the most accurate expectation measure for the price of West Texas Intermediate (WTI) crude oil for the period from January 1992 to June 2014 -an exercise that is of direct relevance to our study since we also focus on the WTI price over approximately the same sample period. The authors consider a battery of expectation measures that are obtained by adjusting the futures price by an estimate of the risk premium. These measures fall into three distinct categories. Following Fama and French (1987, 1988), the first category exploits the information contained in the commodity futures basis to form predictions of the risk premium. The second category uses a variety of macroeconomic and financial variables that have been proposed in the literature on oil return predictability; and the third category utilizes estimates of the risk premium generated by term structure models. In total, the authors examine seventeen expectation measures. In addition to these measures, the authors also investigate the predictive ability of a model-averaging strategy, an encompassing model that includes the entire set of macroeconomic and financial variables, and a ‘pre-test’ model that includes only those macroeconomic and financial variables that are found to be statistically significant. On the basis of their extensive investigation, Baumeister and Kilian (2016b) conclude that, by far, the single most accurate measure of oil price expectations is the financial market expectations measure based on the term structure model of Hamilton and Wu (2014).⁴ This measure yields the largest reductions in MSPE with respect to the MSPE for the futures price benchmark (ranging

from 19 percentage points for the 3-month horizon to 36 percentage points for the 12-month horizon) and it always significantly outperforms the simple no-change model.

INSERT FIGURES 1 and 2

Figures 1 and 2 plot financial market expectations together with futures prices and actual future spot prices for 3- and 12-month horizons, respectively. Two points are particularly noteworthy (for a more thorough discussion, see Baumeister and Kilian, 2016a,b). First, it is evident from the figures that financial market expectations perform remarkably better than futures prices in tracking the evolution of actual spot prices, especially at long horizons. Most notably, during the oil market surge of January 2003 to March 2008, financial market expectations are on average higher than future prices. While, after the oil price collapse, from April to September 2008, they are on average lower than futures prices. This implies that the systematic failure of oil futures prices to predict the behaviour of the actual spot price in the 2000s can be partly attributed to the presence of large and time-varying risk premia. Second, while futures prices exhibit wide fluctuations, similar in magnitude to those of actual prices, financial market expectations suggest an evolution of expected prices which is much smoother. As shown by Baumeister and Kilian (2016b), the evolution of financial market expectations over time is also smoother than that of alternative expectation measures. This makes them more plausible from an economic point of view. On the grounds of its predictive accuracy and economic plausibility, we choose to work with the financial market expectation measure proposed by Baumeister and Kilian (2016b) in our empirical application. Other recent studies that apply the same methodology for measuring financial market expectations include Baumeister et al. (2017) and Newell and Prest (2017).

As a complement to financial market expectations, we adopt a second measure based on survey data. Although survey measures may not perform well in terms of predictive accuracy (see, e.g., Alquist et al., 2013), they have the major advantage over all other expectation measures of not being model dependent. As such, they allow us to remain agnostic about the way market expectations are formed. We obtain survey data from *Consensus Economics Inc.*

This is a world-leading international economic survey organization that polls professional economists from a variety of countries each month to acquire their forecasts for energy and metal prices as well as for key macroeconomic variables.⁵ The organization was established in October 1989 and, from its inception up to December 2012, it collected data on 3- and 12-month expectations for the price of WTI crude oil - which matches the type and grade delivered under the NYMEX futures contract (Alquist et al., 2013). In January 2013, the variable of interest in the survey changed to the price of Brent.

The Consensus Economics dataset includes the arithmetic mean, range, and standard deviation of the cross-sectional professional forecasts. We employ the former statistic as our survey expectation measure. The results in Pesaran and Weale (2006) and Pesaran (2006, Appendix A) indicate that the cross-sectional mean displays some attractive properties. Specifically, they illustrate that average expectations can be rational even if individual expectations are biased and display a high degree of heterogeneity. Thus, the conditions under which average expectations are rational are much looser than those required for individual expectations. We plot 3- and 12-month survey expectations in Figures 1 and 2. A glance at the figures reveals that survey expectations are not much smoother than futures prices and are substantially less smooth in comparison to financial market expectations.

5 ECONOMETRIC METHODS & TECHNICAL DETAILS

In this section, we provide a brief description of the econometric tests used in the empirical analysis in order to facilitate the interpretation of the results. We first discuss unit root tests for bubble detection, and then tests for drawing statistical inference in a predictive regression framework with a persistent regressor.

Unit Root Tests Researchers have employed a variety of unit root tests to examine the presence of speculative bubbles in asset markets. We focus on the recently proposed Generalized Supremum Augmented Dickey Fuller (GSADF) of Phillips et al. (2015a,b). This test has several features that make it particularly attractive in our framework: First, contrary to its predecessor -the Supremum ADF (SADF)- the GSADF test is consistent with the presence of multiple boom-bust episodes, such as those experienced in the oil market over the last decades; second, it has good size properties; third, it displays higher power in detecting bubbles that burst in sample than tests that allow for a single change in persistence; and, fourth, it allows dating episodes of explosive dynamics (Homm and Breitung, 2012; Phillips et al., 2015a,b; Pavlidis et al., 2017).

The GSADF test is based on the following rolling-window type ADF regression

$$\Delta y_t = a_{r_1, r_2} + \gamma_{r_1, r_2} y_{t-1} + \sum_{i=1}^k \psi_{r_1, r_2}^i \Delta y_{t-i} + \epsilon_t, \quad (13)$$

where $\epsilon_t \stackrel{iid}{\sim} N(0, \sigma_{r_1, r_2}^2)$, and r_1 and r_2 denote fractions of the total sample size that specify the starting and ending points of a subsample period. The t -ratio corresponding to the null hypothesis of a unit root, $H_0 : \gamma_{r_1, r_2} = 0$, is given by

$$ADF_{r_1}^{r_2} = \hat{\gamma}_{r_1, r_2} / \text{s.e.}(\hat{\gamma}_{r_1, r_2}),$$

and the alternative hypothesis is that of explosive behaviour, $H_1 : \hat{\gamma}_{r_1, r_2} > 0$. Note that setting the beginning of the sample to the first available observation, $r_1 = 0$, and the end of the sample to the last, $r_2 = 1$, yields the standard ADF test, ADF_0^1 . Following Evans (1991), numerous studies have illustrated that this test displays very low power in the presence of market rallies that are followed by market crashes.

To deal with the effect of a single market collapse on the performance of unit root tests, Phillips et al. (2011) propose a recursive procedure which is based on the estimation of the ADF regression (13) using a forward expanding sample. In this procedure, the beginning

of the sample period is fixed at the first available observation, $r_1 = 0$, and the end of the sample period r_2 varies from a user-specified minimum window size, $r_2 = r_0$, to the end of the entire sample, $r_2 = 1$. The supremum of the ADF statistics is defined as

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

The SADF test is designed to perform well when there is a single boom-bust episode, but is inconsistent when there are more than one episodes. The GSADF test of Phillips et al. (2015a) overcomes this deficiency by allowing both the start, r_1 , and the end, r_2 , of the sample used to estimate regression (13) to vary. The GSADF statistic is defined as

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

The extra flexibility of covering a larger number of subsamples makes the test consistent with multiple boom-bust episodes, and it also translates in significant power gains. Moreover, if the GSADF test rejects the null hypothesis of a unit root, Phillips et al. (2015a) propose an econometric strategy that allows the identification of the periods during which the y_t series displays explosiveness. The strategy consists of, first, estimating the sequence of Backward sup ADF statistics

$$\text{BSADF}_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} \text{BADF}_{r_1}^{r_2}, \quad (14)$$

and then identifying the origination date of the bubble as the first observation that the BSADF statistic exceeds its critical value

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \{r_2 : \text{BSADF}_{r_2}(r_0) > scu_{r_2}^{\beta_T}\},$$

and the termination date as the first date after which the BSADF falls below its critical

value

$$\hat{r}_f = \inf_{r_2 \in [r_0, 1]} \{r_2 : \text{BSADF}_{r_2}(r_0) < scu_{r_2}^{\beta_T}\},$$

where $scu_{r_2}^{\beta_T}$ is the $100\beta_T\%$ critical value of the SADF based on $\lfloor r_2 T \rfloor$ observations and β_T is the significance level.

The implementation of the above unit root tests requires the specification of the lag length k and the limit distributions of the SADF and GSADF test statistics. Following the recommendation of Phillips et al. (2015a), we use a short lag length for the spot, S_t , futures, F_t^n , and expectation, $E_t(S_{t+n})$, series. Specifically, we set k to one. With regard to differences between future spot and expectations, $S_{t+n} - E_t(S_{t+n})$, and differences between market expectations made at different points in time, $E_{t-n_1}(S_t) - E_{t-n_2}(S_t)$, these series follow, by construction, a process with serially correlated errors (see equations (7) and (8)). To account for serial dependence, we use the general-to-specific method of Ng and Perron (1995) which determines k by using a sequential t -test for the significance of the regression coefficient on the last lag. We consider a maximum lag length of $k = 11$, so that the longest lag length specification exploits annual information. With regard to the limit distributions of the SADF and GSADF test statistics, these distributions are non-standard and depend on the minimum window size r_0 . We use the rule-of-thumb of Phillips et al. (2015a), $r_0 = 0.01 + 1.8/\sqrt{T}$, to select the minimum window size, and obtain finite-sample critical values by simulating 2,000 random walk processes with $N(0, 1)$ errors.⁶

Predictive Regressions A source of concern when testing the unbiasedness hypothesis is that the predictor in the Fama regression maybe highly persistent, even explosive under the alternative of bubbles. It is a well-known fact that conventional predictability tests of asset prices based on the regression

$$y_{t+1} = \beta x_t + u_{t+1}, \tag{15}$$

may be invalid in the presence of a persistent regressor whose innovations are correlated with the regressand.⁷ Under local-to-unity asymptotics, the ordinary least squares estimator of

the slope coefficient in (15) is biased and has a non-standard and non-pivotal distribution that depends on nuisance parameters (Elliott and Stock, 1994; Stambaugh, 1999; Phillips, 2014). Several methods have been proposed in the literature to draw valid statistical inference in this setting, such as the efficient Q -test of Campbell and Yogo (2006), the conditional likelihood approach of Jansson and Moreira (2006), and the nearly optimal test of Elliott et al. (2015). We choose to work with the IVX instrumentation method of Phillips and Magdalinos (2009), Phillips and Lee (2013), and Kostakis et al. (2015).

The key idea behind the IVX method is to create an instrument, which relies solely on the regressor (hence the name IVX) but falls within the class of near-stationary processes, and use it to remove endogeneity. Specifically, the IVX instrument is initialized at zero and sequentially computed according to

$$z_t = \delta_z z_{t-1} + \Delta x_t,$$

where $\delta_z = 1 + c_z/T^\zeta$ is an artificial autoregressive scalar, $\zeta \in (0, 1)$, and $c_z < 0$.⁸ Given this instrument, the IVX estimator for β is

$$\hat{\beta}_{\text{IVX}} = \frac{\sum z_t y_{t+1}}{\sum z_t x_t},$$

and the corresponding IVX test statistic

$$\text{IVX} = \frac{\hat{\beta}_{\text{IVX}} - \beta}{\hat{\sigma}_{\text{IVX}}},$$

follows a standard normal distribution (Phillips and Magdalinos, 2009; Kostakis et al., 2015; Phillips and Lee, 2013). In addition to allowing standard hypothesis testing, the IVX has several advantages over alternative tests: First, it is valid for autoregressive processes that cover the whole spectrum from stationary to mildly explosive; second, it displays good finite-sample properties (Kostakis et al., 2015; Pavlidis et al., 2017); third, it is easy to implement

since it is based on simple filtering techniques and linear regression; and, finally, it can be extended to long-horizon predictive regressions (see Phillips and Lee, 2013).

To test for periodically collapsing bubbles, we adopt a rolling-window regression framework, where the window size is set, again, according to the rule of Phillips et al. (2015a), $r_0 = 0.01 + 1.8/\sqrt{T}$. Due to the fact that, in a rolling-regression framework, the null hypothesis of no-bubbles is examined multiple times, the probability of rejecting a true null may exceed the desired significance level. To address this issue, we perform a Bonferroni correction which adjusts the nominal significance level by the number of hypotheses tested. In general, Bonferroni corrections lead to conservative tests. However, simulation results in Pavlidis et al. (2017) indicate that the IVX method still performs comparably or better than recursive unit root tests.

6 DATA & EMPIRICAL RESULTS

We construct a dataset that includes spot prices, 3- and 12-month futures prices, and 3- and 12-month expectations for the price of WTI crude oil. Survey expectations are, as already mentioned, obtained from *Consensus Economics Inc.* Spot and futures price series are downloaded from *Bloomberg*. The sample spans the period from January 1990 to December 2013. The start date is chosen so as to avoid the period of low volume in oil futures markets that commenced in 1983 (see Hamilton and Wu, 2014), and the period of regulation of the WTI market which lasted until the early 1980s (Kilian, 2014). Furthermore, it almost coincides with the first data release of survey expectations, which occurred in October 1989. The end of the sample period, on the other hand, corresponds to the last set of survey expectations.

INSERT FIGURE 3

The period under examination includes several significant historical events that, to different degrees, are believed to have affected the price of crude oil. These include the 1990

Iraq’s invasion of Kuwait and its aftermath, the Iraq war of 2002-03, the financialization of oil futures markets since 2003, the Libyan crisis in 2011, the Arab spring, the oil embargo to Iran in 2012, and, most notably, the price fluctuations of 2008 associated with the financial crisis. Figure 3 displays the spot price of WTI crude oil. A look at the figure reveals why the price fluctuations of 2008 stand out as one of the five most significant oil shocks in history. From 92 dollars in December 2007, the oil price jumped to an all-time-high of 145 dollars in July 2008, and then collapsed to 40 dollars by December 2008. These price fluctuations drew the attention of the popular press and of policy makers. For example, in his testimony before the U.S. Senate Commerce Committee in June 2008,⁹ George Soros described the surge in oil prices as part of a “super-bubble” in the economy.

INSERT TABLE 1

The volatile behaviour of oil markets in the 2000s also motivated researchers to test for periodically collapsing bubbles. A number of studies applied unit root tests to spot and futures oil prices. In Table 1, we report ADF, SADF, and GSADF statistics for these series, as well as for survey expectations from *Consensus Economics Inc.* and financial market expectations, derived as in Baumeister and Kilian (2016b). In line with previous studies, the results for the two recursive tests provide strong evidence in favour of explosive dynamics, with all SADF and GSADF statistics exceeding the 95% critical values. In contrast, the standard ADF test fails to reject the null of a unit root in all cases. The large difference between recursive and whole-sample ADF tests results is expected given the extremely low power of the latter.

INSERT FIGURES 4 and 5

In order to identify the specific periods of explosive dynamics in oil prices, Figures 4 and 5 plot BSADF statistics together with the corresponding 95% critical value sequence. Overall, the BSADF statistics for spot, futures and expected price series exhibit a similar

pattern. Neglecting a few extremely short-lived episodes of explosiveness in the first part of the sample, the BSADF statistics rise above the critical value sequence around the inception of the financialization of oil futures markets in 2003, then drop (in some cases below the critical value sequence), and then rise rapidly again to reach their peak value in 2008 -with the exception of 12-month financial market expectations. Shortly after their 2008 peak, all series fall below the critical value sequence and remain insignificant until the end of the sample.

It follows from the analysis of Section 2 that episodes of explosive dynamics in oil prices *per se* do not constitute evidence in favour of speculative bubbles, since such dynamics may actually reflect changes in fundamental factors. Thus, unit root test results on prices alone should be interpreted with caution. To alleviate this problem, we examine the behaviour of deviations of future spot prices from market expectations.

In order to formally test for the presence of speculative bubbles in the oil market, we run unit root tests on the difference between future spot prices and market expectations, as well as on the difference between market expectations made at different points in time. Table 2 reports the unit root tests results. We observe that the ADF, SADF, and GSADF statistics are below the 95% critical value for all expectation measures and for all horizons. The finding that spot and expected prices display episodes of explosive dynamics but their differences do not is in line with the view that fundamentals are the driving force of oil price movements. In accordance, the fact that there are periods of explosiveness in expected prices but there are no such periods in the difference between expectations made at different points in time implies that revisions to expectations can be attributed to news about fundamentals. With respect to the interpretation of the fluctuations of the 2000s in the difference between actual future spot prices and their corresponding expectations, Kilian and Hicks (2013) report large and persistent positive GDP forecast surprises in the major emerging and industrialised economies for the period June 2003 to June 2008, and large and persistent negative forecast surprises for the period July 2008 to December 2008. Similar results are obtained based

on various proxies for global real economic activity (Kilian, 2009; Kilian and Zhou, 2017). Thus, a possible explanation for those large deviations is that the unexpected component of oil price movements during this period is due to flow demand shocks.

INSERT TABLE 2

As a complement to unit root tests, we run IVX tests using the rolling Fama-type regression framework outlined in the previous section. The IVX test results, reported in Table 3, suggest that the null hypothesis that the slope coefficient in the Fama regression is zero (no bubbles) against the alternative that it is positive cannot be rejected at the 5% significance level. This conclusion holds, again, across expectation measures and horizons.

INSERT TABLE 3

Overall, our findings are consistent with the arguments put forward by Alquist and Gervais (2013), Fattouh et al. (2013), Hamilton (2009), Kilian and Murphy (2014), Kilian and Lee (2014), and Knittel and Pindyck (2016) that economic fundamentals are responsible for the fluctuations in crude oil prices over the sample period under consideration.

7 CONCLUSIONS

The fluctuations in the price of crude oil in the 2000s generated an intense debate among academics, financial analysts and policy makers. Many of the arguments put forward in the debate centred on the issue of whether financialization of the oil futures market and the proliferation of index trading created a non-fundamental pathway, a speculative bubble that subsequently collapsed. Such arguments dominated policy discussions, and efforts to regulate speculation in oil markets soon followed.

From an empirical perspective, investigating whether speculative bubbles drove oil prices away from their fundamental value is complicated. In this paper, we employ market expectations of future oil prices to address this problem. Our framework is similar to that of

Pavlidis et al. (2017), with the difference that Pavlidis et al. use derivative prices to mitigate the joint hypothesis problem. A clear advantage of using market expectations over derivative prices is that the latter is not contaminated by a risk premium. This is particularly relevant for oil markets given the strong evidence in favour of significant time-varying, and persistent movements in the risk premium.

In our empirical analysis, we use two alternative measures of expectations to examine the behaviour of the crude oil market from 1990 until 2013. One measure is directly taken from survey data, while the second one is computed by adjusting futures prices by an estimate of the risk premium. In summary, we do not find any evidence of speculative bubbles despite the fact that the individual spot price and expectation series exhibit episodes of exuberance. The results are similar across expectation measures and horizons. We conclude that fundamental factors were the key drivers of the price of crude oil. This finding complements recent research on structural models of the crude oil market that has explicitly incorporated speculation as an independent contributing factor of prices, but found only a limited quantitative effect.

References

- Alquist, Ron, and Lutz Kilian. (2010) “What do we learn from the price of crude oil futures?” *Journal of Applied Econometrics* 25(4), 539–573
- Alquist, Ron, and Olivier Gervais. (2013) “The role of financial speculation in driving the price of crude oil.” *The Energy Journal* 34(3), 35–54
- Alquist, Ron, Gregory H. Bauer, and Antonio Diez de los Rios. (2014) “What does the convenience yield curve tell us about the crude oil market?” Working Paper 2014-42, Bank of Canada
- Alquist, Ron, Lutz Kilian, and Robert J. Vigfusson. (2013) “Forecasting the price of oil.”

- In *Handbook of Economic Forecasting*, ed. Graham Elliott and Allan Timmermann, vol. 2 (Elsevier) pp. 427 – 507
- Barsky, Robert B., and Lutz Kilian. (2002) “Do we really know that oil caused the great stagflation? A monetary alternative.” *NBER Macroeconomics Annual 2001* 16, 137–183
- Batchelor, Roy. (2001) “How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the Consensus.” *Applied Economics* 33(2), 225–235
- Baumeister, Christiane, and Lutz Kilian. (2016a) “Forty years of oil price fluctuations: Why the price of oil may still surprise us.” *Journal of Economic Perspectives* 30(1), 139–160
- (2016b) “A general approach to recovering market expectations from futures prices with an application to crude oil.” manuscript, University of Michigan
- (2016c) “Lower oil prices and the US economy: Is this time different?” *Brookings Papers on Economic Activity* 2016(2), 287–357
- Baumeister, Christiane, Reinhard Ellwanger, and Lutz Kilian. (2017) “Did the renewable fuel standard shift market expectations of the price of ethanol?” NBER Working Papers 23752, National Bureau of Economic Research, Inc, August
- Blanchard, Olivier J. (1979) “Speculative bubbles, crashes and rational expectations.” *Economics Letters* 3(4), 387–389
- Blanchard, Olivier J., and Jordi Galí. (2010) “The macroeconomic effects of oil price shocks: Why are the 2000s so different from the 1970s?” In *International Dimensions of Monetary Policy*, ed. Jordi Galí and Mark Gertler (National Bureau of Economic Research, Inc) pp. 373–421
- Brooks, Chris, Marcel Prokopczuk, and Yingying Wu. (2015) “Booms and busts in commodity markets: Bubbles or fundamentals?” *Journal of Futures Markets* 35(10), 916–938

- Campbell, John Y., and Motohiro Yogo. (2006) “Efficient tests of stock return predictability.” *Journal of Financial Economics* 81(1), 27–60
- De Roon, Frans A., Theo E. Nijman, and Chris Veld. (2000) “Hedging pressure effects in futures markets.” *The Journal of Finance* 55(3), 1437–1456
- Diba, Behzad T., and Herschel I. Grossman. (1988) “The theory of rational bubbles in stock prices.” *Economic Journal* 98(392), 746–54
- Elliott, Graham, and James H. Stock. (1994) “Inference in time series regression when the order of integration of a regressor is unknown.” *Econometric Theory* 10(3-4), 672–700
- Elliott, Graham, Ulrich K. Müller, and Mark W. Watson. (2015) “Nearly optimal tests when a nuisance parameter is present under the null hypothesis.” *Econometrica* 83(2), 771–811
- Etula, Erkki. (2013) “Broker-dealer risk appetite and commodity returns.” *Journal of Financial Econometrics* 11(3), 486–521
- Evans, George W. (1991) “Pitfalls in testing for explosive bubbles in asset prices.” *American Economic Review* 81(4), 922–930
- Fama, Eugene F. (1984) “Forward and spot exchange rates.” *Journal of Monetary Economics* 14(3), 319–338
- Fama, Eugene F., and Kenneth R. French. (1987) “Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage.” *The Journal of Business* 60(1), 55–73
- (1988) “Business cycles and the behavior of metals prices.” *The Journal of Finance* 43(5), 1075–1093
- Fattouh, Bassam, Lutz Kilian, and Lavan Mahadeva. (2013) “The role of speculation in oil markets: What have we learned so far?” *The Energy Journal* 34(3), 7–33

- Flood, Robert P., and Peter M. Garber. (1980) “An economic theory of monetary reform.” *Journal of Political Economy* 88(1), 24–58
- Frankel, Jeffrey A. (2014) “Effects of speculation and interest rates in a carry trade model of commodity prices.” *Journal of International Money and Finance* 42, 88 – 112
- Gilbert, Christopher L. (2010) “Speculative influences on commodity futures prices 2006-2008.” Discussion Paper 197, United Nations Conference on Trade and Development
- Granger, Clive W. J. (1969) “Prediction with a generalized cost of error function.” *Operational Research Quarterly* 20(2), 199–207
- Gürkaynak, Refet S. (2008) “Econometric tests of asset price bubbles: Taking stock.” *Journal of Economic Surveys* 22(1), 166–186
- Hamilton, James D. (2003) “What is an oil shock?” *Journal of econometrics* 113(2), 363–398
- (2009) “Causes and consequences of the oil shock of 2007-08.” *Brookings Papers on Economic Activity* 40(1), 215–283
- (2013) “Historical oil shocks.” In *Routledge Handbook of Major Events in Economic History*, ed. Randall E. Parker and Robert Whaples (New York: Routledge Taylor and Francis Group) pp. 239–265
- Hamilton, James D., and Charles H. Whiteman. (1985) “The observable implications of self-fulfilling expectations.” *Journal of Monetary Economics* 16(3), 353–373
- Hamilton, James D., and Jing Cynthia Wu. (2014) “Risk premia in crude oil futures prices.” *Journal of International Money and Finance* 42(C), 9–37
- (2015) “Effects of index-fund investing on commodity futures prices.” *International economic review* 56(1), 187–205

- Homm, Ulrich, and Jörg Breitung. (2012) “Testing for speculative bubbles in stock markets: A comparison of alternative methods.” *Journal of Financial Econometrics* 10(1), 198–231
- Jansson, Michael, and Marcelo J. Moreira. (2006) “Optimal inference in regression models with nearly integrated regressors.” *Econometrica* 74(3), 681–714
- Johnson, Leland L. (1960) “The theory of hedging and speculation in commodity futures.” *The Review of Economic Studies* 27(3), 139–151
- Keynes, John Maynard. (1930) *A treatise on money* (Macmillan, London). Vol. II
- Kilian, Lutz. (1999) “Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?” *Journal of Applied Econometrics* 14(5), 491–510
- (2008a) “The economic effects of energy price shocks.” *Journal of Economic Literature* 46(4), 871–909
- (2008b) “Exogenous oil supply shocks: How big are they and how much do they matter for the u.s. economy?” *The Review of Economics and Statistics* 90(2), 216–240
- (2009) “Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market.” *American Economic Review* 99(3), 1053–1069
- (2014) “Oil price shocks: Causes and consequences.” *Annual Review of Resource Economics* 6, 133–154
- Kilian, Lutz, and Bruce Hicks. (2013) “Did unexpectedly strong economic growth cause the oil price shock of 2003-2008?” *Journal of Forecasting* 32(5), 385–394
- Kilian, Lutz, and Daniel P. Murphy. (2014) “The role of inventories and speculative trading in the global market for crude oil.” *Journal of Applied Econometrics* 29(3), 454–478
- Kilian, Lutz, and Mark P. Taylor. (2003) “Why is it so difficult to beat the random walk forecast of exchange rates?” *Journal of International Economics* 60(1), 85–107

- Kilian, Lutz, and Thomas K. Lee. (2014) “Quantifying the speculative component in the real price of oil: The role of global oil inventories.” *Journal of International Money and Finance* 42(C), 71–87
- Kilian, Lutz, and Xiaoqing Zhou. (2017) “Modeling fluctuations in the global demand for commodities.” CEPR Discussion Papers 12357, October
- Knittel, Christopher R., and Robert S. Pindyck. (2016) “The simple economics of commodity price speculation.” *American Economic Journal: Macroeconomics* 8(2), 85–110
- Kostakis, Alexandros, Tassos Magdalinos, and Michalis P. Stamatogiannis. (2015) “Robust econometric inference for stock return predictability.” *Review of Financial Studies* 28(5), 1506–1553
- Lammerding, Marc, Patrick Stephan, Mark Trede, and Bernd Wilfling. (2013) “Speculative bubbles in recent oil price dynamics: Evidence from a Bayesian Markov-switching state-space approach.” *Energy Economics* 36(C), 491–502
- Masters, Michael W. (2008) “Testimony of Michael W. Masters, Managing Member/Portfolio Manager, Masters Management Capital, LLC.” Technical Report, Testimony before the U.S. Senate Committee on Homeland Security and Governmental Affairs, May 20
- (2010) “Testimony of Michael W. Masters, Managing Member/Portfolio Manager, Masters Management Capital, LLC.” Technical Report, Testimony before the Commodity Futures Trading Commission, March 25
- Newell, Richard G., and Brian C. Prest. (2017) “Informing SPR Policy through Oil Futures and Inventory Dynamics.” manuscript, Resources for the Future
- Ng, Serena, and Pierre Perron. (1995) “Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag.” *Journal of the American Statistical Association* 90(429), 268–281

- Obstfeld, Maurice, Kenneth S Rogoff et al. (1996) *Foundations of International Macroeconomics*, vol. 30 (MIT press Cambridge, MA)
- Pagano, Patrizio, and Massimiliano Pisani. (2009) “Risk-adjusted forecasts of oil prices.” *The B.E. Journal of Macroeconomics* 9(1), 1–28
- Patton, Andrew J., and Allan Timmermann. (2007a) “Properties of optimal forecasts under asymmetric loss and nonlinearity.” *Journal of Econometrics* 140(2), 884–918
- (2007b) “Testing forecast optimality under unknown loss.” *Journal of the American Statistical Association* 102, 1172–1184
- Pavlidis, Efthymios, Ivan Paya, and David A. Peel. (2017) “Testing for speculative bubbles using spot and forward prices.” *International Economic Review* 58(4), 1191–1226
- Pesaran, M. Hashem. (2006) “Estimation and inference in large heterogeneous panels with a multifactor error structure.” *Econometrica* 74(4), 967–1012
- Pesaran, M. Hashem, and Martin Weale. (2006) “Survey expectations.” In *Handbook of Economic Forecasting*, ed. G. Elliott, C. Granger, and A. Timmermann (Elsevier) pp. 715–776
- Phillips, Peter C. B. (2014) “On confidence intervals for autoregressive roots and predictive regression.” *Econometrica* 82(3), 1177–1195
- Phillips, Peter C. B., and Ji Hyung Lee. (2013) “Predictive regression under various degrees of persistence and robust long-horizon regression.” *Journal of Econometrics* 177(2), 250–264
- Phillips, Peter C. B., and Jun Yu. (2011) “Dating the timeline of financial bubbles during the subprime crisis.” *Quantitative Economics* 2(3), 455–491
- Phillips, Peter C. B., and Tassos Magdalinos. (2009) “Econometric inference in the vicinity of unity.” CoFie Working Paper 7, Singapore Management University

- Phillips, Peter C. B., Shuping Shi, and Jun Yu. (2015a) “Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500.” *International Economic Review* 56(4), 1043–1078
- (2015b) “Testing for multiple bubbles: Limit theory of real-time detectors.” *International Economic Review* 56, 1079–1134
- Phillips, Peter C. B., Yangru Wu, and Jun Yu. (2011) “Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values?” *International Economic Review* 52(1), 201–226
- Pindyck, Robert S. (1993) “The present value model of rational commodity pricing.” *Economic Journal* 103(418), 511–530
- (2001) “The dynamics of commodity spot and futures markets: A primer.” *The Energy Journal* pp. 1–29
- Shi, Shuping, and Vipin Arora. (2012) “An application of models of speculative behaviour to oil prices.” *Economics Letters* 115(3), 469–472
- Stambaugh, Robert F. (1999) “Predictive regressions.” *Journal of Financial Economics* 54(3), 375–421
- Tsvetanov, Daniel, Jerry Coakley, and Neil Kellard. (2016) “Bubbling over! The behaviour of oil futures along the yield curve.” *Journal of Empirical Finance* 38, Part B, 516 – 533
- Vayanos, Dimitri, and Jean-Luc Vila. (2009) “A preferred-habitat model of the term structure of interest rates.” Technical Report, National Bureau of Economic Research

Notes

1. In the speculative bubbles literature, the joint hypothesis problem typically refers to the difficulty in differentiating between a misspecified model for fundamentals (either in terms of functional form, or omitted variables, or both) and the presence of speculative bubbles (Gürkaynak, 2008). Our approach addresses this problem by exploiting the beliefs of economic agents about the evolution of the intrinsic value of crude oil. However, we still require a good proxy for market expectations and, therefore, we face a -though different-joint hypothesis problem. We examine the issue of measuring market expectations in Section 4.

2. It can be shown that the EMH prediction that $\beta_n = 0$ holds even if fundamentals are explosive. Thus, explosive fundamentals cannot explain the failure of EMH.

3. In a more general setting where the loss function is non-quadratic or asymmetric, expectation uncertainty may also matter (see Patton and Timmermann, 2007a, and Patton and Timmermann, 2007b, for the analysis of the properties of optimal forecasts under general loss functions). Measuring expectations uncertainty is not straightforward because it requires information on the degree of confidence that survey respondents attach to their point expectations (Pesaran and Weale, 2006). There are only a few surveys that gather such data and, to the best of our knowledge, none of them covers the oil market.

4. By adapting the bond market models in Vayanos and Vila (2009) and in Hamilton and Wu (2015) to commodity markets, Hamilton and Wu (2014) show that the interaction between hedging demands from commercial producers or index-fund investors and arbitrageurs that hold optimal mean-variance portfolios generates an affine factor structure for futures prices. The authors also develop estimation algorithms and diagnostic tools for this family of models which they have kindly made available at: <http://econweb.ucsd.edu/~jhamilto/software.htm>. We follow Hamilton and Wu's methodology and employ these algorithms to compute point estimates of the risk premium using weekly data, which we subsequently convert to the monthly frequency.

5. Among others, Frankel (2014) advocates the use of Consensus Economics professional forecasts for measuring expected future changes in spot commodity prices, and Batchelor (2001) their use in macroeconomic forecasting.

6. As a robustness check, we have examined the sensitivity of the unit root test results to the lag length specification by running unit root tests with $k = 1, \dots, 11$ and also by employing an alternative information-based method which utilizes BIC to select k . Both exercises suggest that the results are robust to the choice of the lag length specification. Furthermore, rather than using Monte-Carlo simulations to obtain the limit distributions of the unit root test statistics, we have employed a residual-based bootstrap to draw statistical inference. The advantage of this approach is twofold. First, it takes into account the lag structure of the

time series process, and second it allows for deviations of the error term from normality. The results of the residual-based bootstrap are qualitatively the same as those based on Monte-Carlo simulations. All results are available upon request.

7. Kilian (1999) and Kilian and Taylor (2003) show that bootstrap versions of these tests, if properly implemented, are immune from these problems in practice and have good size, even in small samples.

8. Monte Carlo simulation results in Kostakis et al. (2015) suggest setting $c_z = -1$ and $\zeta = 0.95$.

9. George Soros, Testimony before the U.S. Senate Commerce Committee Oversight Hearing on FTC Advanced Rulemaking on Oil Market Manipulation, Washington DC, 3 June 2008.

TABLES & FIGURES

Table 1: Unit root tests (levels)

| Series/Test | ADF | SADF | GSADF |
|-------------------|--------|--------|--------|
| S_t | -1.370 | 4.934* | 5.407* |
| $E_t^C(S_{t+3})$ | -1.649 | 3.832* | 4.416* |
| $E_t^F(S_{t+3})$ | -0.961 | 3.689* | 4.055* |
| F_t^3 | -1.344 | 4.455* | 4.702* |
| $E_t^C(S_{t+12})$ | -1.089 | 5.146* | 5.318* |
| $E_t^F(S_{t+12})$ | -0.201 | 4.110* | 4.716* |
| F_t^{12} | -1.098 | 4.871* | 4.912* |

Notes: The table reports unit root test statistics for the spot oil price, S , the expected 3- and 12-month-ahead spot prices from *Consensus Economics Inc.*, E^C , the expected 3- and 12-month-ahead spot prices obtained as in Baumeister and Kilian (2016b), E^F , and the 3- and 12-month futures prices, F . An asterisk indicates significance at the 5% level. The lag length is set equal to 1. The minimum window size is 32 months.

Table 2: Unit root tests (differences)

| Series/Test | ADF | SADF | GSADF |
|------------------------------------|--------|--------|-------|
| $S_{t+3} - E_t^C(S_{t+3})$ | -5.655 | 0.802 | 0.989 |
| $S_{t+3} - E_t^F(S_{t+3})$ | -5.090 | -0.136 | 1.510 |
| $S_{t+12} - E_t^C(S_{t+12})$ | -5.169 | 1.460 | 1.482 |
| $S_{t+12} - E_t^F(S_{t+12})$ | -4.885 | -1.199 | 0.520 |
| $E_{t-3}^C(S_t) - E_{t-12}^C(S_t)$ | -4.461 | 0.257 | 1.372 |
| $E_{t-3}^F(S_t) - E_{t-12}^F(S_t)$ | -3.964 | -0.259 | 0.995 |

Notes: The table reports unit root test statistics for the difference between the future spot oil price and i) the expected 3- and 12-month-ahead spot prices from *Consensus Economics Inc.*, E^C , and ii) the expected 3- and 12-month-ahead spot prices obtained as in Baumeister and Kilian (2016b), E^F . It also reports unit root statistics for the difference between i) the expected 3- and 12-month-ahead prices at time t from *Consensus Economics Inc.* and ii) the expected 3- and 12-month-ahead prices at time t obtained as in Baumeister and Kilian (2016b). An asterisk indicates significance at the 5% level. The lag length is selected using the general-to-specific method of Ng and Perron (1995). The minimum window size is 32 months.

Table 3: IVX tests

| Horizon | Survey Expectations | Financial Market Expectations |
|-----------|---------------------|-------------------------------|
| 3 months | 2.794 | 1.866 |
| 12 months | 2.325 | 3.277 |

Notes: The table reports maximum IVX statistics corresponding to the null hypothesis that the slope coefficient in the Fama-type regression (9) is zero against the alternative that it is positive. An asterisk indicates significance at the 5% level. The size of the rolling window is 32 months.

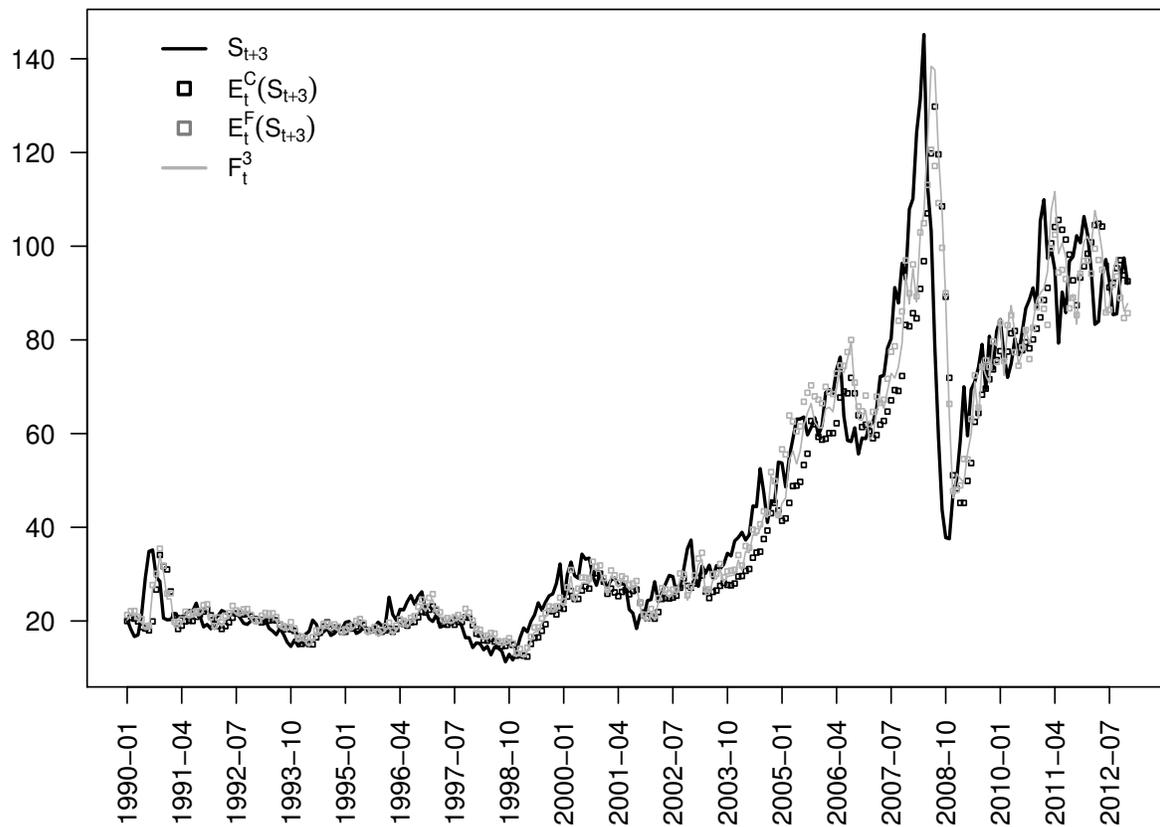


Figure 1: Time series of the actual 3-month-ahead spot oil price, S , the expected 3-month-ahead price from *Consensus Economics Inc.*, E^C , the expected 3-month-ahead spot price obtained as in Baumeister and Kilian (2016b), E^F , and the 3-month futures price, F .

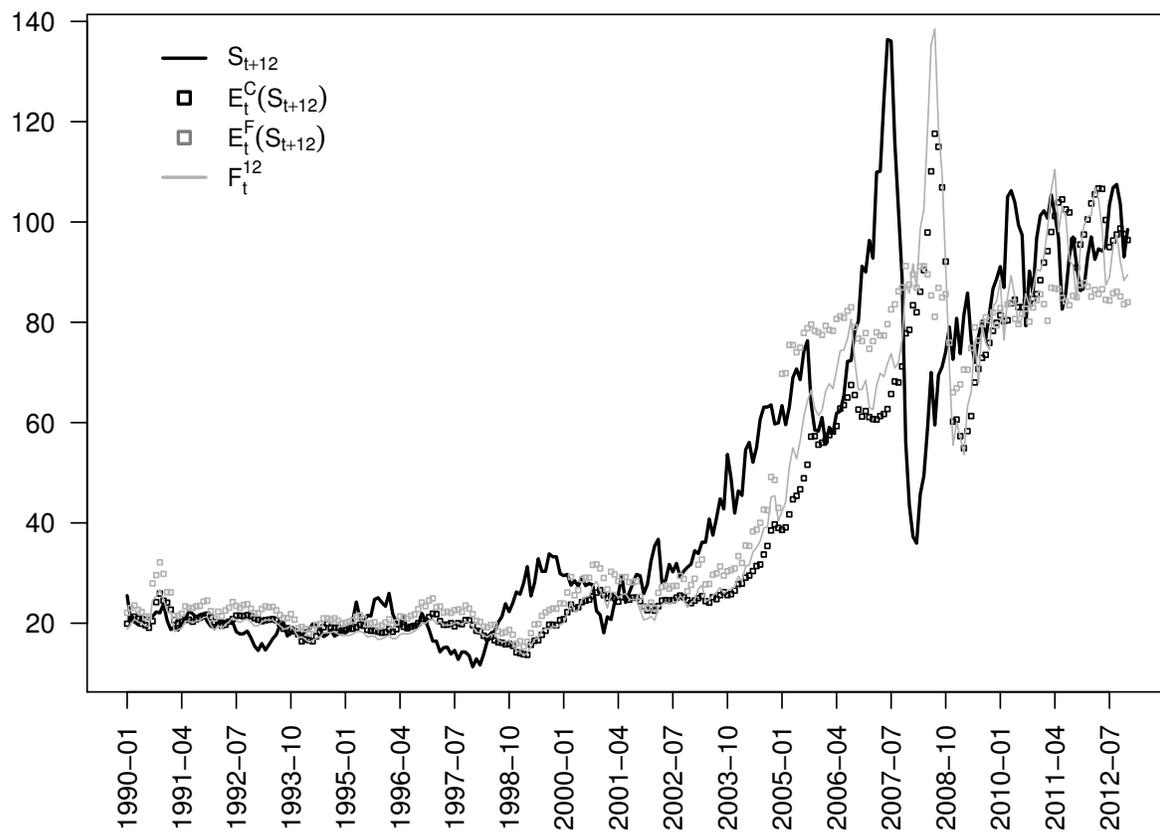


Figure 2: Time series of the actual 12-month-ahead spot oil price, S , the expected 12-month spot price from *Consensus Economics Inc.*, E^C , the expected 12-month spot price obtained as in Baumeister and Kilian (2016b), E^F , and the 12-month futures price, F .

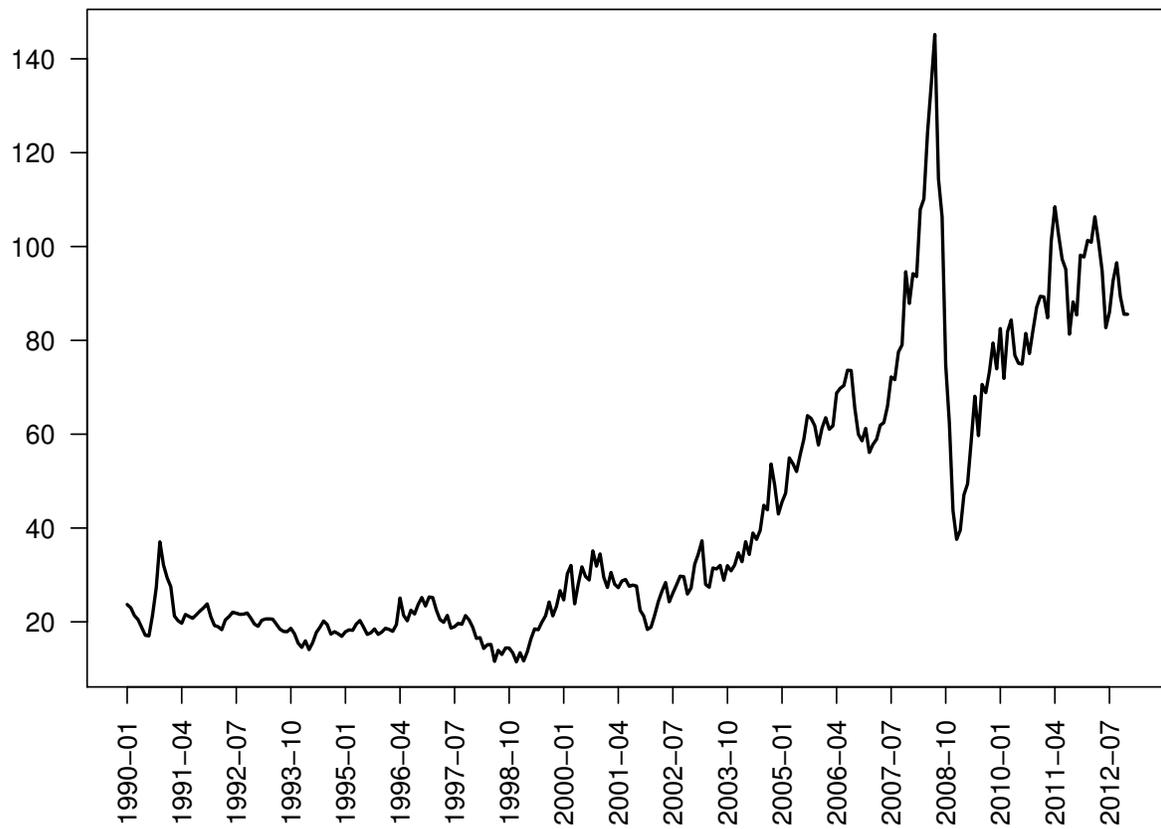


Figure 3: Time series of the spot price of WTI crude oil.

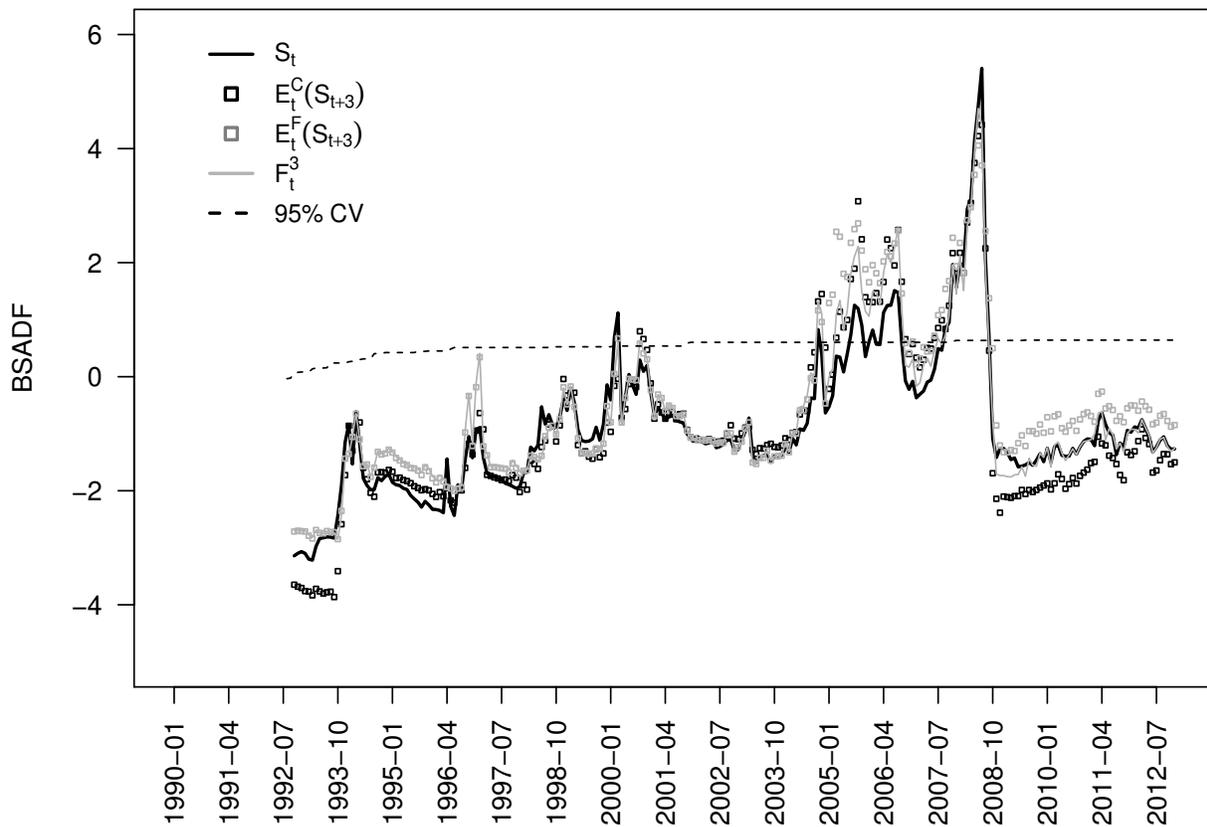


Figure 4: BSADF statistics for the spot oil price, S , the expected 3-month-ahead spot price from *Consensus Economics Inc.*, E^C , the expected 3-month-ahead spot price obtained as in Baumeister and Kilian (2016b), E^F , the 3-month futures price, F , and the 95% BSADF critical value sequence. The lag length is set equal to 1 for all series. The minimum window size is 32 months.

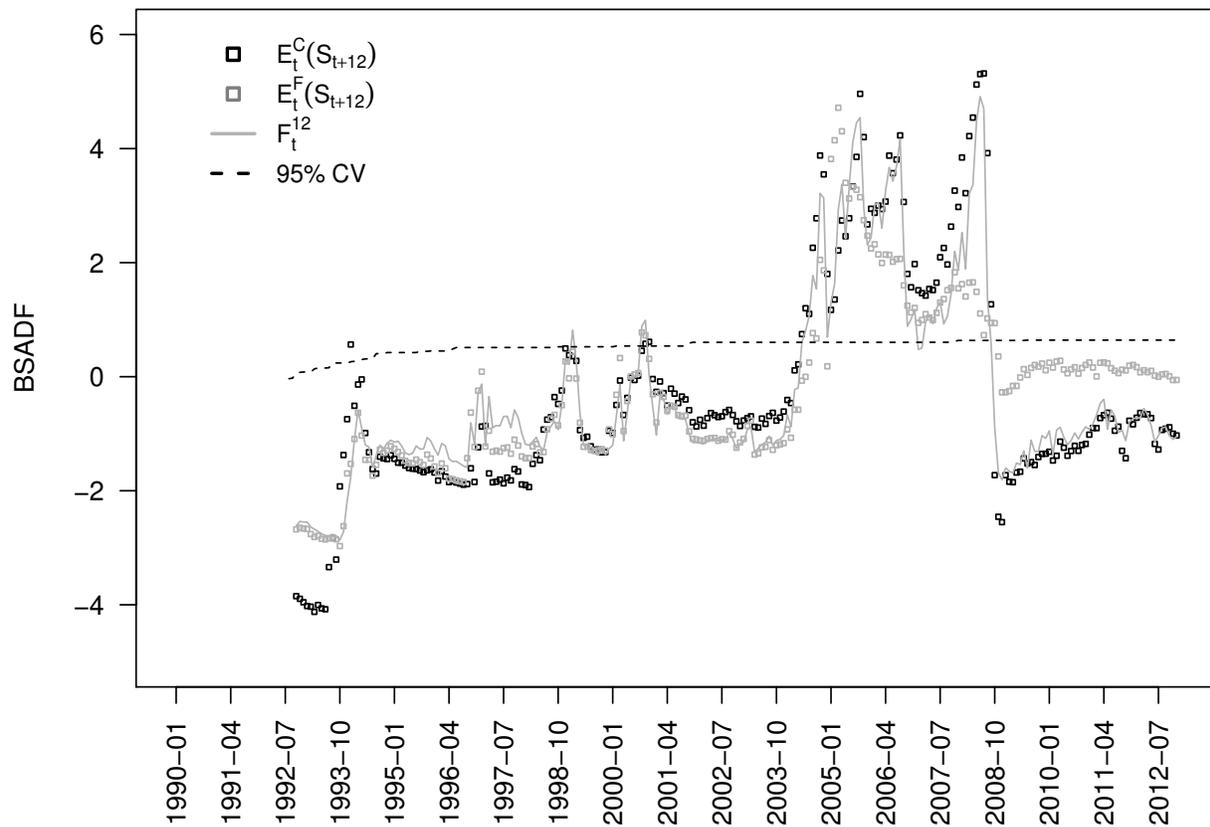


Figure 5: BSADF statistics for the expected 12-month-ahead spot price from *Consensus Economics Inc.*, E^C , the expected 12-month-ahead spot price obtained as in Baumeister and Kilian (2016b), E^F , the 12-month futures price, F , and the 95% BSADF critical value sequence. The lag length is set equal to 1 for all series. The minimum window size is 32 months.