

# IMF Working Paper

---

## Do Commodity Futures Help Forecast Spot Prices?

*David A. Reichsfeld and Shaun K. Roache*

**IMF Working Paper**

Research Department

**Do Commodity Futures Help Forecast Spot Prices?**

**Prepared by David A. Reichsfeld and Shaun K. Roache<sup>1</sup>**

Authorized for distribution by Thomas Helbling

November 2011

**This Working Paper should not be reported as representing the views of the IMF.**

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

**Abstract**

We assess the spot price forecasting performance of 10 commodity futures at various horizons up to two years and test whether this performance is affected by market conditions. We reject efficient markets based on in-sample tests but, out-of-sample, we find that the forecast from the futures market is hard to beat. We find that the forecasting performance of futures does not depend on the slope of the futures curve, in contrast to the predictions of well-known models of commodity markets. We also find futures' forecasting performance to be invariant to whether prices are in an upswing or downswing, casting doubt on aspersions that uninformed investors participating during bull markets impede the price discovery process.

JEL Classification Numbers: G12, G13, G17

Keywords: Asset Pricing, Futures Pricing, Financial Forecasting

Author's E-Mail Address: [dreichsfeld@imf.org](mailto:dreichsfeld@imf.org), [sroache@imf.org](mailto:sroache@imf.org)

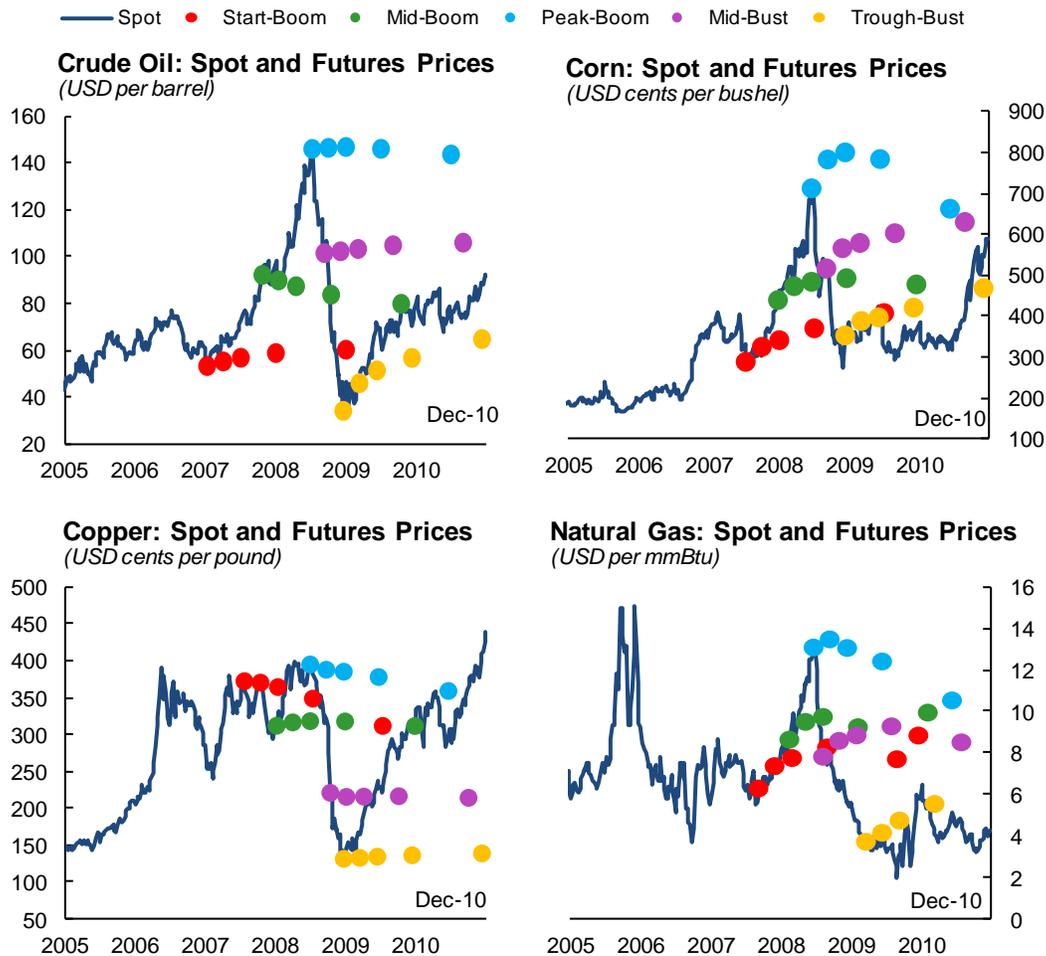
---

<sup>1</sup> We would like to thank Thomas Helbling, Paul Cashin, Menzie Chinn, and colleagues in the Research Department for helpful comments and advice. The usual disclaimer applies.

## 1. Introduction

Why write another paper on the ability of futures to forecast commodity spot prices? There are two main reasons. First, futures prices did a poor job as forecasters during the recent commodity price cycle and it is natural to ask whether we can do better (Figure 1). Second, it is an issue for which matters have yet to be decisively settled. Enhancing the measurement of futures prices, updating the sample period, and broadening the coverage may bring us closer to a definitive answer. Third, forecasting commodity prices is an important, yet often costly, exercise. In particular, for policymakers in countries for which commodity prices can significantly affect the terms of trade, inflation, and poverty levels, obtaining the best possible forecast for commodity prices should be an important priority. Fitting structural or reduced-form models or applying informed judgment to forecast the prices of a wide range of commodities, all of which have different market structures and fundamentals, can be costly and may add no more value than extrapolating from the current price.

Figure 1. Futures Price Curves and Spot Price Developments, 2005-10



Source: Bloomberg.

This paper provides four contributions to the literature on futures as forecasters of commodity spot prices. First, we provide a careful measure of futures prices that exactly matches the horizon of the subsequent change in spot prices and addresses problems posed by illiquid long-dated contracts. Second, we compare the ability of futures prices (and other candidate models) to provide useful forecasts of spot prices at various horizons stretching out two years. For policymakers, the relevant forecast horizon for commodity prices is typically longer than the standard 3 to 12 months tested in much of the literature. Third, we update and broaden the assessment of futures prices as forecasters, including the years since 2002 when commodity market liquidity has greatly increased. Until now, there have been few studies of forecasting by futures markets for a broad range of commodities. Finally, we assess the forecasting ability of futures prices during different market conditions, defined by the futures curve shape (backwardation and contango) and spot price trends (bull and bear markets). This innovation allows us to provide a stricter test of commodity market efficiency and assess the contention that financial investors—attracted by rising commodity prices—have impeded the price discovery process in futures markets.

The plan of this paper is as follows. Section 2 outlines the models to be estimated and the intuition behind the specifications. Section 3 describes the data. Section 4 outlines the main empirical results and discusses their significance. Section 5 provides some brief concluding remarks.

## 2. Model and Empirical Test Specification

### 2.1. Model Specification

This section describes the empirical specifications used in the analysis. The methodology is well-known and we include it here to enhance the clarity of the estimations. We start with a futures pricing model in which the price of a futures contract for a commodity is equal to the discounted value of the expected spot price:

$$F(t, T) = e^{-\rho(T-t)} E^f(t) [S(T)] \quad (1)$$

Where  $F(t, T)$  denotes the futures price in period  $t$  with delivery in period  $T$ ,  $E^f(t)[S(T)]$  is the expectation in the futures market in period  $t$  of the spot price in period  $T$ , and  $\rho$  is the continuously compounded rational expectations risk premium. (In a market without limits to arbitrage, expectations in spot and futures markets will be the same, but we retain this notation for a useful derivation below.) (1) states that a position that requires no upfront investment should deliver an expected return equal to the risk premium of the investor. Taking logs of (1) obtains a linear relationship between the futures price and expected spot price. From this point on, we will consider a forecast length of  $k = T - t$  so that:

$$f_{t, t+k} = -\rho k + E_t^f s_{t+k} + 1/2 \text{var}(s) \quad (2)$$

In (2),  $f$  is the log futures price,  $\rho$  is the (assumed constant) risk premium scaled to the length of the forecast horizon  $k$ ,  $E_t^f s_{t+k}$  is the period  $t$  expectation in the futures market of the log spot

price  $k$  periods ahead, and  $\frac{1}{2} \text{var}(s)$  is Jensen's inequality correction. Henceforth, we will ignore Jensen's term, although it can equivalently be included in the constant term  $\rho k$  in the case of homoscedastic variance, an assumption that we maintain in this paper.

Subtracting the current log spot price  $s_t$  from both sides of (2) gives an expression which states that the current spread between the futures price and spot price (the basis) is equal to the expected change in the spot price for the period until delivery less the risk premium.

$$f_{t,t+k} - s_t = -\rho k + E_t^f (s_{t+k} - s_t) \quad (3)$$

Based on (3), one approach to testing for market efficiency is to estimate the following regression.

$$s_{t+k} - s_t = \alpha + \beta (f_{t,t+k} - s_t) + \varepsilon_{t+k} \quad (4)$$

The parameter  $\alpha$  can be interpreted as the constant component of the risk premium. If the basis provides an unbiased forecast of the future spot price then  $\alpha = 0$ ,  $\beta=1$  and  $\varepsilon_{t+k}$  has a conditional mean of zero. This regression is typically estimated in the market efficiency literature, including for commodities (e.g., Chinn and Coibion, 2010 and Reeve and Vigfusson, 2011). However, this approach glosses over some important details. In particular, what are the implications of assuming rational expectations? How should we test for rational expectations?

Our focus in this paper is on the usefulness of futures as forecasters of spot prices; we are less interested in testing in-sample prediction properties such as efficiency and unbiasedness. Nonetheless, as noted by Clements and Hendry (1998), given that these properties are often considered to be minimum requirements for optimal or rational forecasts (assuming a typical symmetric loss function), this is a good place to begin our assessment of futures prices as forecasters.

It is worth noting at the outset, however, that in-sample tests of efficiency and bias can tell us whether certain models may be useful for forecasting, but they do not evaluate their forecast accuracy. As Granger and Newbold (1977) argue, the distributional properties of forecasters and actual values are almost always different; as a result, a direct comparison of the two is of limited use. More useful forecast evaluation techniques are based on out-of-sample forecast errors. This is the subject of section 4.

## 2.2. In-sample rationality tests

We discuss and test three notions of rationality in this paper. First and weakest of these notions is that the futures market does not make persistent in-sample prediction errors and that, over the "long run", expected errors are zero. In other words, while the prediction errors may exhibit some persistence and serial correlation, they eventually converge to zero. A stronger assumption is that the in-sample futures market prediction error  $v$  is an independent serially-uncorrelated white noise process:

$$s_{t+k} - E_t^f s_{t+k} = v_{t+k} \quad E_t v_{t+k} = 0 \quad E_t v_{t+k} v_{t-j} = 0 \quad j = 0, k, 2k, 3k, \dots \quad (6)$$

Substituting (6) into (2) obtains a linear relationship between the futures price in period  $t$  with delivery  $k$  periods ahead and the realized spot price in period  $t+k$ .

$$f_{t,t+k} = -\rho k + \beta s_{t+k} + \omega_{t+k} \quad \omega_{t+k} = -\beta v_{t+k} \quad (7)$$

(7) states that the current futures price is a linear function of the realized spot price and that, if  $f$  and  $s$  are I(1) processes, then the two series must be cointegrated with a cointegrating vector  $(1, -\beta)$ , a restriction that can be tested, where the coefficient  $\beta$  generalizes the specification.

The second notion of rationality is weak-form efficiency. This means that the current futures price incorporates all the information useful for in-sample prediction. To assess efficiency, (7) can be rearranged to bring the realized spot price change to the left-hand side. This is similar, but different in an important way, to (4):

$$s_{t+k} - s_t = \phi_0 + \phi_1 (f_{t,t+k} - s_t) + (\phi_1 - 1) s_t + v_{t+k} \quad (8)$$

where  $\phi_0 = \rho k / \beta$ ,  $\phi_1 = 1 / \beta$  and the residual  $v$  is the prediction error from (7). Note that (4) implies one important restriction on (8), namely that the cointegrating vector is  $(1, -1)$  so that the coefficient on the lagged spot price level is zero and can therefore be excluded from the relationship. A general representation of (8) that is useful if  $f$  and  $s$  are I(1) processes is an error-correction specification. This represents the realized spot price change as a process that allows for the presence of lagged price changes as shown below:

$$\Delta s_{t+k} = \alpha + \delta (\phi_0 + \phi_1 f_{t-k,t} - s_t) + \mu \Delta f_{t,t+k} + \sum_{j=1}^K \phi_{1+j} \Delta f_{t-kj,t-k(j-1)} + \sum_{l=0}^M \gamma_l \Delta s_{t-lk} + v_{t+k} \quad (9)$$

In this model, cointegration requires that  $\delta \neq 0$  to ensure that the realized spot price change responds to deviations in the long-run equilibrium. As shown by Beck (1994), weak-form short-run market efficiency requires the following conditions:

$$\delta = 1 \quad \delta \phi_1 = \mu \neq 0 \quad \phi_{1+j} = \gamma_l = 0 \quad \forall j > 0, l \geq 0 \quad (10)$$

This assertion can be checked by rewriting (10) as:

$$s_{t+k} = (\alpha + \delta \phi_0) + (1 - \delta) s_t + (\delta \phi_1 - \mu) f_{t-k,t} + \mu f_{t,t+k} + \sum_{j=1}^K \phi_{1+j} \Delta f_{t-kj,t-k(j-1)} + \sum_{l=0}^M \gamma_l \Delta s_{t-lk} + u_{t+k} \quad (11)$$

If we reject the null hypothesis (10), then the coefficients on the lagged levels of the spot price and the futures price (i.e., the futures price at period  $t-k$ ) would help predict the spot price,

violating the efficiency condition.

The final and strictest notion of rationality is that the futures market is unbiased. This incorporates weak-form efficiency but also requires that there is no risk premium. In other words, unbiasedness imposes a stricter set of restrictions that are similar to the efficiency condition, with the difference that the coefficient on the futures price in the cointegrating equation is now equal to unity:

$$\delta = 1 \quad \delta\phi_1 = \mu = 1 \quad \phi_{1+j} = \gamma_l = 0 \quad \forall j > 0, l \geq 0 \quad (12)$$

We will follow Beck (1994) and test both sets of restrictions, for efficiency (10) and unbiasedness (12) using the error-correction specification (9), but in contrast to her analysis, we use overlapping weekly observations, a wider set of commodities, and longer horizons.

For those commodities for which spot and futures prices are I(1), we assessed the first notion of rationality by conducting cointegration tests. We applied the Engle-Granger procedure to (6) in which the realized spot price is the dependent variable:

$$s_{t+k} = -\frac{\rho k}{\beta} + \frac{1}{\beta} f_{t,t+k} + v_{t+k} \quad (13)$$

$$\Delta \hat{v}_t = \phi \hat{v}_{t-1} + \sum_{j=1}^k a_j \Delta \hat{v}_{t-j}$$

The lag length for each test equation of the residuals was determined by Bayesian information criteria. Given our use of overlapping observations, we control for residual correlation by using Newey-West HAC standard errors with a bandwidth equal to the futures contract horizon (i.e. days to maturity) minus one.

The lag specification  $K$  and  $M$  for each commodity at each forecast horizon  $k$  in (9) is decided on the basis of Bayesian information criterion. Again, given our use of overlapping observations, we control for residual correlation by using Newey-West HAC standard errors. We assessed efficiency and unbiasedness by applying the Wald test and using the robust estimate of the covariance matrix

To test for in-sample efficiency, for commodities for which both realized spot and futures prices were both I(0) variables, we used an equation in log levels rather than an ECM:

$$s_{t+k} = \alpha + \mu f_{t,t+k} + \sum_{j=1}^K \phi_j f_{t-kj,t-k(j-1)} + \sum_{l=0}^M \gamma_l s_{t-lk} + v_{t+k} \quad (14)$$

Lag length and robust standard error procedures were the same as described for the ECM approach. In this case, the restrictions implied by efficiency are that  $\phi_j = \gamma_l = 0$  for  $j \geq 1$  and  $l \geq 0$ , with the additional restriction of unbiasedness  $\mu = 1$ .

### 2.3. Out-of-sample forecasting ability test specifications

In-sample tests tell us something about the “rationality” of futures markets, but they do not indicate their usefulness as forecasters of spot prices. In this section, we describe tests of out-of-sample performance of futures prices as forecasters against a random walk benchmark.

The appropriate metric to assess forecasting ability will depend upon the forecaster’s loss function. For policymaking purposes, this function can and often should be asymmetric. A commodity exporting country would likely be much more concerned about lower-than-expected prices than an upside surprise. This could be because a downside surprise would lead to a shortfall in commodity-related government revenues and a need to increase public borrowing or reduce spending. In contrast, revenues from an upside surprise could be saved. Similar, but opposite, considerations would hold for commodity importers. In practice, it is difficult to specify a loss function and conventional practice has been to assume a symmetric loss function. The optimal forecaster in this case will be that which minimizes the mean squared forecast error (MSFE) and this is the approach we use here.

We compare the forecasting power of futures prices and other candidate models to the random walk without drift benchmark. The use of the random walk “no change” benchmark of forecast accuracy has a long history, stretching at least as far back as Theil (1966). What should be our prior be for the relative performance of futures contracts against spot markets? If we assume that current spot and futures prices incorporate the same information that is relevant for forecasting future spot prices, then futures prices should be at least as accurate as spot prices, on average. To see this, consider the cost-of-carry relationship for a financial asset, ignoring coupon or dividend payments, in a market without frictions. This states that the log price of a futures contract at time  $t$  which specifies delivery at  $t+k$  denoted by  $f_{t,t+k}$  is equal to the current log spot price  $s_t$  plus the continuously compounded constant interest rate  $r$  for the period  $t+k$ :

$$f_{t,t+k} = r_{t+k} + s_t \quad (15)$$

This relationship tends not to hold for commodities for two reasons. Spot and futures prices must take into account the costs of holding physical inventory, e.g., warehousing and insurance, which increases the cost of carry. Also, market participants may hold physical inventory of a commodity for its value as a consumption good, rather than as a financial asset. The benefit that accrues to the inventory holder is often referred to as the marginal convenience yield. We incorporate the physical storage costs, denoted by  $m$ , as a constant proportion of the spot price. The inclusion of the convenience yield for the marginal unit of inventory, denoted by  $\psi$ , leads to more profound changes. In particular, as the level of current and expected future inventories (denoted by  $N$ ) falls, the probability of experiencing a physical “stock out” increases, and  $\psi$  should rise, at an increasing rate as inventories fall towards their zero bound. (This nonlinearity is a key feature of important theoretical models such as Williams and Wright, 1991.) Incorporating these two features of commodity markets, storage costs and marginal convenience yields, into (15) obtains the arbitrage condition:

$$f_{t,t+k} = (r_{t,t+k} + m - \psi_{t,t+k}(N)) + s_t \quad (16)$$

Just as we can write the asset pricing equation (2) for futures markets, we can do the same for spot markets, assuming that the risk premium in both markets is the same:

$$s_t = -(r_{t,t+k} + m + \rho k - \psi_{t,t+k}(N)) + E_t^s s_{t+k} \quad (17)$$

In (17), investors holding a position in the spot market must be compensated for the physical cost of carry and the risk premium, less the marginal convenience yield. The forecast “errors” made by the spot price (denoted  $\varepsilon^s$ ) and futures price (denoted  $\varepsilon^f$ ), respectively, are:

$$\begin{aligned} \varepsilon_t^s &= s_t - s_{t+k} \\ \varepsilon_t^f &= f_{t,t+k} - s_{t+k} \end{aligned} \quad (18)$$

Using (2), (17) and (18), we can then write the difference in the squared forecast errors (denoted by  $d_t$ ) made by spot prices and futures prices as:

$$d_t = (\varepsilon_t^f)^2 - (\varepsilon_t^s)^2 = -(r_{t,t+k} + m - \psi_{t,t+k}(N))^2 + (E_t^f s_{t+k} - s_{t+k})^2 - (E_t^s s_{t+k} - s_{t+k})^2 \quad (19)$$

In most commodity pricing models with arbitrage, expectations in both spot and futures markets are the same and linked by the market for storage (e.g., Pindyck, 2001). In other words, market participants with different forecast would trade inventories until the difference in the spot and futures prices are explained by (16). This means the final two terms on the right hand side of (19) cancel out and that the squared forecast error of the futures price must always be less than or equal to the squared prediction error of the spot price; that is:

$$(\varepsilon_t^f)^2 \leq (\varepsilon_t^s)^2 \quad (20)$$

The actual value of (19) will clearly depend upon the quantity  $(r+m-\psi)^2$ . In turn, this should reflect the importance of current market conditions (rather than expectations for the future) in determining current spot prices. When current conditions exert a large influence on spot markets (for example, a period of temporary physical scarcity and low  $N$ ),  $\psi$  is high, the market is backwardated, and the futures price (which is less influenced by current market conditions) should provide, on average, a better forecast than the spot price. (We say on average because any particular observation may be affected by shocks that cause both spot and futures prices to rise or fall.) Conversely, if spot markets are driven by interest rate arbitrage with the forward-looking futures market (typically when inventories are plentiful),  $\psi$  must be low, and the forecasts of both spot and futures prices should be relatively close (assuming  $r$  and  $m$  are not too large). These assertions, particularly that  $\psi$  can be relatively large in backwardated markets, are confirmed by the data (Roache and Erbil, 2010).

It follows that a stronger test of (20) is one in which the loss function  $d$  is conditioned on the slope of the curve (and implicitly the value of  $\psi$ ). We estimate a regression in which the loss function  $d_t$  is the dependent variable and a constant and dummy variable are independent variables (with the dummy taking a value of 1 when the futures curve is backwardated and 0

otherwise, at the time the “forecast” is made). Given the quantity  $(r+m-\psi)^2$  is larger, on average, in backwardation, (20) predicts that the coefficient on the dummy variable is negative. In other words, the forecast ability of futures relative to a random walk benchmark should be better in backwardation. We also use our dataset to test the related assertion of Reeve and Vigfusson (2011) that what matters is the difference between the spot and futures price with a dummy variable that takes a value of 1 when the spot price is more than 5 percent higher than the futures price. By definition,  $\psi$  is very high during these periods and the spot market is largely driven by current conditions.

Our other candidate models include univariate ARIMA(1,1,1) and ARMA(1,1) with and without drift, depending upon the order of integration of the spot price. We also consider an exponential smoother, a futures price with risk premium model, a basis model, an error correction model, and a model in levels that includes lagged spot and futures prices (Table 1). The parameterized models allow us to test the forecasting performance of models that optimized in-sample fit.

Each model produces one-step ahead forecasts, hence, the notation  $t+1$  for the realized spot and error terms in Table 1. This means that the lags used in the time series models match the length of the forecast horizon. For example, for the 91 day spot price ARIMA (1,1,1) forecasting model, the lag of the change in the spot price would be for the previous 91 day period, and so on for 182, 364, and 728 days. The exceptions to this approach include the “weekly” ARIMA (labeled as W-ARIMA) and the Holt-Winters exponential smoothing models that produce  $k$ -step ahead forecasts, with  $t-j$  lags representing variable from the  $j$  weeks previous.

Table 1. Candidate and Benchmark Forecasting Models

Benchmark model	
Random walk	$s_{t+1} = s_t + \varepsilon_{t+1}$
Candidate models	
ARIMA (1,1,1)	$\Delta s_{t+1} = a_0 + a_1 \Delta s_t + a_2 \varepsilon_t + \varepsilon_{t+1}$
ARMA (1,1)	$s_{t+1} = b_0 + b_1 s_t + b_2 \varepsilon_t + \varepsilon_{t+1}$
W-ARIMA (1,1,1)	$\Delta s_{t+k} = a_0 + a_1 \Delta s_t + a_2 \varepsilon_t + \varepsilon_{t+k}$
Holt-Winters	$s_{t+k} = (a + bk) c_{t+k}$
Futures price	$s_{t+1} = f_{t,t+1} + \varepsilon_{t+1}$
Futures with risk premium	$s_{t+1} = \alpha + \beta f_{t,t+1} + \varepsilon_{t+1}$
Basis	$\Delta s_{t+1} = \alpha + \beta (f_{t,t+1} - s_t) + \varepsilon_{t+1}$
Error correction	$\Delta s_{t+1} = \alpha + \delta (\phi_0 + \phi_1 f_{t-1,t} - s_t) + \mu \Delta f_{t,t+1} + \phi_2 \Delta f_{t-1,t} + \gamma_0 \Delta s_t + \varepsilon_{t+1}$
Levels with lags	$s_{t+1} = \alpha + \beta f_{t,t+1} + \phi_0 f_{t-1,t} + \gamma_0 s_t + \varepsilon_{t+1}$

We test the null hypothesis that the squared forecast errors from the candidate models and the random walk benchmark are equal using the Diebold-Mariano (DM) procedure. When our candidate model is the futures price, a left-tailed rejection of the null would provide evidence in support of (20). This test statistic is:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\omega}/T}} \quad \text{where} \quad \bar{d} = T^{-1} \sum_{t=1}^T d_t \quad (21)$$

$\hat{\omega}$  is a consistent estimate of the long-run variance of  $T^{1/2} \bar{d}$  that takes account of the serial correlation introduced by overlapping observations. Diebold and Mariano (1995) show that under the null that both predictors are equally accurate, this statistic is asymptotically normally distributed with  $DM \sim N(0,1)$ . For models that nest the random walk and estimate additional parameters, we also report the results of hypothesis tests based on the adjusted MSFE of Clarke and West (2005). This test takes account of the noise introduced when, under the null hypothesis, parameters that are zero in the population affect the forecasts of models estimated from finite samples.

Our final test is not guided by any firm predictions from theory, although it is related to the “noise trader” perspective of commodity markets that has become increasingly prevalent (Vansteenkiste, 2011). The essence of this story is that financialization—defined as the influx of financial investors into commodity markets—has encouraged the participation of investors in futures markets that either have less information than traditional market participants or ignore fundamentals altogether. There is likely to be less of an effect on spot market activity because these investors often do not have the capacity to hold physical positions (which would involve storage and insurance). The participation of these types of investors is sometimes believed to increase during periods when prices are rising because they trade on price momentum; in other words, the more that prices increase, the greater their expectations of further price gains. In this scenario, we should expect that the forecasting performance of futures markets relative to spot markets should deteriorate during bull markets (i.e., when prices are rising) compared to bear markets (when prices are falling).

We test this hypothesis by estimating a regression in which the variable  $d$  is the dependent variable and a constant and a dummy representing bull markets are the independent variables. Bull and bear markets are identified using the Bry-Boschan algorithm that has been used previously to identify turning points in commodity price cycles (Cashin, McDermott, and Scott, 1999). The null hypothesis is that the relative forecasting ability of futures prices is the same in both bull and bear markets, or that the coefficient on the dummy variable is zero.

### 3. Data

#### 3.1. Overview of data

Our sample of spot and futures prices begins in January 1990, ends in June 2011, and is sampled at a weekly frequency. Although futures prices are available for some commodities before 1990,

this strikes a balance between a sufficiently long sample period and breadth of coverage. Prices for futures contracts are taken from Bloomberg. We use the set of the first 24 contracts, ordered by days to delivery. For those commodities with equally-spaced contracts with delivery dates for each month of the year, this implies that the futures curve stretches out for two years. In some cases, the number of traded contracts is less than 24, but the futures curve may stretch out further than two years; this is typically the case for agricultural commodities for which there is not a delivery date each month. In addition, there were insufficient data to undertake analysis at the two year horizon for aluminum and gasoline. Table 3 provides a detailed overview of the futures contracts used in the empirical analysis.

Table 3. Commodity Price Data: Specifications

Commodity	Spot price	Futures contract
Aluminum	London Metal Exchange-Aluminum 99.7% Cash	London Metal Exchange
Copper	London Metal Exchange-Copper, Grade A Cash	London Metal Exchange
Corn	Corn Number 2 Yellow	Chicago Mercantile Exchange
Cotton	Cotton, 1 1/16Str Low - Middling, Memphis	ICE: Intercontinental Exchange
Crude Oil	West Texas Intermediate Spot Cushing	Chicago Mercantile Exchange
Gasoline	Unleaded Regular Oxygenated New York	New York Mercantile Exchange
Gold	Gold Bullion London Bullion Market	Chicago Mercantile Exchange
Natural Gas	Natural Gas-Henry Hub	Chicago Mercantile Exchange
Soy	Soyabbeans, Number 1 Yellow	Chicago Mercantile Exchange
Wheat	Wheat, Number 2 Hard (Kansas)	Chicago Mercantile Exchange

### 3.2. Estimating fitted futures prices

One contribution of this paper is to provide a careful measure of the futures and spot price of each commodity, ensuring to the fullest extent possible that the horizon of the futures price matches that for the corresponding spot price change. Irregularly spaced futures contracts, even for those with delivery for each month of the year, make it difficult to ensure an exact match for high frequency observations. For example, there is no commodity for which there always exists a futures contract with a delivery date exactly 28 days forward that can be sampled at weekly intervals.

We overcome this constraint by estimating a smoothed futures curve for each week with a third order polynomial regression. This will provide a continuous curve from which we will always be able to extract an implied futures price at some exactly specified future date. In other words, at the end of each trading week, we estimate the following cross-sectional regression:

$$F_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 + v_t \quad (22)$$

where  $F$  is the vector of futures prices for a particular commodity at the close of Friday trading with increasing maturity,  $\gamma_0$  is a constant, and  $t$  represents the number of days until delivery for each contract. The constant and the coefficients  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$ , are the parameters to be estimated for each commodity each week. To obtain the implied futures price for a notional contract with delivery exactly  $t_1$  days forward, we use the estimated coefficients from this equation and

substitute in  $t_l$  for  $t$ . (In our case, we use  $t_l = 0, 91, 182, 364,$  and  $728$ .) One result from this approach is that the implied spot price ( $t=0$ ) equals  $\gamma_0$ . This is useful when contract specification differences make it problematic to compare futures to actual spot prices. Another useful result is that it provides an implied price for illiquid long-dated contracts based on the liquid part of the curve rather than non-tradable price quotes.

To assess the average fit of (22) we report average R-squared statistics for each commodity (Table 4). These results indicate that for most commodities, a third-order polynomial provides an acceptable fit. The average R-squared for all commodities stands at a respectable 0.8. In some cases, the fit is below 0.8 and this reflects one of two causes. First, for some commodities seasonal factors mean that the curve may at times have more than one local maximum or exhibit a number of inflection points. In almost all of these cases, (22) still does a relatively good job (e.g., corn and wheat). Second, quoted futures prices at the back end of 2-year curve may be quite illiquid and not fully representative of market conditions. This is true, for example, for gold for which (22) should prove to be a very good fit (given the interest-arbitrage determined curve). Our least-squares method by definition tends to put more weight on the liquid and smooth part of the curve where arbitrage considerations (and actual market conditions) are more important.

Table 4. Estimated Commodity Futures Price Curves: Measures of Fit  
(R-squared statistics) 1/

Aluminum	0.96
Copper	0.92
Corn	0.80
Cotton	0.89
Gasoline	0.75
Gold	0.72
Natural gas	0.54
Soybeans	0.72
Wheat	0.78
WTI crude oil	0.93

1/ Average R-square for all equations per commodity.

## 4. Empirical results and Discussion

### 4.1. In-sample rationality

We first examine the weakest notion of rationality—whether futures markets make persistent in-sample prediction errors. As the preliminary step before testing for cointegration, we found that most spot and futures prices were nonstationary. These results, shown in Table 5, are the result of augmented Dickey-Fuller unit root tests with the lag length selected by Bayesian information criteria. Given our use of overlapping observations, we control for residual correlation by using Newey-West HAC standard errors with a bandwidth equal to the futures contract horizon (i.e. days to maturity) minus one.

Table 5. Commodity Spot and Futures Prices: Unit Root Tests, Jan-1990 to Jun-2011  
(t-statistics for the null hypothesis that the variable is a unit root process)

Horizon	Aluminum	Copper	Corn	Cotton	Gasoline	Gold	Natural gas	Soybeans	Wheat	Crude oil
Spot prices										
3 months	-1.91	-0.61	-0.89	-2.55	-0.14	2.94	<b>-3.24</b>	-1.33	-1.73	-0.26
6 months	-2.04	-0.48	-1.63	<b>-2.99</b>	0.00	3.00	-2.65	-1.48	-1.87	-0.18
12 months	-0.34	-0.59	-2.15	<b>-3.63</b>	0.23	0.63	-2.21	-2.34	0.11	0.25
24 months	...	<b>-4.83</b>	-2.83	<b>-3.28</b>	...	-1.03	<b>-3.90</b>	-2.60	<b>-4.05</b>	-0.01
Futures prices										
3 months	-1.27	-0.58	-0.73	-2.46	0.09	2.89	-1.90	-1.15	-1.30	-0.06
6 months	-1.37	-0.34	-1.03	-1.78	0.12	2.79	-1.83	-0.95	-1.09	0.12
12 months	-1.31	-0.31	-0.69	<b>-4.00</b>	2.33	0.58	<b>-3.17</b>	-0.85	-0.68	0.30
24 months	...	<b>-4.63</b>	-1.12	-1.79	...	-1.59	<b>-3.72</b>	-2.16	-0.78	0.39

Source: Authors' estimates.

1/ Bold figures indicate that the null hypothesis of a unit root was rejected at the 5 percent level using Dickey-Fuller critical values and t-statistics based on robust standard errors.

For those commodities with I(1) spot and futures prices, we were able to reject the null hypothesis that the coefficient  $\phi = 0$  from equation (13) in all cases (Table 6). In other words, the commodity futures price and the realized spot price are cointegrated and futures markets do not make persistent forecasting errors.

Table 6. Hypothesis Tests of Cointegration Between Realized Spot and Futures Prices 1/  
(Engle-Granger test statistics)

Horizon	Aluminum	Copper	Corn	Cotton	Gasoline	Gold	Natural gas	Soybeans	Wheat	Crude oil
Spot prices										
3 months	<b>-8.78</b>	<b>-6.76</b>	<b>-9.99</b>	<b>-9.06</b>	<b>-7.36</b>	<b>-13.08</b>	...	<b>-7.01</b>	<b>-4.94</b>	<b>-4.99</b>
6 months	<b>-7.94</b>	<b>-7.55</b>	<b>-9.43</b>	...	<b>-8.83</b>	<b>-3.45</b>	<b>-5.60</b>	<b>-8.85</b>	<b>-3.11</b>	<b>-4.63</b>
12 months	<b>-4.37</b>	<b>-5.39</b>	<b>-6.85</b>	...	<b>-4.29</b>	<b>-3.82</b>	...	<b>-7.09</b>	<b>-4.59</b>	<b>-6.32</b>
24 months	...	...	<b>-3.07</b>	...	...	<b>-6.19</b>	...	<b>-3.29</b>	...	<b>-6.05</b>

Source: Authors' calculations.

1/ Based on Engle-Granger tests of the residuals from equation # with the number of lags selected by Bayesian information criteria. MacKinnon (1991) critical values.

Wald test  $p$ -value results for the null of efficiency (10) and the joint null of efficiency and unbiasedness (12) are shown in Table 7. The efficient market hypothesis was rejected for

most commodities and at most horizons. In particular, at the 91 and 182 day horizons, less than half of the commodities are efficient at the 5 percent significance level. We can also rule out efficiency for all commodities at the one and two year horizon except for cotton.

Table 7. Hypothesis Tests for the Efficiency and Unbiasedness of Futures Markets  
(null hypothesis that market is efficient and unbiased, p-values)

Commodity	Market Efficiency Null Hypothesis Horizon (days)				Unbiasedness Hypothesis Horizon (days)			
	91	182	364	728	91	182	364	728
Nonstationary I(1) spot and futures prices								
Aluminum	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	...	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	...
Copper	0.06	<b>0.00</b>	<b>0.00</b>	...	0.07	<b>0.00</b>	<b>0.00</b>	...
Corn	<b>0.04</b>	...	...	...	0.09	...	...	...
Cotton	...	...	...	...	...	...	...	...
Gasoline	<b>0.01</b>	0.10	<b>0.00</b>	...	<b>0.01</b>	0.08	<b>0.00</b>	...
Gold	0.93	<b>0.00</b>	<b>0.00</b>	...	0.97	<b>0.00</b>	<b>0.00</b>	...
Natural gas	...	0.47	<b>0.00</b>	...	...	0.63	<b>0.00</b>	...
Soybeans	0.57	<b>0.04</b>	...	...	0.57	<b>0.01</b>	...	...
Wheat	<b>0.00</b>	<b>0.00</b>	...	...	<b>0.00</b>	<b>0.00</b>	...	...
WTI crude oil	<b>0.00</b>	<b>0.01</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Stationary I(0) spot and futures prices								
Aluminum	...	...	...	...	...	...	...	...
Copper	...	...	...	<b>0.00</b>	...	...	...	<b>0.00</b>
Corn	...	...	...	...	...	...	...	...
Cotton	0.09	<b>0.04</b>	0.14	...	0.13	<b>0.02</b>	<b>0.00</b>	...
Gasoline	...	...	...	...	...	...	...	...
Gold	...	...	...	...	...	...	...	...
Natural gas	...	...	...	...	...	...	...	...
Soybeans	...	...	...	<b>0.00</b>	...	...	...	<b>0.00</b>
Wheat	...	...	...	...	...	...	...	...
WTI crude oil	...	...	...	...	...	...	...	...

Source: Authors' calculations.

Why was rationality, defined by efficiency and unbiasedness, so easy to reject? In most cases, lagged values of spot and futures prices provide significant in-sample explanatory power for realized spot prices. This finding was particularly true for the longer horizons of one and two years. Interpreting the coefficients on lags—which varied widely in sign and size across commodities—is difficult largely due to high degree of multicollinearity between the lags of realized spot prices and futures prices. One clear result, however, is that Bayesian criteria typically selected quite long lag lengths for the test equations, often of five or six periods, especially for horizons beyond 3 months. In other words, the evolution of spot prices seems to depend not only on lagged spot and futures prices, but on lags that stretch back some way into the past. The signs on these coefficients varied by commodity, which rules out a simple interpretation such as long-run mean reversion.

#### 4.2. Out-of-sample forecasting

Root mean square errors (RMSEs) for each of the candidate models and the random walk benchmark are presented in Table 8. One striking result is the large size of these errors for all of the approaches considered, confirming the extent to which commodity prices are volatile and difficult to forecast. For example, the RMSE for crude oil, copper, and corn futures—three of the most liquid contracts in the energy, metals, and food groups, respectively—range between 15 to 19 percent at the three month horizon. Absolute forecasting performance, in most cases, deteriorates with the length of the horizon; for example, for the same three contracts the RMSE range was 31 to 45 percent at the two year horizon.

The second notable result from the RMSEs is that the futures price and random walk models appear to outperform all of the other models, particularly for horizons from 91 to 364 days. The futures price and random walk RMSEs averaged across all 10 commodities is 16.1 percent and 16.4 percent, respectively at the 91 day horizon. In contrast, the RMSEs for the time series models are about 9 percentage points higher and for the models producing  $k$ -step ahead forecasts (including the exponential smoother and the weekly ARIMA) about 2 percentage points higher. These gaps only narrow significantly at the 2 year horizon. The one exception to this is the good performance of the Holt-Winters exponential smoother at the longer horizons, suggesting some trend persistence that is not picked up by other models.

In Figure 2, we show extended results for crude oil. This includes a larger range of models (specifically ARIMA and ARMA without drift) and extends the forecast horizon to 3 years (or 1092 days). The deterioration in the forecasting performance of futures the longer the forecast horizon is particularly clear from this analysis.

Figure 2. Root Mean Square Errors, 1990-2011  
(percent)

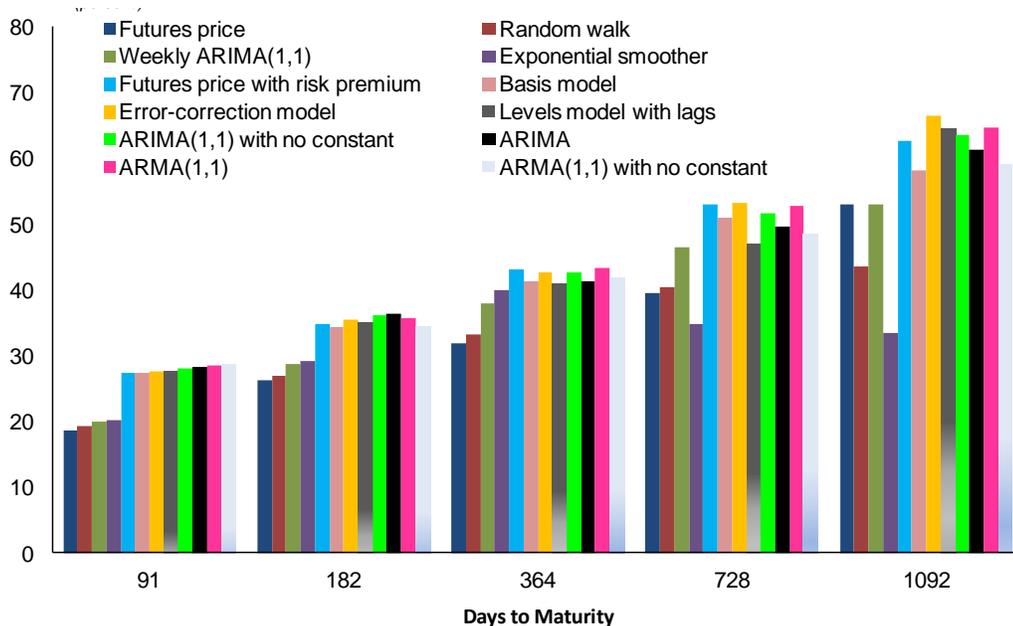


Table 8. Root Mean Square Errors, 1990-2011  
(percent)

	Horizon (days)				Horizon (days)			
	91	182	364	728	91	182	364	728
	Aluminum				Copper			
Random walk	11.2	17.5	23.6	...	15.5	23.3	31.7	44.5
ARIMA	21.1	26.8	37.3	...	25.3	33.2	48.3	61.0
ARMA(1,1)	21.1	27.3	35.1	...	25.6	34.4	47.9	57.3
Weekly ARIMA(1,1)	14.1	22.3	28.5	...	17.5	26.2	35.1	51.8
Exponential smoother	14.6	24.0	33.1	...	18.6	29.0	41.7	52.7
Futures price	11.6	18.2	23.7	...	15.3	23.3	32.5	45.4
Futures price with risk premium	21.6	28.1	37.7	...	25.3	34.5	49.4	59.6
Basis model	21.7	28.3	38.3	...	25.0	33.6	48.0	58.7
Error-correction model	21.0	27.4	35.7	...	25.6	34.4	48.4	57.6
Levels model with lags	21.7	27.8	35.5	...	25.6	34.8	49.0	58.1
	Corn				Cotton			
Random walk	16.4	23.8	29.4	36.8	14.9	22.6	30.8	38.7
ARIMA	26.0	32.8	40.9	41.2	22.8	33.7	38.5	31.2
ARMA(1,1)	26.2	31.3	36.9	36.0	22.4	31.3	35.8	30.1
Weekly ARIMA(1,1)	19.0	26.6	31.9	46.6	15.7	23.7	34.5	65.3
Exponential smoother	20.0	29.9	42.1	58.5	15.8	24.0	35.3	37.7
Futures price	16.7	23.0	27.1	31.3	14.8	22.2	29.5	36.0
Futures price with risk premium	25.2	30.8	35.3	32.8	22.8	32.2	35.7	31.8
Basis model	24.2	29.2	35.5	33.2	22.5	32.0	36.6	31.6
Error-correction model	25.0	30.4	38.2	34.0	22.5	33.7	35.4	30.8
Levels model with lags	25.3	30.8	39.7	35.9	22.3	32.7	35.6	25.6
	Crude oil				Gasoline			
Random walk	19.3	26.9	33.2	40.3	20.9	28.3	32.4	...
ARIMA	28.1	36.3	41.3	49.7	29.4	35.2	42.2	...
ARMA(1,1)	28.6	35.7	43.2	52.8	30.1	34.0	43.1	...
Weekly ARIMA(1,1)	19.9	28.6	37.9	46.6	22.9	31.0	37.0	...
Exponential smoother	20.1	29.1	39.9	34.7	22.9	31.1	36.9	...
Futures price	18.6	26.2	31.9	39.5	19.6	25.8	37.0	...
Futures price with risk premium	27.4	34.2	41.4	51.0	27.7	32.7	43.7	...
Basis model	27.7	34.9	40.8	46.8	28.1	33.1	42.9	...
Error-correction model	27.4	34.8	43.2	53.0	29.7	34.5	43.6	...
Levels model with lags	27.6	35.4	42.6	53.3	29.2	33.0	42.9	...

Hypothesis tests of out-of-sample forecast accuracy are presented in Table 9. The null hypothesis in these tests is that there is no difference in forecasting accuracy between the candidate model and the benchmark. On the basis of DM tests, it was possible to reject the null and conclude that the random walk is a better forecaster than most of the naïve reduced form models for the majority of commodities in our sample. Taking into account parameter uncertainty and on the basis of Clarke and West adjusted mean squared errors, however, the tests were less conclusive. In most cases, it was not possible to reject the null and ARIMA specifications also appear to outperform the random walk, especially at longer horizons for some agricultural commodities.

Table 8 (cont). Root Mean Square Errors, 1990-2011

	Horizon (days)				Horizon (days)			
	91	182	364	728	91	182	364	728
	Gold				Natural gas			
Random walk	7.0	9.9	15.2	25.5	29.5	38.6	49.2	52.7
ARIMA	10.6	14.9	22.3	34.6	37.4	45.9	56.2	55.8
ARMA(1,1)	10.5	15.2	24.3	36.6	39.3	50.2	54.2	63.4
Weekly ARIMA(1,1)	7.5	10.8	16.4	28.6	28.8	40.6	54.3	65.2
Exponential smoother	7.7	11.2	17.1	16.8	28.8	40.5	54.5	55.2
Futures price	6.9	9.7	14.7	24.4	27.3	34.5	43.8	53.1
Futures price with risk premium	10.6	15.7	26.0	38.2	36.0	47.8	54.9	66.5
Basis model	10.5	15.8	26.2	38.4	36.0	46.7	51.2	59.3
Error-correction model	10.6	15.4	22.8	34.7	36.1	49.2	57.3	64.9
Levels model with lags	10.3	14.7	22.9	34.7	36.5	50.7	54.6	65.0
	Soybeans				Wheat			
Random walk	13.2	18.8	24.1	32.7	14.3	20.4	27.9	34.4
ARIMA	20.9	26.4	33.1	37.0	21.2	29.6	35.9	35.3
ARMA(1,1)	20.8	25.9	34.3	37.3	21.2	29.9	35.4	26.9
Weekly ARIMA(1,1)	15.2	21.5	27.6	42.2	15.5	22.0	31.2	65.1
Exponential smoother	15.2	21.5	27.6	25.3	15.4	21.8	31.0	50.0
Futures price	12.4	17.1	21.3	29.8	18.0	21.9	28.1	38.9
Futures price with risk premium	19.7	25.0	32.0	35.3	20.9	30.5	37.4	27.0
Basis model	19.3	24.4	32.5	35.4	20.6	29.8	36.6	30.5
Error-correction model	20.2	25.3	32.8	34.5	20.5	29.8	38.4	31.5
Levels model with lags	19.8	25.2	34.5	36.4	20.8	29.7	37.8	29.2

Clearer results were obtained for futures prices. In almost all cases, the futures price outperformed the random walk, although statistical significance (at the 90 percent and 95 percent levels) of this result was less common. Futures prices did better at shorter horizons and became less accurate relative to a random walk at the two-year horizon. This suggests that lower liquidity at the back end of the curve may be impeding the price discovery process and reducing their use as forecasters, particularly relative to the Holt-Winters smoother. Time series models, on the other hand, did much worse than the random walk.

To illustrate some of these results in more detail, we focus on crude oil. Futures prices forecast better than a random walk at all horizons up to 2 years, although this result is statistically significant (at the 10 percent level) only at the 91 day horizon. Notwithstanding high liquidity in crude oil markets, the back end of the futures curve underperforms just as it does with many other commodities. One striking result is that, based on the DM test, we can reject the null at the 2 year and 3 year horizon and conclude that that the exponential smoother is a better forecaster than the random walk beyond 2 years. It is beyond the scope of our paper to understand the characteristics of commodity price dynamics and the smoother that produces this result, but medium-term reversion to persistent trends in prices seems to be playing some role in this result.

Table 9. Relative Forecast Evaluation: Diebold-Mariano and Clarke-West Test Statistics

	Horizon (days)				Horizon (days)			
	91	182	364	728	91	182	364	728
	Aluminum				Copper			
<b>Diebold-Mariano statistics</b>								
ARIMA (1,1,1)	2.42	1.58	3.01	0.80	3.78	3.18	2.26	1.16
ARMA (1,1)	2.74	1.86	1.19	0.04	3.96	3.06	1.87	0.84
Weekly ARIMA(1,1,1)	1.69	1.02	0.46	0.49	2.86	2.09	1.59	1.70
Exponential smoother	2.04	1.41	1.18	1.00	2.36	1.66	1.41	0.22
Futures price	0.17	0.00	0.19	-0.66	-1.17	-0.01	0.45	0.22
Futures price with risk premium	2.58	1.98	2.45	1.18	3.76	3.28	2.29	1.05
Basis model	2.54	1.88	2.96	0.96	3.73	2.92	2.07	1.00
Error correction model	2.55	1.96	1.77	0.54	3.94	3.20	2.47	0.95
Levels model with lags	2.81	2.02	1.94	0.68	4.00	3.08	2.06	0.97
<b>Clarke and West adjusted statistic</b>								
ARIMA (1,1,1)	1.34	-1.60	-0.41	-0.26	1.59	-0.77	0.04	0.24
ARMA (1,1)	1.56	-1.12	-0.59	-0.77	1.73	-0.38	-0.03	-0.18
Futures price with risk premium	1.66	-0.96	-0.31	-0.06	1.47	-0.37	0.26	0.09
Basis model	1.64	-0.99	-0.26	-0.30	1.42	-0.63	0.00	-0.01
Error correction model	1.37	-1.31	-0.67	-0.62	1.67	-0.37	0.09	-0.16
Levels model with lags	1.74	-1.06	-0.66	-0.53	1.73	-0.23	0.17	-0.09
	Corn				Cotton			
<b>Diebold-Mariano statistics</b>								
ARIMA (1,1,1)	4.48	2.58	1.95	0.25	4.83	3.39	1.00	-0.47
ARMA (1,1)	4.43	1.75	1.18	-1.13	3.84	2.71	0.69	-0.52
Weekly ARIMA(1,1,1)	4.52	3.19	2.42	2.30	4.55	4.41	3.69	1.92
Exponential smoother	3.39	2.06	2.07	-0.30	3.89	3.79	2.73	-2.48
Futures price	0.27	-0.50	-0.83	-1.11	-0.13	-0.25	-0.57	-0.80
Futures price with risk premium	4.24	1.88	0.91	-1.80	3.90	3.14	0.71	-0.29
Basis model	3.85	1.30	0.92	-1.60	3.61	3.13	0.78	-0.34
Error correction model	4.63	1.74	1.60	-1.42	4.04	3.27	0.65	-0.46
Levels model with lags	4.38	1.98	1.80	-0.95	3.75	3.34	0.63	-1.48
<b>Clarke and West adjusted statistic</b>								
ARIMA (1,1,1)	0.91	-0.73	-0.09	-1.49	1.96	1.12	-0.72	-1.37
ARMA (1,1)	0.92	-1.13	-1.08	-3.28	1.27	0.24	-1.62	-1.32
Futures price with risk premium	0.34	-1.55	-1.71	-3.45	1.61	0.67	-1.65	-1.13
Basis model	-0.52	-2.25	-1.59	-3.22	1.31	0.60	-1.23	-1.18
Error correction model	0.21	-1.74	-0.84	-3.05	1.43	1.22	-1.78	-1.28
Levels model with lags	0.41	-1.68	-0.39	-2.58	1.15	0.95	-1.72	-2.04

Table 9 (cont). Relative Forecast Evaluation: Diebold-Mariano and Clarke-West Test Statistics

	Horizon (days)				Horizon (days)			
	91	182	364	728	91	182	364	728
	Crude oil				Gasoline			
<b>Diebold-Mariano statistics</b>								
Arima (1,1,1)	4.24	3.76	2.58	2.15	3.44	1.40	0.72	-0.14
Arma (1,1)	4.06	3.58	3.06	2.11	4.16	2.07	0.39	1.23
Weekly ARIMA(1,1,1)	1.94	-0.16	-0.73	0.59	2.18	1.53	1.37	1.55
Exponential smoother	2.25	1.08	0.51	-2.42	2.65	1.73	1.59	-3.41
Futures price	-0.66	-0.72	-0.70	-0.61	-1.55	-2.22	-2.16	0.02
Futures price with risk premium	4.30	3.94	3.33	2.19	3.11	1.83	0.49	2.08
Basis model	4.17	3.70	3.62	2.64	3.06	1.44	-0.08	0.34
Error correction model	4.32	3.86	2.81	1.82	3.16	2.15	0.97	1.84
Levels model with lags	3.87	3.02	2.47	1.66	3.22	2.38	0.50	1.75
<b>Clarke and West adjusted statistic</b>								
Arima (1,1,1)	0.70	0.05	-0.01	0.95	-0.27	-1.69	-1.86	-2.52
Arma (1,1)	0.43	0.40	1.29	1.47	0.57	-0.58	-2.20	-1.39
Futures price with risk premium	0.65	0.95	1.86	1.66	-0.88	-1.28	-1.80	-0.70
Basis model	0.55	1.07	2.98	2.49	-0.91	-1.57	-2.74	-1.60
Error correction model	0.68	0.70	0.38	1.00	-0.80	-0.98	-1.66	-1.06
Levels model with lags	0.15	-0.15	0.41	0.92	-0.62	-0.52	-2.04	-0.95
	Gold				Natural gas			
<b>Diebold-Mariano statistics</b>								
Arima (1,1,1)	4.24	3.76	2.58	2.15	3.44	1.40	0.72	-0.14
Arma (1,1)	4.06	3.58	3.06	2.11	4.16	2.07	0.39	1.23
Weekly ARIMA(1,1,1)	1.94	-0.16	-0.73	0.59	2.18	1.53	1.37	1.55
Exponential smoother	2.25	1.08	0.51	-2.42	2.65	1.73	1.59	-3.41
Futures price	-0.66	-0.72	-0.70	-0.61	-1.55	-2.22	-2.16	0.02
Futures price with risk premium	4.30	3.94	3.33	2.19	3.11	1.83	0.49	2.08
Basis model	4.17	3.70	3.62	2.64	3.06	1.44	-0.08	0.34
Error correction model	4.32	3.86	2.81	1.82	3.16	2.15	0.97	1.84
Levels model with lags	3.87	3.02	2.47	1.66	3.22	2.38	0.50	1.75
<b>Clarke and West adjusted statistic</b>								
Arima (1,1,1)	0.70	0.05	-0.01	0.95	-0.27	-1.69	-1.86	-2.52
Arma (1,1)	0.43	0.40	1.29	1.47	0.57	-0.58	-2.20	-1.39
Futures price with risk premium	0.65	0.95	1.86	1.66	-0.88	-1.28	-1.80	-0.70
Basis model	0.55	1.07	2.98	2.49	-0.91	-1.57	-2.74	-1.60
Error correction model	0.68	0.70	0.38	1.00	-0.80	-0.98	-1.66	-1.06
Levels model with lags	0.15	-0.15	0.41	0.92	-0.62	-0.52	-2.04	-0.95

Table 9 (cont). Relative Forecast Evaluation: Diebold-Mariano and Clarke-West Test Statistics

	Horizon (days)				Horizon (days)			
	91	182	364	728	91	182	364	728
	Soybeans				Wheat			
<b>Diebold-Mariano statistics</b>								
Arima (1,1,1)	4.07	2.09	2.02	-0.05	4.61	3.48	1.46	1.41
Arma (1,1)	4.20	1.74	2.40	0.04	4.12	3.03	1.53	-0.34
Weekly ARIMA(1,1,1)	2.94	2.95	3.17	2.63	4.13	3.44	2.67	2.50
Exponential smoother	2.80	2.94	3.03	-2.35	3.42	2.62	1.80	1.95
Futures price	-1.70	-1.69	-1.64	-0.90	3.64	1.15	0.10	0.55
Futures price with risk premium	4.13	1.81	1.85	-0.75	4.49	3.85	1.74	-0.33
Basis model	3.69	1.90	1.98	-0.87	4.20	3.71	1.59	0.51
Error correction model	4.33	2.29	1.90	-1.18	4.29	3.76	1.70	0.79
Levels model with lags	4.09	2.15	2.31	-0.34	4.23	3.45	2.15	0.21
<b>Clarke and West adjusted statistic</b>								
Arima (1,1,1)	1.02	-0.55	-0.60	-2.08	0.31	0.12	-1.11	-0.13
Arma (1,1)	0.85	-0.67	-0.27	-2.04	0.23	0.19	-1.25	-2.15
Futures price with risk premium	-0.07	-1.12	-0.96	-2.89	-0.05	0.61	-0.73	-2.19
Basis model	-0.57	-1.65	-0.79	-2.99	-0.36	0.25	-0.95	-1.34
Error correction model	0.45	-1.22	-0.68	-3.28	-0.39	0.24	-0.50	-1.19
Levels model with lags	0.01	-1.22	-0.18	-2.51	-0.13	0.15	-0.77	-1.81

Table 10 shows the results from a conditional test of (20) that the futures price should be a better forecaster during periods when the market is backwardated (as discussed in section 2.2). The figures are the coefficients on a dummy variable which takes a value of 1 when the market is backwardated in a regression in which the dependent variable is the loss differential  $d$ . (19) predicts that this coefficient will be negative. In most cases, we cannot reject the null hypothesis that both prices are equally good forecasters in both contango and backwardation. In some cases, notably crude oil, we find that futures prices are worse forecasters in these markets. These results have important implications for forecasters. In particular, they caution against disregarding spot prices when making forecasts in a backwardated (and tight) market.

We also tested the relative performance of futures when the market is in significant backwardation, which we define here as a spot price more than 5 percent higher than the corresponding futures price. (We use 2 percent in the case of corn due to insufficient observations at the higher threshold. We also had to drop aluminum and gold from this analysis as there are very few occurrences of such extreme backwardation in these markets.) The results shown in Table 10 show that futures are somewhat better forecasters relative to a random walk in extreme backwardation, but this is not a consistent result across commodities or forecast horizons, in contrast to the findings of Reeve and Vigfusson (2011).

Table 10. Relative Forecasts Futures Prices During Contango and Backwardation 1/  
(coefficient on dummy variable=1 in backwardation)

	Horizon (days)				Horizon (days)			
	91	182	364	728	91	182	364	728
Backwardation test (dummy variable = 1 if spot price > futures price)								
	Aluminum				Copper			
Coefficient	-0.04	0.00	0.62	...	0.10	0.57	2.28	7.04
Standard error	(0.07)	(0.25)	(0.73)	...	(0.11)	(0.48)	(2.06)	(5.81)
	Corn				Cotton			
Coefficient	-4.04	-6.09	-8.94	1.33	<b>-1.02</b>	-1.71	-1.09	1.44
Standard error	(2.62)	(3.66)	(4.73)	(6.84)	(0.33)	(1.09)	(2.04)	(5.08)
	Crude oil				Gasoline			
Coefficient	0.42	<b>1.84</b>	<b>5.40</b>	<b>12.19</b>	<b>1.37</b>	<b>2.60</b>	<b>12.74</b>	...
Standard error	(0.32)	(0.92)	(2.22)	(6.06)	(0.58)	(1.11)	(4.58)	...
	Gold				Natural gas			
Coefficient	<b>0.04</b>	<b>0.18</b>	<b>0.44</b>	0.91	<b>-5.40</b>	<b>-8.31</b>	-6.46	12.26
Standard error	(0.02)	(0.05)	(0.21)	(0.96)	(2.64)	(3.54)	(5.81)	(12.94)
	Soybeans				Wheat			
Coefficient	-0.91	-1.20	-0.07	3.52	<b>2.25</b>	<b>2.41</b>	3.67	4.76
Standard error	(0.58)	(1.20)	(1.83)	(4.08)	(0.64)	(1.05)	(2.26)	(4.49)
Backwardation test (dummy variable = 1 if log(spot price/futures price) > 0.05)								
	Aluminum				Copper			
Coefficient	...	...	...	...	0.16	1.37	3.11	7.09
Standard error	...	...	...	...	(0.51)	(0.99)	(2.98)	(6.63)
	Corn				Cotton			
Coefficient	<b>-5.74</b>	<b>-9.62</b>	<b>-11.43</b>	0.98	<b>-1.98</b>	<b>-3.38</b>	-2.45	0.50
Standard error	(3.43)	(4.36)	(4.63)	(7.25)	(0.39)	(1.18)	(2.21)	(4.91)
	Crude oil				Gasoline			
Coefficient	-0.47	1.38	<b>5.01</b>	11.07	0.98	1.79	<b>13.09</b>	...
Standard error	(0.63)	(1.08)	(2.29)	(6.37)	(0.76)	(1.19)	(4.89)	...
	Gold				Natural gas			
Coefficient	...	...	...	...	<b>-10.55</b>	<b>-12.10</b>	-10.44	12.86
Standard error	...	...	...	...	(4.31)	(4.25)	(7.35)	(14.76)
	Soybeans				Wheat			
Coefficient	<b>-3.49</b>	-4.26	-1.25	3.21	<b>2.04</b>	<b>2.14</b>	2.85	4.40
Standard error	(1.69)	(2.55)	(3.24)	(4.46)	(0.51)	(0.94)	(1.96)	(4.90)

1/ Robust standard errors in parentheses. Figures in bold significant at the 95 percent level.

What can explain the results in Table 10? One possible reason is that commodity markets are segmented and that spot market participants make better forecasts than their future market counterparts (and offset the “error” caused by the cost of carry as in (19)). Segmentation is likely to be true, but it does not explain why market participants cannot arbitrage both markets; for example, spot market participants should sell spot and buy futures if they expect prices to rise in a backwardated market. Our inability to reject the null, particularly when conditioning on curve shape, is a puzzle when considered against conventional commodity pricing models.

Table 11 shows the results from the test of the null hypothesis that the forecasting performance of futures prices relative to a random walk does not depend on whether spot prices are in a bull or bear market (i.e., there is no difference in the loss function  $d$  in either case). Proponents of the view that financial speculation during bull markets leads to futures prices being driven away from fundamental levels should anticipate that futures prices do a worse job of forecasting during these periods. In Table 11, this would mean that the coefficients are positive (see section 2.2 for details).

We find very little evidence to support this view. In most cases, the relative forecasting ability of futures prices does not depend on the phase of the market (the exceptions are the 1-year crude oil and gasoline contracts and near-dated wheat contracts). One explanation for this result is that speculation affects both spot and futures markets and this is likely to be partially true. But at the same time, during a bull market, this would require a large rise in inventories as spot market participants anticipate persistent increases in prices. The evidence of much of the last decade—in which inventories declined as prices rose for many commodities—is inconsistent with this story. We conclude that the results in Table 11 undermine the view that index investing (which takes positions across a basket of commodities) distorts the price discovery process in futures markets.

Table 11. Relative Forecasts of Futures Prices During Bull and Bear Markets 1/  
(coefficient on dummy variable=1 in bull market)

91	Horizon (days)			91	Horizon (days)		
	182	364	728		182	364	728
	Aluminum			Copper			
<b>-0.29</b> (0.11)	-0.45 (0.24)	0.50 (0.53)	1.79 (1.61)	0.18 (0.10)	0.70 (0.40)	1.83 (1.66)	6.23 (4.91)
	Corn			Cotton			
<b>-1.81</b> (0.45)	<b>-3.49</b> (1.03)	<b>-6.21</b> (2.33)	-5.58 (4.51)	<b>-1.72</b> (0.35)	<b>-3.80</b> (1.04)	<b>-5.33</b> (1.90)	-6.23 (7.56)
	Crude oil			Gasoline			
-0.19 (0.25)	0.25 (0.64)	<b>3.37</b> (1.62)	6.08 (5.02)	0.65 (0.55)	0.98 (1.09)	<b>12.57</b> (4.01)	14.39 (8.20)
	Gold			Natural gas			
<b>-0.07</b> (0.02)	<b>-0.30</b> (0.06)	<b>-1.09</b> (0.27)	<b>-3.31</b> (1.19)	-2.09 (1.49)	<b>-5.33</b> (2.29)	<b>-8.55</b> (3.17)	-6.72 (11.89)
	Soybeans			Wheat			
-0.09 (0.28)	-0.40 (0.49)	-1.64 (1.15)	<b>-7.00</b> (2.96)	<b>2.14</b> (0.57)	<b>2.57</b> (0.82)	2.54 (1.82)	4.39 (3.97)

1/ Robust standard errors in parentheses. Figures in bold significant at the 95 percent level.

## 5. Conclusion

We arrive at four main conclusions regarding the forecasting performance of commodity futures prices in this paper. First, futures price-based forecasts are hard to beat. Futures prices perform at least as well as a random walk for most commodities and at most horizons and, in some cases, do significantly better. But the second result is that the failure of futures prices to clearly (and statistically significantly) outperform this benchmark in almost all cases is a puzzle. The spot price reflects the cost of carry and is more influenced by current physical market conditions (and less by expectations of the future) than is the futures price. In the absence of constraints on arbitrage, this should mean that futures prices outperform the random walk, on average. Third, many other naïve time series models, including some that maximize in-sample fit, tend to do much worse than a random walk. Parameter instability renders many time-series models as useless, at best. Fourth, the relative forecasting ability of futures prices deteriorates the longer the forecast horizon, which likely reflects lower liquidity at the back end of futures curves.

We also assessed the forecasting performance of futures prices relative to a random walk during different market conditions. Theory predicts that futures prices should do much better

than the random walk when the market is in backwardation because the influence of current market conditions on spot prices is particularly strong during these periods. However, we do not find a significant difference in forecasting ability between periods of backwardation and contango. This result holds even when spot prices are significantly above futures prices, in strong backwardation. What can explain this result? Over small sample periods it is possible that permanent shocks that increase prices could lead to better “forecasts” by spot prices. But over long periods and assuming a symmetric distribution of shocks, this cannot be the explanation. Segmented markets could also explain this result, with backwardated markets reflecting different and better information in the spot market about future spot prices than futures markets. But this would require strict and unrealistic limits to arbitrage; what would prevent spot market participants from simply buying cheaper futures contracts? We believe this apparent puzzle would benefit from further research.

We also do not find a significant difference in the forecasting ability of futures markets during bull and bear markets, defined as when spot prices are trending higher or lower. This new result suggests that the recent period of financialization has not distorted the futures price discovery process.

## References

- Beck, Stacie E., 1994, "Cointegration and Market Efficiency in Commodities Futures Markets," *Applied Economics*, Vol. 26, pp. 249-257.
- Cashin, Paul Anthony, C. John McDermott and Alasdair M. Scott, 1999, "Booms and Slumps in World Commodity Prices," *IMF Working Paper*, No. 99/155.
- Clements, Michael and David Hendry, 1998, *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.
- Chinn, Menzie D., and Coibion, Olivier, 2010, "The Predictive Content of Commodity Futures," *NBER working paper* 15830.
- Clarke, Todd E., and Kenneth D. West, 2007, "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, Vol. 138, Issue 1, pp. 291-311.
- Diebold, Francis X., and Roberto S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, Vol. 13, No. 3, pp. 253-263.
- Granger, C. W. J., and Paul Newbold, 1977, *Forecasting Economic Time Series*, Academic Press, New York.
- Pindyck, Robert S., 2001, "The Dynamics of Commodity Spot and Futures markets: A Primer," *Energy Journal*, Vol. 22, No. 3, pp. 1-29.
- Reeve, Trevor A. and Robert J. Vigfusson, 2011, "Evaluating the Forecasting Performance of Commodity Futures Prices," *International Finance Discussion Papers*, Number 1025.
- Roache, Shaun and Neşe Erbil, 2010, "How Commodity Price Curves and Inventories React to a Short-Run Scarcity Shock," *IMF Working Paper*, No. 10/222.
- Theil, H., 1966, *Applied Economic Forecasting*, North-Holland, Amsterdam.
- Vansteenkiste, Isabel, 2011, "What is Driving Oil Futures Prices? Fundamentals Versus Speculation," *ECB Working Paper*, No. 1371
- Williams, J. C., and B. D. Wright, B. D., 1991, *Storage and commodity markets*, Cambridge University Press,