WP/04/41

IMF Working Paper

Forecasting Commodity Prices: Futures Versus Judgment

Chakriya Bowman and Aasim M. Husain

INTERNATIONAL MONETARY FUND

© 2004 International Monetary Fund WP/04/41

IMF Working Paper

Research Department

Forecasting Commodity Prices: Futures Versus Judgment[1]

Prepared by Chakriya Bowman and Aasim M. Husain

March 2004

Abstract

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.

This paper assesses the performance of three types of commodity price forecasts—those based on judgment, those relying exclusively on historical price data, and those incorporating prices implied by commodity futures. For most of the 15 commodities in the sample, spot and futures prices appear to be nonstationary and to form a cointegrating relation. Spot prices tend to move toward futures prices over the long run, and error-correction models exploiting this feature produce more accurate forecasts. The analysis indicates that on the basis of statistical- and directional-accuracy measures, futures-based models yield better forecasts than historical-data-based models or judgment, especially at longer horizons.

JEL Classification Numbers: E37, G13, O13

Keywords: Commodity prices, futures, cointegration, error correction, forecast

Author's E-Mail Addresses: chakriya.bowman@absalom.com.au; ahusain@imf.org

---

[1] The authors would like to thank, without implicating, Paul Cashin, David Hallam, Sam Ouliaris, George Rapsomanikis, and participants at the Symposium on State of Research and Future Directions in Agricultural Commodity Markets and Trade, organized by the Food and Agriculture Organization, for extremely helpful comments on earlier drafts of the paper. The paper is to be published in a symposium proceedings volume.

Contents Page

I. Introduction

Although the share of primary commodities in global output and trade has declined over the past century, fluctuations in commodity prices continue to affect global economic activity. For many countries, especially developing countries, primary commodities remain an important source of export earnings, and commodity price movements have a major impact on overall macroeconomic performance. Hence, commodity-price forecasts are a key input to macroeconomic policy planning and formulation.

Forecasting commodity prices with reasonable accuracy is complicated by their considerable variability. Even the long-run trend behavior of commodities prices has generated debate, as typified by the important work of Cuddington (1992), who found little evidence to support the widely held Prebisch-Singer view that prices of primary commodities were on a declining path over the long term. More recently, however, Cashin and McDermott (2002) find some support for small and variable long-run downward trends in commodity price data, although they also find that such trends are swamped by the consistently high volatility of commodity prices.

This paper aims to assess the accuracy of alternative price forecasts for 15 primary commodities over the past decade. In view of the difficulties in accurately forecasting future price movements, the assessment of forecast performance has to be a relative one—measured by how certain types of forecasts perform in relation to others. For this purpose, three types of forecasts are considered: (i) judgmental forecasts, or those based on quantitative and qualitative analysis of a variety of factors—including, possibly, analysis of supply and demand fundamentals—thought to determine the price of the commodity in question; (ii) forecasts based on statistical models relying exclusively on historical price information; and (iii) forecasts based on models that purport to systematically incorporate all available information—as captured by commodity futures prices—at the time of the forecast, together with historical price data. A number of alternate measures of forecast performance, having to do with statistical as well as directional accuracy, are employed.[2]

The analysis indicates that although judgmental forecasts tend to outperform the model-based forecasts over short horizons of one quarter for several commodities, models incorporating futures prices generally yield superior forecasts over horizons of one year or longer. Spot and futures prices were generally found to be nonstationary and, in most cases, spot and futures prices appear to be cointegrated. Although there is considerable comovement between spot and futures prices, futures prices tend to exhibit less variability than spot prices. Hence, futures prices tend to act as an anchor for spot prices, and error-

[2] The ability of a forecasting methodology to predict adverse movements is perhaps a more relevant measure of accuracy in the context of commodity forecasts. Granger and Pesaran (1999) note that the literature on forecast evaluation has been biased toward statistical accuracy measures while neglecting measures of the economic importance of the forecasts.

correction models that exploit the long-run cointegrating relationship provide better forecasts of future spot-price developments.

The remainder of this paper is organized as follows. Section II summarizes key developments in the prices—spot and futures—of the commodities analyzed in the study. Section III presents tests of stationarity for spot and futures prices, as well as tests of cointegration between spot and futures prices where these are found to be nonstationary. Section IV describes the models that are used to generate forecasts, while Section V considers alternate measures of assessing forecast performance. The data are described in Section VI, and the results are summarized in Section VII. Section VIII concludes.

II. Commodity Price Developments: Some Facts

The analysis reported below covers 15 primary commodities which are part of the IMF's commodities price index and for which 3-month (or longer horizon) futures price data were available for the past decade. The commodities include six industrial metals (aluminum, copper, lead, nickel, tin, and zinc) as well as nine agricultural items (wheat, maize, soybeans, soybean meal, soybean oil, sugar, cotton, coffee—other milds, and coffee—robusta).

Prices of each of these commodities have declined considerably in real terms over the past three decades. Since 1970, the average quarterly change in the real price of each commodity has been negative (Table 1). On a cumulative basis, the real decline for coffee, copper, and tin has been especially large—about 70 percent or more—while sugar (U.S. market), wheat, and zinc prices have declined by 23–27 percent.

Futures prices tend to fluctuate in step with spot prices (Table 2), although the volatility of futures is markedly lower for virtually all commodities. Generally speaking, metals prices have tended to be less volatile than prices of agricultural commodities, possibly because agricultural commodities are more susceptible to weather-related shocks. Figures 1 and 2, which illustrate movements in cotton and copper prices, capture the relatively lower variability of futures prices.

Researchers have come to varying conclusions regarding the efficiency of commodity futures markets and whether futures prices are unbiased predictors of future spot prices. For example, Moosa and Al-Loughani (1994) find evidence of a risk premium in crude oil futures markets and conclude that futures prices are not efficient forecasters of future spot prices. On the other hand, Kumar (1992) presents evidence to support market efficiency and finds in favor of futures prices as unbiased forecasters of crude oil prices. Brenner and Kroner (1995) suggests that the inconsistencies observed between futures and spot prices may be as the result of carrying costs rather than a failing of the efficient market hypothesis, while Avsar and Goss (2001) observe that inefficiencies are likely to be exacerbated in relatively young and shallow futures markets such as the electricity market, where forecast errors may indicate a market still coming to terms with the true market model. Inefficiencies could also be exacerbated in markets with thin trading issues, or at time-to-maturity horizons that are relatively long, as market liquidity is also likely to affect risk premia (Kaminsky and Kumar, 1990b).

Rather than test for market efficiency directly, the objective here is to investigate simply whether futures prices can help predict developments in spot prices up to two years in the future. If spot and futures prices of a commodity are found to be nonstationary, and if there is evidence to suggest a cointegrating relationship between the two series, it would be expected that the addition of futures prices to a forecasting model will improve the performance of model forecasts. A related exercise was conducted by Kaminsky and Kumar (1990a), who looked into the power of futures prices to forecast future spot prices for seven commodities at horizons of up to nine months, although they did not exploit potential cointegrating relationships between spot and futures prices. Beck (1994), on the other hand, used cointegration techniques to test for market efficiency and the presence of risk premia in five commodity markets at the 8- and 24-week horizons. McKenzie and Holt (2002) employed cointegration and error correction models to test market efficiency and unbiasedness in four agricultural commodity markets, finding that for two of the four commodities in their sample, statistical model-based forecasts outperformed futures in a statistical sense.

Previous studies examining the performance of forecasts implied by futures prices versus those generated by models or "expert" opinion come to mixed conclusions about the performance of futures-based forecasts relative to judgmental or models-based forecasts. For example, Bessler and Brandt (1992) found that their expert opinion livestock forecaster performed significantly better in a statistical sense at the one-quarter horizon than the futures market for cattle but not for hogs, while Irwin, Gerlow, and Liu (1994) concluded that their expert opinion forecaster failed to perform significantly better than the futures market at the one- and two-quarter horizons, both for cattle and for hogs. It should be noted, however, that because of the time-restricted nature of futures contracts, futures prices have not been used to generate longer-term forecasts (1–5 years). Hence, the performance of such forecasts, especially in relation to judgmental forecasts, has not been consistently examined at the longer horizons for a reasonably wide set of commodities. Moreover, these studies do not assess directional performance—the ability to predict turning points—across different types of forecasts.

III. Stationarity and Cointegration

Commodity prices have generally been found to be nonstationary, although the precise nature of the trend—deterministic, stochastic, or containing structural breaks—is open to debate (Cashin, Liang and McDermott, 2000). The Prebisch-Singer hypothesis posits that there is a general downward trend in primary commodity prices, a thesis supported by many subsequent researchers[3]—with the important exception of Cuddington (1992)—who

[3] See, for example, Lutz (1999) and Cashin and McDermott (2002).

generally find a small but long-term negative deterministic trend in commodity price series,[4] and some cyclical movement.[5] This trend is typically augmented by long-lasting price shocks[6] and there is a significant degree of variability in the commodity prices that has increased over time.[7]

The overwhelming majority of commodity prices analyzed in this study were found to have nonstationary characteristics (Tables 3 and 4). The time series properties of commodity prices—spot and futures—were assessed by performing unit root tests. Rejection of the null hypothesis of a unit root under both the Augmented Dickey Fuller (ADF) test and the Phillips-Perron (PP) test was taken as evidence of stationarity. As the tables indicate, stationarity cannot be rejected only for soybeans, soybean meal, and soybean oil spot prices. Among futures prices, only tin, wheat, maize, and soybean prices appear to be stationary.

Most commodity prices appear to be cointegrated with at least their 3-month or 6-month futures price series. Results of cointegration testing using the Johansen test for cointegration are summarized in Table 5.[8] In the cases where no evidence is found for cointegration with any of the relevant futures price series (lead and coffee—other milds), this may be due to structural breaks in the series, which would result in this form of the Johansen test becoming biased against the rejection of the null hypothesis of no cointegration. As the results presented below indicate, error correction models tend to perform relatively well for virtually all commodities, suggesting that spot and futures prices are of the same order of integration and are cointegrated.

IV. FORECASTING MODELS

The simplest form of a forecasting model is the unit root model with trend and drift, which may be written as:

$$S_t = \alpha + \beta S_{t-1} + \gamma T + e_t, \tag{1}$$

[4] See, for example, Helg (1991), León and Soto (1997), Cashin and McDermott (2002).

[5] See, for example, Cashin and McDermott (2002).

[6] See, for example, Helg (1991) , Cuddington (1992), León and Soto (1997), and Cashin, Liang and McDermott (2000)

[7] See Cashin and McDermott (2002).

[8] For each commodity, the appropriate lag length was determined by minimizing the Akaike information criteria for each set of spot and futures prices, with a maximum of 6 lags tested.

where S_t is the natural logarithm of the commodity spot price at time *t* and T is a time trend variable. The error term, e_t , is assumed to be white noise. If the commodity price series contains a unit root, then a difference stationary model (or cointegration) should be used to model prices, otherwise the basic trend stationary model is appropriate. This simple model can serve as a useful benchmark for comparison with other, more sophisticated models.

An alternative forecasting model could be one that allows for an autoregressive process in the first difference of S_t and a moving average model for the errors. A suitable time series model of this form, the ARMA model, may be written as:

$$\Delta S_t = \alpha + \sum_{j=1}^{p} \beta_j \Delta S_{t-j} + e_t, \tag{2a}$$

with errors given by

$$e_t = \sum_{t=1}^{q} \gamma_i u_{t-i} + u_t, \tag{2b}$$

and where u_t is white noise. Such a model may be particularly appropriate for commodities where prices are mean reverting (see Irwin, Zulauf, and Jackson, 1996, for a discussion).

If markets are efficient, futures prices should be unbiased predictors of future spot prices and a simple prediction model should give superior results to those using alternative variables. The general futures forecast model is:

$$S_t = \alpha + \beta F_{t|t-k} + e_t, \tag{3}$$

where $F_{t|t-i}$ is the price for period t implied by futures markets in period t-k. Rather than testing market efficiency, which would imply α=0 and β=1, the aim here is to examine whether futures prices can enhance the forecasting ability of simple models.[9] To that end, futures prices can be added to the unit root model and ARMA specifications in an effort to obtain more accurate forecasts.

Finally, if commodity spot and futures prices are cointegrated, an error-correction model (ECM) can be used to capitalize on this relationship. Engle and Granger (1987) show that a system of two cointegrated series implies an error-correcting equation. Assuming that futures prices are weakly exogenous,[10] the general form of the ECM is:

[9] Efficiency tests would require careful matching of futures contract horizons and expiry dates with actual spot prices. As described below, the averaging of futures and spot price data in our dataset does not permit such tests with reasonable accuracy.

[10] This was verified during cointegration testing. Results are available on request.

$$\Delta S_t = \alpha + \beta_0 \varepsilon_{t-1} + \sum_{i=1}^{m} \beta_i \Delta F_{t-i|t-k} + \sum_{j=1}^{n} \gamma_j \Delta S_{t-j} + u_t, \tag{4}$$

where ε_t is the lagged residual of the cointegrating equation (i.e. equation 3). The ECM is used in this study as a contrast to the best forecast obtained from the simple unit root and ARMA models (with and without futures), as well as judgmental forecasts.

More complex models may, of course, be developed, such as that of Heaney (2002) which incorporates cost-of-carry into a forecasting model for lead prices and hence contains an interest rate component, or GARCH models (Morana, 2002) and probability-based forecast models (Abramson and Finizza, 1995). However, for the purposes of this study, where the objective is to gauge whether the incorporation of futures prices potentially yields superior forecast performance, forecasts use only historical spot prices and futures prices in an effort to identify simple models which may be successfully applied to a wide range of commodities, rather than to specific commodities.

V. ASSESSING FORECAST PERFORMANCE

When evaluating the *ex-post* effectiveness of forecasts, standard statistical measures are commonly used. Mean pricing error, mean absolute pricing error, mean absolute relative pricing error (*MARPE*), median absolute relative pricing error and root mean squared error (*RMSE*) are typically calculated and the results used to generate conclusions about the accuracy of forecasts.[11] This research will focus primarily on *RMSE*, which gives a measure of the magnitude of the average forecast error, as an effectiveness measure. It may be noted, however, the *RMSE* is a measure that is commodity specific, and cannot be readily used for comparison across commodities.

The *RMSE* may be defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(S_i - FC_i\right)^2} \qquad (5))$$

where S_i is the actual (spot) commodity price, and FC_i is the forecast price.

As the magnitude of the *RMSE* is specific to each price series, it can be difficult to quickly assess the performance of a model from this statistic. Hence in this application, the *RMSE* result is displayed relative to the *RMSE* of either the random walk model or the judgmental forecast, to facilitate comparison between models. The base model (the judgmental forecast) will have a value of unity. If a comparison model has a relative *RMSE*

---

[11] See for example Just and Rausser (1981), Leitch and Tanner (1991), Bessler and Brandt (1992), and Gerlow, Irwin, and Liu (1993).

value greater than unity, it may be considered to underperform the base model in terms of statistical accuracy. On the other hand, a relative *RMSE* value less than unity would indicate superior *RMSE* performance in relation to the base model.

Directional accuracy is also relevant to commodity forecasts, where the ability to identify future turning points is of particular importance. When assessing forecast performance, identification of directional changes may indeed be more important than the actual magnitude of error. Two methods are used to assess directional accuracy in this study. The first is the Harding and Pagan (2002) test of concordance, which seeks to identify synchronicity in the turning points of two series. The Harding–Pagan test is a statistical measure that casts no preference on the ability of the model to predict "important" changes as opposed to small but directionally accurate changes. This measure is augmented by the Cumby and Modest (1987) test, which weights the prediction of significant turning points more highly and hence is often used as a measure of the profitability of a prediction.

A rough measure of directional accuracy can be obtained by simply counting the number of times the forecast and actual prices move in the same direction. From this, a percentage of accurate directional forecasts may be calculated for each model. On average, a random walk model should pick the direction successfully around 50 percent of the time, and that more accurate forecast models should improve on this. Harding and Pagan (2002) extend this concept of directional accuracy, creating a measure of synchronicity that may be used to determine whether forecasts are "in synch" with actual price movements, or whether the confluence of prediction and reality is simply luck. This test is generated by creating two series, X_F for the forecasted (or futures) series and X_S for the actual spot price series:

$$
\begin{array}{lll}
 & X_{F,t} = 0 & \text{if } F_{t+n|t} - S_t < 0 \\
 & X_{F,t} = 1 & \text{if } F_{t+n|t} - S_t \geq 0 \\
\text{and} & X_{S,t} = 0 & \text{if } S_{t+n} - S_t < 0 \\
 & X_{S,t} = 1 & \text{if } S_{t+n} - S_t \geq 0,
\end{array}
$$

where F and S are the futures and spot price series, respectively, and n is the forecast horizon.

The Concordance statistic, for a given forecast horizon, is determined by:

$$C_{S,F} = T^{-1}\left[\sum_{t=1}^{T}(X_{S,t}X_{F,t}) + \sum_{t=1}^{T}(1-X_{S,t})(1-X_{F,t})\right]. \tag{6}$$

Hence, this statistic measures how closely—in directional terms—prices implied by futures move with actual spot prices. As noted above, forecasts from a random walk model would be expected, on average, to yield Concordance statistics of about 0.5. To obtain a sense of the statistical significance of the synchronicity between the series, a regression of the form

$$X_{S,t} = \alpha + \beta X_{F,t} + u_t \tag{7}$$

is run using Newey-West heteroskedastic autocorrelated consistent standard errors. If the series are not synchronous, the Harding-Pagan statistic (β) will equal zero. Hence, the estimated t-statistic for the β coefficient can be considered to yield a measure of the statistical significance of the synchronicity.

Another test of the directional performance of forecast models is the Cumby and Modest (1987) test for market timing ability, which is an extension of the Merton (1981) market timing test and was designed to use information about the magnitude of change as well as the direction of change to generate a performance statistic. The Cumby-Modest test is obtained from the estimated β coefficient from a regression of the form

$$\Delta S_t = \alpha + \beta X_t + e_t, \tag{8}$$

where S is the (natural logarithm of the) actual spot price and X is a dummy variable that takes the value of zero if the forecast anticipates a price decline for period t, and a value of unity if the forecast anticipates a price increase (or no change) for period t.[12] In essence, this differs from the Harding-Pagan statistic in that the dependent variable incorporates both the magnitude as well as the direction of the change. Hence, the Cumby-Modest statistic gives extra weight to situations under which the forecast would have correctly predicted the direction of large actual changes in spot prices, and when a forecast misses a directional change in prices that is small in magnitude, it is not penalized as heavily by the Cumby-Modest statistic as it is by the Harding-Pagan statistic.

VI. DATA

As noted above, the objective of this study is to compare the performance of three alternative types of commodity price forecasts—those based on judgment, those relying on statistical models using only historical price data, and those incorporating both futures prices as well as historical spot prices to yield statistical forecasts. Before turning to the assessment of the performance of the various forecasts, however, some explanation of how the forecasts were obtained and/or constructed is in order.

For the judgmental forecasts, commodity price projections prepared by the IMF, in collaboration with the World Bank, were used. These projections are prepared about once a quarter for each of the roughly 50 commodities in the IMF's primary commodity price index. The projections are for quarterly average prices, typically for the subsequent 5–8 quarters, and are available from 1993Q4. To the extent that judgmental forecasts incorporate information contained in futures prices, albeit not in a systematic fashion, they may be expected to be at least as accurate as futures-based forecasts.

[12] The estimates apply the White (1980) adjustment for heteroskedasticity.

The statistical forecasts were generated using the models described in equations (1) and (2), both with and without futures prices.[13] The estimated equations were used to generate forecasts as of each quarter for one-, four-, and eight-quarter horizons. Of the four statistical forecasts for each commodity, the best performing model in terms of statistical as well as directional accuracy was selected as the "best model" for comparison against the judgmental forecasts and the ECM forecasts.

For 8 of the 15 commodities, the best model at the one-quarter horizon incorporated futures.[14] For most of the metals (copper, lead, nickel, and tin), as well as wheat and cotton, this took the form of a unit root with futures prices as an additional exogenous model (i.e. equation 1 in first differences with an additional explanatory variable for futures prices), while for zinc and soybean oil the best model was an ARMA model with futures (i.e. equation (2) with an additional explanatory variable for futures prices). At the four-quarter horizon, the best model for 6 of the 15 commodities (tin, zinc, wheat, maize, soybean meal, and soybean oil) included futures, in most cases in a unit root model framework. At the eight-quarter horizon, the best model incorporated futures for 10 commodities (aluminum, copper, lead, nickel, tin, zinc, wheat, soybean oil, cotton, and coffee—robusta).

Similarly, the quarterly ECM forecasts were generated at the one-, four-, and eight-quarter horizons using estimated versions of equation (4). Figures 3 and 4, which illustrate the judgment and ECM forecasts generated in the third quarter of 1994 against actual price developments, indicate reasonable convergence between the forecasts at the one-quarter horizon for aluminum but not for coffee—other milds. By eight quarters, however, the opposite holds. Judgmental and ECM forecasts for coffee appear to converge while the forecasts for aluminum seem to move apart. The next section describes the extent to which both types of forecasts, as well as the best unit root/ARMA forecasts, deviate on average from actual price developments in directional and statistical accuracy terms.

Quarterly futures price series were constructed to facilitate comparability to the quarter average price projections in the judgmental forecasts. Monthly futures price quotes from Bloomberg are available for contracts with maturity dates near the end of the

[13] *ARMA(p,q)* models were generated from an iterative test of the combination of *p,q* that generated the lowest Schwarz criterion per Mills (1999), with $\max(p,q) \le 10$ for the standard series, and $\max(p,q) \le 6$ for the futures series (due to limits on the size of most futures series). This test was run for a model that fit the full range of data (ie. start date to 2003), and the parameters determined were then applied over the appropriate out-of-sample testing and forecasting windows.

[14] Information on which unit root/ARMA model performs best for each commodity at each horizon is contained in Tables 6–8.

subsequent 1-5 months for all 15 commodities in our sample. The one-quarter ahead price implied by futures was thus taken as the average of the prices prevailing at the end of each month in the current quarter of the contracts maturing in the next quarter. This procedure allowed the construction of one-quarter ahead prices for all 15 commodities. For wheat, maize, soybean oil, sugar, cotton, coffee—other milds, and coffee—robusta, we were able to also construct two-quarter ahead futures prices. Up to three-quarter ahead prices were constructed for soybeans, soybean meal, and soybean oil, while up to four-quarter ahead futures were constructed for copper.[15]

In terms of market depth, most of the futures contracts used are liquid, with open interest of over 100,000 contracts and with over 15,000 contracts normally traded on any given day. The exceptions to this are sugar (U.S.), coffee—robusta and cotton, which usually have less than 10,000 trades on any given day.[16] The London Metals Exchange gives monthly volume figures for the metals forwards, with aluminum and zinc being the most heavily traded metals (over one million trades per month) and tin being the least traded (around 100,000 trades per month).

VII. Results

The various directional and statistical accuracy measures tend to favor forecasts incorporating futures prices, particularly at the four- and eight-quarter horizons. At the shorter horizon of one quarter ahead, however, futures price based models performed at least as well as the judgment based models only for six of the fifteen commodities in the sample (Table 6). For nickel and zinc, the ECM outperforms the judgmental and the best unit root/ARMA forecasts in both directional and statistical terms. For soybean meal and cotton, the ECM forecast does at least as well as the other forecasts from either the statistical accuracy or the directional accuracy standpoint. For lead and soybean oil, the best unit root/ARMA forecast—in both cases based on models incorporating futures data—outperforms other forecasts in terms of directional accuracy. Judgmental forecasts for the remaining eight commodity prices, however, outperform the models-based forecasts—with or without futures prices—at the one-quarter horizon.

At the four-quarter horizon, judgmental forecasts outperform the models-based forecasts for only four of the fifteen commodities (Table 7). Among the remainder, the ECM

[15] Futures prices that most closely matched the forecast horizon (one, four, or eight quarters) were used in the econometric models for the model-based forecasts.

[16] The coffee commodities and tin are the least traded contracts (about 8–10,000 trades per day), while wheat futures are traded somewhat more (about 24,000 trades per day) and maize futures are a liquid market (about 62,000 trades per day). It is therefore not surprising that wheat and maize futures tend to be part of the “best” model, while coffee futures do not.

forecast does best for four commodities (aluminum, tin, zinc, and maize), while the best unit root/ARMA model does best for coffee—other milds. For the remaining six commodities (lead, nickel, soybean meal, soybean oil, sugar, coffee—other, and coffee—robusta), no single type of forecast consistently outperforms the others, although forecasts that incorporate futures—either in the ECM or in the best unit root/ARMA framework—do at least as well as other forecasts in five of these six cases.

The ECM forecasts outperform the other types of forecasts for eight of the fifteen commodities at the eight quarter horizon (Table 8). In some of these cases, the ECM forecast performance is superior in both statistical and directional terms (wheat, soybeans, and soybean meal), although for several commodities the ECM yields significantly better directional accuracy at the expense of somewhat lower statistical accuracy (aluminum, lead, nickel, zinc, and maize). For another four commodities (tin, soybean oil, sugar, and cotton), the ECM performs about as well as judgment at the eight-quarter horizon, and both perform better than the best unit root/ARMA forecasts. Of the remaining three commodities, the best unit root/ARMA framework yields the best forecasts. In the case of coffee—robusta, the model includes futures, while in the case of coffee—other it does not. Only for copper does judgment outperform the other forecasts by a sizable margin, although without significantly better directional accuracy. In sum, then, for thirteen of the fifteen commodity prices, models incorporating futures prices in either an error correction or unit root/ARMA framework produce forecasts that are at least as good as—and in most cases better than—forecasts that do not explicitly incorporate futures, including judgmental forecasts, at the eight-quarter horizon.

It may be noted that the model forecasts for coffee—robusta encounter problems due to a spike in futures prices. As a result, forecast prices increase rapidly and over longer horizons become very large. This contributes to the very low statistical accuracy of the forecasts. Hence, from a practical perspective, futures- and model-based forecasts may need to be "sanity checked" to ensure that short-term price panics do not create model forecasts that are unrealistic. Alternatively, discretionary inclusion of dummy variables in the estimated equations to adjust for such spikes may be appropriate in improving forecast accuracy.

VIII. Conclusion

The results suggest that futures prices can provide reasonable guidance about likely developments in spot prices over the longer term, at least in directional terms. For most of the commodities analyzed in this study, the incorporation of futures prices in an error-correction framework yields superior forecast performance at the two-year horizon. Since spot and futures prices are cointegrated for most commodities, and with futures prices exhibiting lower variability, longer-term spot price movements appear to be anchored by futures prices.

The generally superior performance of models with futures prices is somewhat surprising in light of the procedure employed to construct futures-price series that were comparable to those forecasted by the judgmental approach, particularly in view of the potential incorporation of futures-price information, albeit not systematically, in the judgmental forecasts. The averaging across various futures contracts and over various dates at which these contracts were priced may have resulted in a significant loss of information contained in the futures prices. Further research that more fully exploits this information by matching the dates of futures contracts with forecast horizons would clearly be desirable and may yield even stronger performance of futures-based models. Indeed, more careful date matching may well produce futures-based forecasts that more consistently outperform judgment at even the shorter horizons. The predictive capacity of the models may also be enhanced by incorporating variables capturing the demand for individual commodities, possibly via an economic-activity variable, and perhaps by pooling forecasts generated by various alternative statistical models or by employing more sophisticated time series techniques—such as ARCH, GARCH, or those incorporating structural breaks—to generate the models-based forecasts. These also remain on the agenda for future work.

Table 1. Select Commodities Spot and Futures Price Variability

Commodity	Real Price Decline Since 1970 (percent)	Standard Deviation[1] Spot	Standard Deviation[1] Futures	Futures Position	Start Period[2]
Aluminum	39.3	0.096	0.082	3-months forward	87: Q2
Coffee, other milds	68.0	0.164	0.140	6-months forward	87: Q1
Coffee, robusta	79.6	0.155	0.171	6-months forward	91: Q3
Copper	69.5	0.088	0.064	12-months forward	89: Q1
Cotton	56.0	0.112	0.071	6-months forward	86: Q3
Lead	58.9	0.095	0.091	3-months forward	87: Q1
Maize	53.1	0.106	0.078	3-months forward	72: Q2
Nickel	33.6	0.144	0.140	3-months forward	88: Q1
Soybean meal	52.8	0.085	0.066	9-months forward	82: Q4
Soybean oil	56.4	0.101	0.073	9-months forward	80: Q2
Soybeans[3]	49.5	0.063	0.054	9-months forward	75: Q1
Sugar, U.S.	23.0	0.036	0.023	6-months forward	88: Q1
Tin	69.5	0.070	0.070	3-months forward	89: Q3
Wheat	53.1	0.087	0.068	6-months forward	76: Q4
Zinc	27.3	0.093	0.080	3-months forward	89: Q1

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.

[1] Standard deviation of nominal dollar prices.

[2] All commodities series--with the exception of maize--end in 2003: Q1. The maize series ends in 2002: Q3.

[3] Outlier datapoint for 1994: Q3 removed.

Table 2. Correlation of Spot and Futures Prices, 1991: Q3–2003: Q1

(correlation of log first differences, in percent)

	Futures Horizon			
Commodity	3-month	6-month	9-month	12-month
Aluminum	95.44			
Coffee, other milds	93.22	91.70		
Coffee, robusta	94.18	93.20		
Copper	93.15			90.76
Cotton	62.73	74.07		
Lead	96.80			
Maize	79.70	81.80		
Nickel	94.97			
Soybean meal	76.78		69.96	
Soybean oil	74.54		61.07	
Soybeans[1]	85.75		83.24	
Sugar, U.S.	82.49	84.73		
Tin	93.48			
Wheat	70.82	79.23		
Zinc	93.03			

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.

[1] The reported correlation of nine-month soybeans futures with spot prices removes an outlier for 1994: Q1.

Table 3. Unit Root Tests: Spot Prices

Commodity	Augmented Dickey-Fuller[1] *t*-Statistic [*p*-value]	Sample Period [Lag Length]	Phillips-Perron[2] *t*-Statistic [*p*-value]	Sample Period [Bandwidth]
Aluminum	-2.9963 [0.14]	1970:1 – 2003:1 [1]	-2.8927 [0.17]	1970:1 – 2003:1 [4]
Copper	-2.5907 [0.29]	1970:1 – 2003:1 [0]	-2.9371 [0.15]	1970:1 – 2003:1 [3]
Lead	-2.9726 [0.14]	1970:1 – 2003:1 [1]	-2.5636 [0.30]	1970:1 – 2003:1 [2]
Nickel	-4.4299** [0.00]	1970:1 – 2003:1 [3]	-3.3012 [0.07]	1970:1 – 2003:1 [5]
Tin	-2.0282 [0.58]	1970:1 – 2003:1 [2]	-1.9105 [0.64]	1970:1 – 2003:1 [4]
Zinc	-3.5017* [0.04]	1970:1 – 2003:1 [1]	-2.9183 [0.16]	1970:1 – 2003:1 [6]
Wheat	-3.0235 [0.13]	1970:1 – 2003:1 [0]	-3.239 [0.08]	1970:1 – 2003:1 [6]
Maize	-3.6710* [0.03]	1970:1 – 2003:1 [1]	-3.2955 [0.07]	1970:1 – 2003:1 [1]
Soybean	-4.1025** [0.01]	1970:1 – 2003:1 [1]	-3.5506* [0.04]	1970:1 – 2003:1 [7]
Soybean meal	-4.5161** [0.00]	1970:1 – 2003:1 [1]	-3.7611* [0.02]	1970:1 – 2003:1 [5]
Soybean oil	-4.3439** [0.00]	1970:1 – 2003:1 [3]	-3.5153* [0.04]	1970:1 – 2003:1 [5]
Sugar (U.S.)	-5.0234** [0.00]	1970:1 – 2003:1 [1]	-3.2125 [0.09]	1970:1 – 2003:1 [10]
Cotton	-3.7266* [0.02]	1970:1 – 2003:1 [1]	-3.03 [0.13]	1970:1 – 2003:1 [1]
Coffee (other milds)	-3.2426 [0.08]	1970:1 – 2003:1 [1]	-2.5763 [0.29]	1970:1 – 2003:1 [0]
Coffee (robusta)	-3.2867 [0.07]	1970:1 – 2003:1 [3]	-2.5705 [0.29]	1970:1 – 2003:1 [2]

Sources: IMF Primary Commodity Prices Database; and authors' estimates.
Notes: ** indicates rejection of unit root hypothesis at 1 percent; * indicates rejection of unit root hypothesis at 5 percent. Rejection of the null hypothesis by both tests is regarded as evidence of stationarity. There is evidence here that soybean, soybean meal, and soybean oil are I(0), albeit not at the 1 percent significance level.

[1] The Augmented Dickey-Fuller statistic is used to test the null hypothesis of a unit root. Lag lengths were determined by minimizing the Schwarz information criteria.

[2] The Phillips-Perron statistic tests the null hypothesis of a unit root, and adjusts the standard Dickey-Fuller statistic for the presence of serial correlation using nonparametric procedures. Bartlett kernel estimation is used and bandwidth estimations made according to the Newey-West (1994) procedure.

Table 4. Unit Root Tests: Futures Prices

Commodity	Augmented Dickey-Fuller[1] t-Statistic [p-value]	Sample Period [Lag Length]	Phillips-Perron[2] t-Statistic [p-value]	Sample Period [Bandwidth]
Aluminum[3]	-3.02 [0.07]	1987:2 – 2003:1 [1]	-2.6 [0.14]	1987:2 – 2003:1 [3]
Copper[4]	-2.91 [0.08]	1989:1 – 2003:1 [1]	-2.53 [0.16]	1989:1 – 2003:1 [1]
Lead[3]	-2.66 [0.13]	1987:1 – 2003:1 [1]	-2.98 [0.07]	1987:1 – 2003:1 [2]
Nickel[3]	-3.89* [0.01]	1987:1 – 2003:1 [1]	-3.31 [0.04]	1987:1 – 2003:1 [3]
Tin[3]	-5.21** [0.00]	1989:2 – 2003:1 [0]	-4.90** [0.00]	1989:2 – 2003:1 [4]
Zinc[3]	-3.02 [0.07]	1989:1 – 2003:1 [1]	-3.34 [0.03]	1989:1 – 2003:1 [0]
Wheat[5]	-5.06** [0.00]	1972:1 – 2003:1 [4]	-4.028* [0.01]	1972:1 – 2003:1 [7]
Maize[5]	-4.18** [0.00]	1972:1 – 2003:1 [1]	-3.71* [0.01]	1972:1 – 2003:1 [0]
Soybean[5]	-3.68* [0.01]	1975:1 – 2003:1 [0]	-3.86* [0.01]	1975:1 – 2003:1 [4]
Soybean meal[5]	-3.87* [0.01]	1978:1 – 2003:1 [2]	-3.26 [0.04]	1978:1 – 2003:1 [1]
Soybean oil[5]	-2.83 [0.01]	1979:2 – 2003:1 [0]	-3.18 [0.05]	1979:2 – 2003:1 [3]
Sugar (US)[6]	-3.60* [0.02]	1988:1 – 2003:1 [1]	-2.62 [0.14]	1988:1 – 2003:1 [0]
Cotton[6]	-3.36 [0.03]	1986:2 – 2003:1 [1]	-2.99 [0.07]	1986:2 – 2003:1 [2]
Coffee (other milds)[6]	-2.82 [0.10]	1986:3 – 2003:1 [0]	-3.01 [0.07]	1986:3 – 2003:1 [2]
Coffee (robusta)[7]	-2.22 [0.23]	1991:3 – 2003:1 [1]	-1.97 [0.30]	1991:3 – 2003:1 [2]

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.
Notes: ** indicates rejection of unit root hypothesis at 1 percent; * indicates rejection of unit root hypothesis at 5 percent. Tests are for quarterly 3-month futures series. Rejection of the null hypothesis by both tests is regarded as evidence of stationarity. The evidence suggests that soybean, tin, wheat, and maize futures are *I(0)* .

[1] The Augmented Dickey-Fuller statistic is used to test the null hypothesis of a unit root. Lag lengths were determined by minimizing the Schwarz information criteria.

[2] The Phillips-Perron statistic tests the null hypothesis of a unit root, and adjusts the standard Dickey-Fuller statistic for the presence of serial correlation using nonparametric procedures. Bartlett kernel estimation is used and bandwidth estimations made according to the Newey-West (1994) procedure.

[3] Contract is listed on the London Metals Exchange.

[4] Contract is listed on the New York Mercantile Exchange (NYMEX) or Commodity Exchange Inc. (COMEX).

[5] Contract is listed on the Chicago Board of Trade.

[6] Contract is listed on the New York Board of Trade Coffee, Sugar and Cocoa Exchange or Cotton Exchange.

[7] Contract is listed on the London International Financial Futures and Options Exchange.

Table 5. Johansen Cointegration Test Results

Commodity	Sample Period	Test Statistic $k = 0$	$k <= 1$	Lags[4]
Aluminum 3-month[1]	1987:2 – 2003:1	19.04**	0.27	6
Copper 3-month[1]	1989:1 – 2003:1	19.26**	0.74	6
Copper 6-month[3]	1989:1 – 2003:1	28.73*	5.5	6
Copper 9-month[3]	1989:1 – 2003:1	21.36	8.08	1
Copper 12-month[3]	1989:1 – 2003:1	22.99	8.53	1
Lead 3-month[3]	1987:1 – 2003:1	18.76	7.97	2
Nickel 3-month[3]	1987:1 – 2003:1	58.14**	10.84	6
Tin 3-month[1]	1989:2 – 2003:1	14.62*	0.87	2
Zinc 3-month[3]	1989:1 – 2003:1	18.71	5.61	6
Wheat 3-month[2]	1972:1 – 2003:1	21.20*	6.63	6
Wheat 6-month[3]	1976:4 – 2003:1	16.59	5.27	1
Maize 3-month[3]	1972:1 – 2003:1	29.34*	10.45	3
Maize 6-month[3]	1972:1 – 2003:1	30.05*	11.4	2
Soybean 3-month[2]	1975:1 – 2003:1	23.26*	5.12	5
Soybean 6-month[2]	1975:1 – 2003:1	22.18*	4.54	3
Soybean 9-month[1]	1989:2 – 2003:1	15.64*	0.09	1
Soybean meal 3-month[2]	1978:1 – 2003:1	25.19**	8.57	1
Soybean meal 9-month[2]	1982:3 – 2003:1	14.25	3.97	6
Soybean oil 3-month[2]	1979:2 – 2003:1	26.16**	8.24	1
Soybean oil 6-month[2]	1979:4 – 2003:1	21.23*	7.31	1
Soybean oil 9-month[2]	1980:2 – 2003:1	22.91*	8.14	2
Sugar (US) 3-month[3]	1988:1 – 2003:1	20.36	8.01	2
Sugar (US) 6-month[3]	1988:1 – 2003:1	19.32	5.81	2
Cotton 3-month[3]	1986:2 – 2003:1	16.14	3.3	4
Cotton 6-month[1]	1986:3 – 2003:1	14.48*	0.09	5
Coffee (other milds) 3-month[3]	1986:3 – 2003:1	17.03	6.3	1
Coffee (other milds) 6-month[3]	1987:1 – 2003:1	15.89	6.53	1
Coffee (robusta) 3-month[1]	1991:3 – 2003:1	14.34*	0.48	1
Coffee (robusta) 6-month[1]	1991:3 – 2003:1	11.33	0.48	1

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.
Notes: ** indicates rejection of unit root hypothesis at 1 percent; * indicates rejection of unit root hypothesis at 5 percent. Evidence of cointegration between spot and futures prices was found for most commodities with 3-month futures and several with later-dated contracts. The exceptions were lead, zinc, sugar (U.S.), and coffee (other milds), for which no evidence of cointegration was found. This may be due to a variety of factors, including the presence of structural breaks in the series. Contracts with evidence of cointegration are highlighted in bold.

[1] Results found no deterministic trend in the data, and no intercept or trend in the cointegrating equation.

[2] Results found no deterministic trend in the data, and an intercept but no trend in the cointegrating equation.

[3] Results found a linear trend in the data, and both an intercept and a trend in the cointegrating equation.

[4] Lag length determined by minimizing the Akaike information criteria for a maximum of 6 lags. Weak exogeneity was confirmed using restriction testing.

Table 6. Forecast Performance at One-Quarter Horizon

	RMSE Ratio (RMSE actual value)			Harding Pagan Statistic (β) (*t*-statistic) Concordance Statistic			Cumby-Modest Statistic (*t*-statistic)		
Commodity	Judgment	Best UR/ ARMA	ECM	Judgment	Best UR/ ARMA	ECM	Judgment	Best UR/ ARMA	ECM
Aluminum [1]	1.00	1.31	1.15	0.36*	0.24	0.34*	0.07**	0.04*	0.07**
	(82.58)	(108.43)	(95.00)	(2.08)	(1.30)	(1.91)	(3.60)	(1.76)	(3.43)
				[0.68]	[0.62]	[0.68]			
Copper [2]	1.00	1.38	1.26	0.48**	0.30**	0.41**	0.09**	0.07**	0.07**
	(124.72)	(172.08)	(157.34)	(3.22)	(2.32)	(2.54)	(3.70)	(2.49)	(2.32)
				[0.73]	[0.65]	[0.70]			
Lead [2]	1.00	1.07	1.03	0.21	0.28*	0.19	0.03	0.02	0.03
	(36.78)	(39.24)	(37.91)	(1.37)	(2.02)	(1.10)	(1.16)	(0.75)	(1.19)
				[0.59]	[0.62]	[0.59]			
Nickel [2]	1.00	1.04	0.90	0.30	0.40**	0.47**	0.10**	0.09**	0.13**
	(670.36)	(697.62)	(603.99)	(1.62)	(2.27)	(3.05)	(2.60)	(2.29)	(4.41)
				[0.65]	[0.70]	[0.70]			
Tin [2]	1.00	3.31	3.28	0.89**	-0.36**	0.07	0.10**	-0.01	-0.00
	(99.66)	(329.64)	(327.22)	(11.87)	(-4.71)	(0.40)	(7.86)	(-0.76)	(-0.20)
				[0.95]	[0.62]	0.46			
Zinc [4]	1.00	1.32	1.24	0.19	0.34**	0.31*	0.07**	0.08**	0.08**
	(71.72)	(94.87)	(89.20)	(0.92)	(2.44)	(2.25)	(2.73)	(3.08)	(2.61)
				[0.59]	[0.68]	[0.62]			
Wheat [2]	1.00	1.54	1.48	0.42**	0.44**	0.42*	0.10**	0.07*	0.09**
	(11.38)	(17.47)	(16.80)	(2.51)	(3.94)	(2.13)	(3.13)	(2.14)	(2.51)
				[0.70]	[0.70]	[0.70]			
Maize [1]	1.00	1.26	1.17	0.44**	0.08	0.24*	0.09**	0.08*	0.06
	(11.11)	(13.95)	(13.00)	(2.44)	(0.61)	(1.92)	(2.35)	(1.92)	(1.48)
				[0.70]	[0.54]	[0.62]			
Soybean [3]	1.00	1.08	1.16	0.56 **	0.24*	0.29*	0.07**	0.02	0.05*
	(14.71)	(15.86)	(16.99)	(3.71)	(1.77)	(1.89)	(3.46)	(1.12)	(2.19)
				[0.78]	[0.62]	[0.65]			
Soybean meal [1]	1.00	1.00	1.01	0.32**	0.18	0.18	0.07**	0.01	0.09**
	(18.01)	(17.96)	(18.13)	(2.58)	(1.27)	(0.99)	(2.35)	(0.36)	(2.42)
				[0.65]	[0.59]	[0.59]			
Soybean oil [4]	1.00	1.13	1.01	0.13	0.32**	-0.03	0.04	0.01	0.05
	(44.51)	(50.10)	(45.17)	(0.86)	(2.64)	(-0.18)	(1.44)	(0.43)	(1.47)
				[0.57]	[0.68]	[0.51]			
Sugar (U.S.) [3]	1.00	1.41	1.39	0.51**	0.31*	0.26	0.04**	0.03*	0.03*
	(0.72)	(1.02)	(1.00)	(4.15)	(2.09)	(1.50)	(3.02)	(1.93)	(2.10)
				[0.76]	[0.62]	[0.62]			
Cotton [2]	1.00	0.97	0.96	0.45**	0.34**	0.29**	0.11**	0.11**	0.10**
	(5.81)	(5.64)	(5.59)	(3.35)	(2.52)	(2.58)	(3.42)	(3.75)	(3.12)
				[0.73]	[0.68]	[0.65]			
Coffee (other) [3]	1.00	1.43	1.38	0.22	0.20	0.11	0.10	0.07	0.09
	(18.26)	(26.16)	(25.15)	(1.17)	(1.34)	(0.57)	(1.36)	(1.14)	(1.35)
				[0.68]	[0.59]	[0.59]			
Coffee (robusta) [3]	1.00	2.11	4.04	0.73**	0.29*	0.19	0.20**	0.13**	0.08
	(7.97)	(16.82)	(32.20)	(5.48)	(1.71)	(1.05)	(3.86)	(2.49)	(1.64)
				[0.86]	[0.65]	[0.57]			

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.
Notes: Significant statistics in bold; best forecasts in italics; * and ** denote significance at 5 percent and 1 percent, respectively.

[1] Best UR/ARMA model was standard unit root model.

[2] Best UR/ARMA model was unit root model with lagged futures prices.

[3] Best UR/ARMA model was ARMA (p,q) model with (p,q) equal to (7,8) for soybeans; (2,5) for sugar; (1,1) for coffee--other milds; and (2,10) for coffee--robusta.

[4] Best UR/ARMA model was ARMA (p,q) model with futures, with (p,q) equal to (1,2) for zinc and (3,3) for soybean oil.

Table 7. Forecast Performance at Four-Quarter Horizon

	RMSE Ratio (RMSE actual value)			Harding Pagan Statistic (β) (t-statistic) Concordance Statistic			Cumby-Modest Statistic (t-statistic)		
Commodity	Judgment	Best UR/ ARMA	ECM	Judgment	Best UR/ ARMA	ECM	Judgment	Best UR/ ARMA	ECM
Aluminum [3]	__1.00__	__1.18__	___0.94___	__0.30*__	__0.50**__	___0.63**___	0.05	__0.14**__	___0.18**___
	(257.23)	(303.09)	_(241.22)_	(2.11)	(2.97)	_(5.04)_	(0.92)	(2.43)	_(4.22)_
				[0.61]	[0.76]	_[0.76]_			
Copper [3]	___1.00___	__1.08__	__1.18__	___0.34*___	-0.10	0.12	___0.14*___	-0.06	-0.01
	_(468.65)_	(508.36)	(554.36)	_(2.05)_	(-0.49)	(0.56)	_(2.29)_	(-0.90)	(-0.10)
				_[0.70]_	[0.43]	[0.54]			
Lead [3]	___1.00___	___1.14___	__1.02__	_0.15_	___0.29**___	-0.06	___0.11*___	_0.05_	0.01
	_(94.00)_	_(107.06)_	(95.84)	_(1.33)_	_(2.78)_	(-0.26)	_(2.03)_	_(0.99)_	(0.16)
				_[0.52]_	_[0.65]_	[0.43]			
Nickel [3]	__1.00__	___0.85___	___0.93___	__0.38*__	___0.45 **___	___0.52**___	0.05	___0.27**___	___0.44**___
	(2006.69)	_(1703.79)_	_(1858.68)_	(1.83)	_(2.41)_	_(2.85)_	(0.37)	_(2.90)_	_(7.24)_
				[0.73]	_[0.65]_	_[0.62]_			
Tin [4]	__1.00__	__1.13__	___1.16___	0.26	0.19	___0.35*___	-0.01	-0.00	_0.02_
	(585.76)	(660.10)	_(679.95)_	(1.35)	(0.91)	_(1.68)_	(-0.26)	(-0.09)	_(0.51)_
				[0.67]	[0.57]	_[0.65]_			
Zinc [2]	__1.00__	__1.15__	___0.98___	0.23	-0.21	___0.54**___	__0.16*__	__-0.16*__	___0.22**___
	(185.69)	(213.09)	_(181.40)_	(1.31)	(-0.88)	_(4.17)_	(1.69)	(-1.88)	_(3.48)_
				[0.55]	[0.54]	_[0.68]_			
Wheat [4]	___1.00___	__1.28__	__0.93__	_0.24_	__-0.35**__	0.21	___0.15*___	__-0.21**__	0.12
	_(31.69)_	(40.51)	(29.40)	_(1.20)_	(-2.42)	(0.96)	_(1.73)_	(-3.31)	(1.51)
				_[0.61]_	[0.35]	[0.59]			
Maize [4]	__1.00__	__1.06__	___0.93___	__0.30*__	0.09	___0.27*___	__0.21*__	0.05	___0.16*___
	(26.08)	(27.77)	_(24.14)_	(1.88)	(0.57)	_(1.78)_	(1.87)	(0.69)	_(1.79)_
				[0.58]	[0.57]	_[0.62]_			
Soybean [1]	___1.00___	__1.31__	__1.32__	___0.63**___	0.12	__0.36*__	___0.19**___	__-0.05*__	0.08
	_(29.36)_	(38.48)	(38.65)	_(5.05)_	(0.41)	(1.90)	_(4.14)_	(-1.75)	(1.44)
				_[0.82]_	[0.49]	[0.68]			
Soybean meal [4]	___1.00___	___1.03___	___1.01___	___0.51**___	___0.73**___	___0.53**___	___0.27**___	___0.33**___	___0.28**___
	_(38.09)_	_(39.16)_	_(38.47)_	_(3.29)_	_(6.96)_	_(3.21)_	_(2.78)_	_(4.72)_	_(3.30)_
				_[0.76]_	_[0.86]_	_[0.76]_			
Soybean oil [2]	___1.00___	__1.09__	___0.94___	___0.32*___	-0.26	_0.30_	___0.20**___	__-0.19**__	___0.16*___
	_(102.88)_	(112.06)	_(96.84)_	_(2.29)_	(-1.08)	_(1.13)_	_(2.61)_	(-2.69)	_(1.99)_
				_[0.67]_	[0.38]	_[0.68]_			
Sugar (U.S.) [3]	___1.00___	___1.14___	___0.89___	___0.38*___	___0.45**___	___0.34**___	_0.01_	___0.07*___	_0.05_
	_(1.69)_	_(1.93)_	_(1.50)_	_(2.24)_	_(3.83)_	_(2.29)_	_(0.17)_	_(1.92)_	_(1.61)_
				_[0.70]_	_[0.54]_	_[0.65]_			
Cotton [3]	___1.00___	__1.08__	__1.25__	___0.36*___	0.06	0.16	_0.10_	-0.04	-0.01
	_(13.49)_	(14.51)	(16.88)	_(2.72)_	(0.32)	(0.77)	_(1.64)_	(-0.52)	(-0.19)
				_[0.58]_	[0.51]	[0.59]			
Coffee (other) [1]	__1.00__	___1.21___	__0.96__	-0.22	___0.37**___	-0.12	-0.06	___0.22*___	0.02
	(49.14)	_(59.48)_	(47.33)	(-1.23)	_(2.39)_	(-0.58)	(-0.54)	_(2.02)_	(0.21)
				[0.45]	_[0.70]_	[0.49]			
Coffee (robusta) [3]	___1.00___	___1.43___	- [5]	_0.06_	___0.42**___	___0.38*___	_-0.03_	___0.33**___	- [6]
	_(31.52)_	_(44.99)_	-	_(0.27)_	_(3.12)_	_(1.86)_	_(-0.29)_	_(3.43)_	-
				_[0.58]_	_[0.70]_	_[0.49]_			

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.

Notes: Significant statistics in bold; best forecasts in italics; * and ** denote significance at 5 percent and 1 percent, respectively.

[1] Best UR/ARMA model was standard unit root model.

[2] Best UR/ARMA model was unit root model with lagged futures prices.

[3] Best UR/ARMA model was ARMA (p,q) model with (p,q) equal to (1,1) for aluminum; (4,4) for copper; (2,6) for lead; (2,3) for nickel; (2,5) for sugar; (1,2) for cotton; and (2,10) for coffee--robusta.

[4] Best UR/ARMA model was ARMA (p,q) model with futures, with (p,q) equal to (4,4) for tin; (2,3) for wheat; (2,2) for maize; and (3,1) for soybean meal.

[5] Results were affected by a spike in the data; hence statistical results are grossly inaccurate. Direction statistics were unaffected.

[6] Results indicated a similar price movement over all estimation periods; hence it was not possible to calculate Cumby-Modest and/or Harding-Pagan statistics.

Table 8. Forecast Performance at Eight-Quarter Horizon

	RMSE Ratio (RMSE actual value)			Harding Pagan Statistic (β) (*t*-statistic) Concordance Statistic			Cumby-Modest Statistic (*t*-statistic)		
Commodity	Judgment	Best UR/ ARMA	ECM	Judgment	Best UR/ ARMA	ECM	Judgment	Best UR/ ARMA	ECM
Aluminum [4]	1.00	1.60	1.04	- [5]	-0.22	0.35**	- [5]	-0.05	0.11**
	(264.28)	(422.95)	(275.69)	-	(-0.15)	(2.66)	-	(-0.75)	(2.74)
				[0.43]	[0.62]	[0.59]			
Copper [2]	1.00	1.22	1.19	- [5]	-0.12	0.23	0.08	-0.24**	0.12
	(542.42)	(660.28)	(643.98)	-	(-0.86)	(1.14)	(0.75)	(-3.32)	(1.33)
				[0.86]	[0.62]	[0.62]			
Lead [2]	1.00	1.32	1.04	- [5]	0.11	0.07	0.21	-0.12	0.18*
	(126.12)	(167.09)	(131.47)	-	(0.40)	(0.34)	(1.97)	(-0.98)	(2.24)
				[0.71]	[0.70]	[0.54]			
Nickel [2]	1.00	1.37	1.13	0.50	-0.01	0.44**	0.20	-0.23	0.24*
	(1971.64)	(2710.85)	(2221.59)	(1.20)	(-0.03)	(2.99)	(0.50)	(-1.52)	(2.28)
				[0.86]	[0.51]	[0.62]			
Tin [4]	1.00	1.50	1.11	0.50	0.21	0.17	0.20	0.05	-0.00
	(733.11)	(1093.49)	(810.76)	(1.20)	(1.41)	(1.08)	(1.53)	(0.81)	(-0.03)
				[0.86]	[0.54]	[0.60]			
Zinc [4]	1.00	1.52	1.35	0.33	-0.47**	0.68**	0.44**	-0.29**	0.29**
	(174.64)	(264.90)	(236.38)	(1.04)	(-3.48)	(5.36)	(5.46)	(-4.85)	(4.21)
				[0.71]	[0.46]	[0.81]			
Wheat [2]	1.00	1.20	0.80	0.50	0.55**	0.56**	0.23	- [5]	0.38**
	(42.59)	(51.18)	(34.11)	(1.15)	(3.05)	(2.83)	(0.93)	-	(3.91)
				[0.83]	[0.65]	[0.78]			
Maize [3]	1.00	1.16	1.15	0.50	0.09	0.39*	0.08	0.10	0.32**
	(22.85)	(26.48)	(26.38)	(1.15)	(0.63)	(2.54)	(0.49)	(0.93)	(3.23)
				[0.83]	[0.62]	[0.70]			
Soybean [1]	1.00	1.16	0.91	-0.17	0.41**	0.31*	0.13	- [5]	0.21**
	(47.86)	(55.34)	(43.42)	(-0.91)	(2.72)	(1.67)	(1.07)	-	(2.94)
				[0.71]	[0.54]	[0.68]			
Soybean Meal [1]	1.00	1.33	0.80	- [5]	0.27	0.67*	- [5]	-0.15	0.55**
	(56.66)	(75.24)	(45.14)	-	(1.15)	(4.17)	-	(-1.41)	(5.18)
				[0.00]	[0.59]	[0.84]			
Soybean Oil [2]	1.00	1.65	1.25	0.80**	-0.15	0.58**	0.32	-0.31**	0.37**
	(101.56)	(167.92)	(126.81)	(3.66)	(-0.64)	(2.58)	(1.13)	(-2.68)	(3.11)
				[0.86]	[0.43]	[0.81]			
Sugar (US) [3]	1.00	1.54	1.01	- [5]	0.46**	0.59**	0.24*	- [6]	0.06
	(1.48)	(2.29)	(1.50)	-	(2.72)	(4.34)	(3.68)	-	(2.24)
				[1.00]	[0.59]	[0.81]			
Cotton [2]	1.00	1.28	1.00	- [5]	0.13	0.11	-0.15	0.05	0.06
	(18.98)	(24.35)	(18.93)	-	(1.41)	(0.92)	(-1.29)	(1.20)	(1.11)
				[0.57]	[0.46]	[0.68]			
Coffee (other) [3]	1.00	1.39	1.11	-0.17	0.32*	-0.16	-0.62	0.23 *	-0.14
	(50.20)	(70.02)	(55.93)	(-0.91)	(2.15)	(-1.00)	(-2.56)	(2.06)	(-1.15)
				[0.71]	[0.59]	[0.54]			
Coffee (robusta) [4]	1.00	1.91	1.61	-0.17	0.32**	0.11	-0.92*	0.44**	- [5]
	(46.75)	(89.28)	(75.26)	(-0.91)	(2.80)	(0.99)	(-5.55)	(3.54)	-
				[0.71]	[0.65]	[0.39]			

Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.
Notes: Significant statistics in bold; best forecasts in italics; * and ** denote significance at 5 percent and 1 percent, respectively.

[1] Best UR/ARMA model was standard unit root model.

[2] Best UR/ARMA model was unit root model with lagged futures prices.

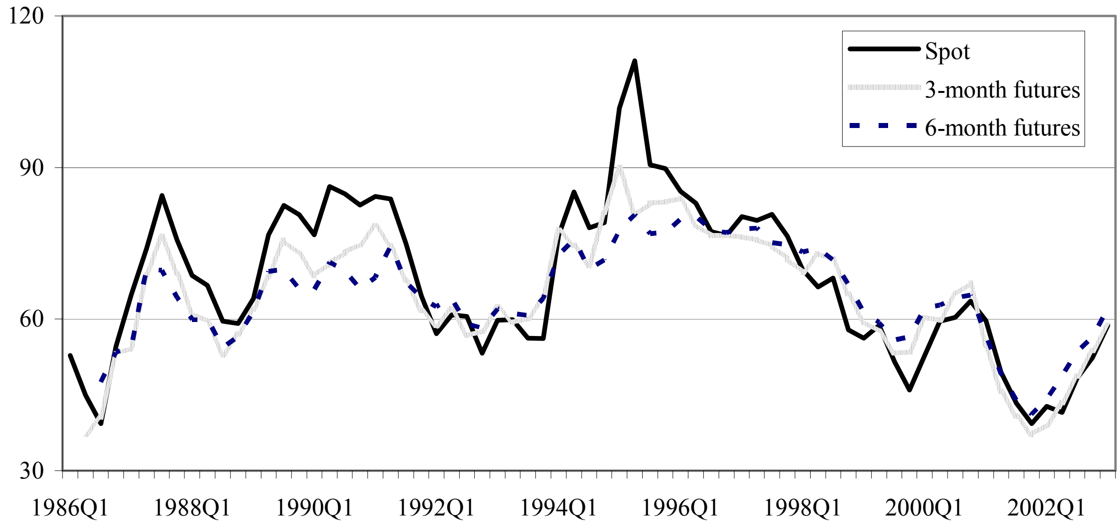
[3] Best UR/ARMA model was ARMA (p,q) model with (p,q) equal to (6,10) for maize; (2,5) for sugar; and (1,1) for coffee--other milds.

[4] Best UR/ARMA model was ARMA (p,q) model with futures, with (p,q) equal to (4,3) for aluminum; (4,4) for tin; (1,2) for zinc; and (2,2) for coffee--robusta.

[5] Results indicated a similar price movement over all estimation periods; hence it was not possible to calculate Cumby-Modest and/or Harding-Pagan statistics.

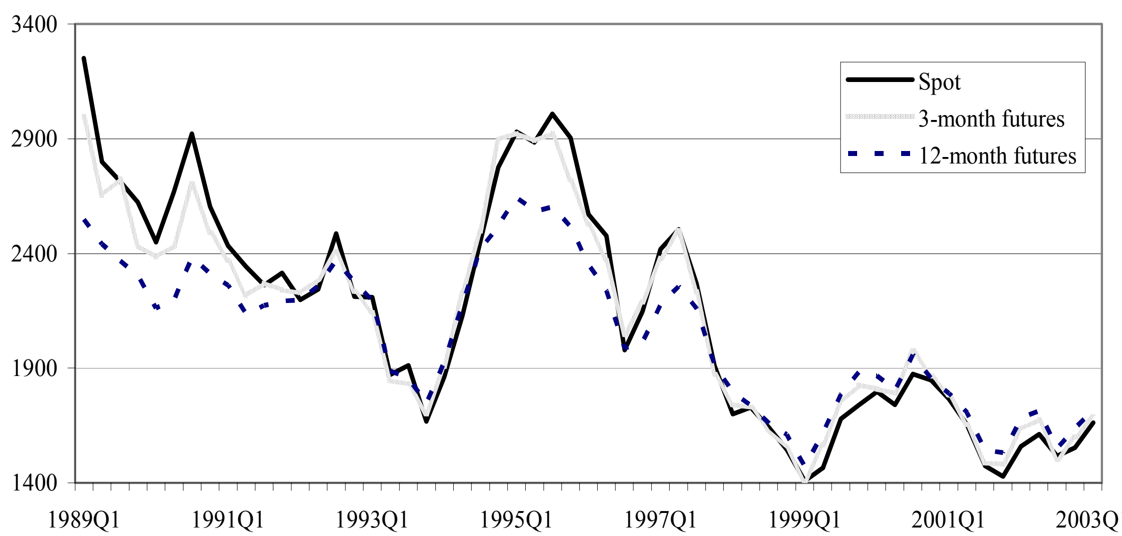
[6] Results demonstrated no heteroskedasticity; hence it was not possible to calculate statistics.

Figure 1. Cotton: Spot and Futures Prices, 1986-2003
(Cents per pound)



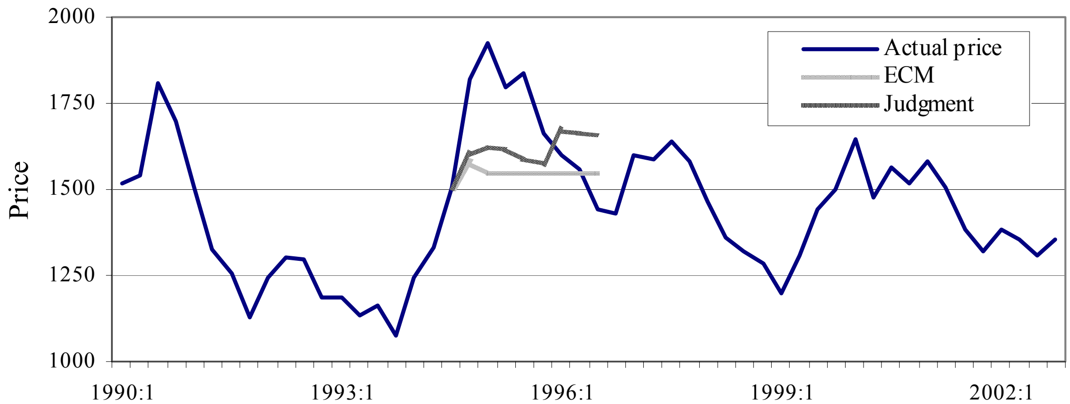
Sources: IMF Primary Commodity Prices Database and Bloomberg Financial, LP.

Figure 2. Copper: Spot and Futures Prices, 1989-2003
(Dollars per metric ton)



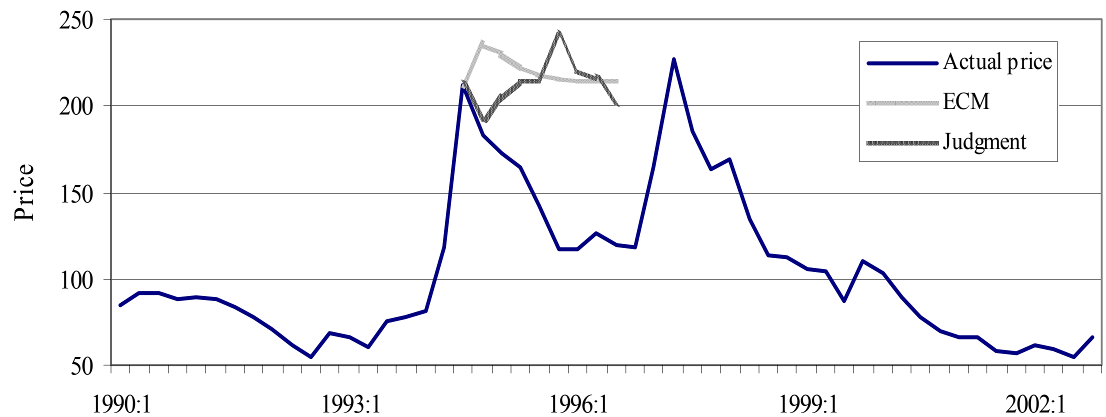
Sources: IMF Primary Commodity Prices Database and Bloomberg Financial, LP.

Figure 3. Aluminum: Judgmental and ECM Forecasts, 1994: Q3



Sources: IMF Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.

Figure 4. Coffee (Other Milds): Judgmental and ECM Forecasts, 1994: Q3



Sources: IMF Primary Commodity Prices Database; Bloomberg Financial, LP; and authors' estimates.

REFERENCES

Abramson, B., and A. Finizza, 1995, "Probabilistic Forecasts from Probabilistic Models: A Case Study in the Oil Market," *Journal of Forecasting*, Vol. 11(1), pp. 63–72.

Avsar, S.G., and B.A. Goss, 2001, "Forecast Errors and Efficiency in the U.S. Electricity Futures Market," *Australian Economic Papers*, Vol. 40(4), pp. 479–499.

Bessler, D.A., and J.A. Brandt, 1992, "An Analysis of Forecasts of Livestock Prices," *Journal of Economic Behavior and Organizations*, Vol. 18(2), pp. 249–263.

Brenner, R.J., and F.K. Kroner, 1995, "Arbitrage, Cointegration and Testing the Unbiasedness Hypothesis in Financial Markets," *Journal of Financial and Quantitative Analysis*, Vol. 30(1), pp. 23–42.

Cashin, P., H. Liang, and C.J. McDermott, 2000, "How Persistent are Shocks to World Commodity Prices?," *IMF Staff Papers*, Vol. 47(2), pp. 177–217.

Cashin, P., and C.J. McDermott, 2002, "The Long-run Behavior of Commodity Prices: Small Trends and Big Variability, *IMF Staff Papers*, Vol. 49(2), pp. 175–199.

Cashin, P., C.J. McDermott, and A. Scott, 1999, "The Myth of Commoving Commodity Prices," IMF Working Paper 99/169.

Cuddington, J.T., 1992, "Long-run Trends in 26 Primary Commodity Prices: A Disaggregated Look at the Prebisch-Singer Hypothesis," *Journal of Development Economics*, Vol. 39(2), pp. 207–227.

Cumby, R.E., and D.M. Modest, 1987, "Testing for Market Timing Ability: A Framework for Forecast Evaluation," *Journal of Financial Economics*, Vol. 19(1), pp. 169–189.

Engle, R.F., and C.J.W. Granger, 1987, "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, Vol. 55(2), pp. 251–276.

Gerlow, M.E., S.H. Irwin, and T. Liu, 1993, "Economic Evaluation of Commodity Price Forecasting Models," *International Journal of Forecasting*, Vol. 9(3), pp. 387–397.

Granger, C.W.J., and M.H. Pesaran, 1999, "Economic and Statistical Measures of Forecast Accuracy," University of Cambridge, DAE Working Paper 9910.

Harding, D., and A. Pagan, 2002, "Synchronisation of Cycles," mimeo., Australian National University.

Heaney, R., 2002, "Does Knowledge of the Cost of Carry Model Improve Commodity Futures Price Forecasting Ability? A Case Study Using the London Metal Exchange Lead Contract," *International Journal of Forecasting*, Vol. 18(1), pp. 45–65.

Irwin, S.H., M.E. Gerlow, and T. Liu, 1994, "The Forecasting Performance of Livestock Futures Prices: A Comparison to USDA Expert Predictions," *Journal of Futures Markets*, Vol. 14(7), 861–875.

Irwin, S.H., C.R. Zulauf, and T.E. Jackson, 1996, "Monte Carlo Analysis of Mean Reversion in Commodity Futures Prices," *American Journal of Agricultural Economics*, Vol. 78(2), pp. 397–399.

Just, R.E., and G.C. Rausser, 1981, "Commodity Price Forecasting with Large-Scale Econometric Models and the Futures Market," *American Journal of Agricultural Economics*, Vol. 63(2), pp. 197–208.

Kaminsky, G., and M.S. Kumar, 1990a, "Efficiency in Commodity Futures Markets," *IMF Staff Papers*, Vol. 37(3), pp. 671–699.

———, 1990b, "Time-varying Risk Premia in Futures Markets," IMF Working Paper 90/116.

Kumar, M.S., 1992, "The Forecasting Accuracy of Crude Oil Futures Prices," *IMF Staff Papers*, Vol. 39(2), pp. 432–461.

Leitch, G., and J.E. Tanner, 1991, "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, Vol. 81(3), pp. 580–590.

León, J., and R. Soto, 1997, "Structural Breaks and Long-run Trends in Commodity Prices," *Journal of International Development*, Vol. 9(3), pp. 347–366.

Lutz, M.G., 1999, "A General Test of the Prebisch-Singer Hypothesis," *Review of Development Economics*, Vol. 3(1), pp. 44–57.

McKenzie, A.M., and M.T. Holt, 2002, "Market Efficiency in Agricultural Futures Markets," *Applied Economics*, Vol. 34(12), pp. 1519–1532.

Merton, R.C., 1981, "On Market Timing and Investment Performance I : An Equilibrium Theory of Value for Market Forecasts," *Journal of Business*, Vol. 54(3), pp. 363–406.

Mills, T.C., 1999, *The Econometric Modeling of Financial Time Series*, Cambridge University Press, Cambridge, United Kingdom.

Moosa, I.A., and N.E. Al-Loughani, 1994, "Unbiasedness and Time-varying Risk Premia in the Crude Oil Futures Market," *Energy Economics*, Vol. 16(2), pp. 99–105.

Morana, C., 2001, "A Semiparametric Approach to Short-term Oil Price Forecasting," *Energy Economics*, Vol. 23(3), pp. 325–338.

Newey, W., and K. West, 1994, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, Vol. 61(4), pp. 631–653.

White, H., 1980, "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, Vol. 48(4), pp. 817–835.