The Relative Efficiency of Commodity Futures Markets

NEIL KELLARD
PAUL NEWBOLD
TONY RAYNER
CHRISTINE ENNEW*

The ability of futures markets to predict subsequent spot prices has been a controversial topic for a number of years. Empirical evidence to date is mixed; for any given market, some studies find evidence of efficiency, others of inefficiency. In part, these apparently conflicting findings reflect differences in the time periods analyzed and the methods chosen for testing. A limitation of existing tests is the classification of markets as either efficient or inefficient with no assessment of the degree to which efficiency is present. This article presents tests for unbiasedness and efficiency across a range of commodity and financial futures markets, using a cointegration methodology, and de-

The authors wish to thank two anonymous referees for helpful comments and suggestions. Neil Kellard also gratefully acknowledges the financial support provided by a MAFF research studentship. The research was completed under a grant, F/114/BH, from The Leverhulme Trust.
*Correspondence author, Business School, The University of Nottingham, Social Sciences Building, University Park, Nottingham, NG7 2RD, United Kingdom.

- *Neil Kellard is in the Department of Economics at the University of Nottingham in the United Kingdom.*
- *Paul Newbold is a Professor in the Department of Economics at the University of Nottingham in the United Kingdom.*
- *Tony Rayner is in the Department of Economics at the University of Nottingham in the United Kingdom.*
- *Christine Ennew is a Professor in the University of Nottingham Business School in the United Kingdom.*

The Journal of Futures Markets, Vol. 19, No. 4, 413–432 (1999)

CCC 0270-7314/99/040413-20

velops a measure of relative efficiency. In general, the findings suggest that spot and futures prices are cointegrated with a slope coefficient that is close to unity, so that the postulated long-run relationship is accepted. However, there is evidence that the long-run relationship does not hold in the short run; specifically, changes in the spot price are explained by lagged differences in spot and futures prices as well as by the basis. This suggests that market inefficiencies exist in the sense that past information can be used by agents to predict spot price movements. A measure of the relative degree of inefficiency (based on forecast error variances) is then used to compare the performance of different markets. Jrl Fut Mark 19: 413–432, 1999

INTRODUCTION

The value of futures markets arises from their ability to forecast cash prices at a specified future date and thus provide agents with a means of managing the risks associated with trading in a given commodity. In an efficient commodity market the futures price will be an optimal forecast of the spot price at contract termination in the sense that it will only be proved wrong to the extent of a random unpredictable zero-mean error. In its simplest form the Efficient Market Hypothesis (EMH) can be reduced to the joint hypothesis that agents are, in an aggregate sense, endowed with rational expectations and are risk neutral so that the futures price is an unbiased estimator of the future spot price (Taylor, 1995). The conventional econometric evaluation of EMH is normally carried out within a cointegration and error correction framework (Fujihara & Mougoué, 1997), recognizing that the time series for spot and futures are usually nonstationary variables (Shen and Wang, 1990). A number of reasons have been advanced for market inefficiency in cases where tests support rejection of these hypotheses, some of which include the presence of a risk premium (Krehbiel & Adkins, 1993), the inability of the futures price to reflect all publicly available information (Beck, 1994), and the inefficiency of agents as information processors (Kaminsky & Kumar, 1990). Also, as noted for example by Fortenbery and Zapata (1993), cointegration may not be found in efficient markets for commodities in which returns to storage or transportation are nonstationary.

The tests of EMH are a useful first step in the evaluation of the social utility of futures markets. However, as argued cogently by Stein (1986), such tests do not provide information about the degree of inefficiency in any specific market and, consequently, do not allow a quantitative comparison of the functioning of different futures markets. In this paper, a quantitative measure of inefficiency is proposed—based on the notion of the futures price as a forecast of the subsequent spot price, and calculated from the results of the econometric analysis of a futures market. Specifically, inefficiency is estimated as a ratio of the forecast error variance from the "best-fitting" quasi-error correction model (ECM) to the forecast error variance of the futures price as predictor of the spot price. This ratio equals unity in the efficient market case, but is less than unity for an inefficient market—that is, the lower the value the greater the implied inefficiency.

In order to illustrate the potential usefulness of this approach, a comparative analysis is carried out on a number of rather different futures markets, specifically soybeans, live cattle, live hogs, gasoil, crude oil (Brent) and the Deutsch mark/dollar exchange rate, employing a carefully selected and standardized data set. All commodities selected were characterized by a regular settlement pattern, thus avoiding any of the problems that arise in the time series analysis of unevenly spaced data. For each market, data are sampled on each futures price at 28 days to maturity on the nearest contract; the choice of a date close to contract maturity helps to minimize inefficiency arising from a risk premium and from learning errors. The common forecast horizon of 28 days is less than the contract period (one or two months depending on the commodity) and this is explicitly accounted for in the time series modeling. For the four non-oil contracts, where the contract period is two months, our analysis is extended to a forecast horizon of 56 days. The analysis proceeds by first testing for market efficiency within the framework of a quasi-ECM and, second, estimating the measure of inefficiency for each market. Finally, a comparison is made of the relative inefficiencies of the various markets and conclusions are drawn.

DATA AND SAMPLING ISSUES

To analyze the efficiency of futures markets for each commodity selected, time series are required of the future spot price and futures prices that correspond to those spot prices, conventionally known as S_t and F_{t-1} respectively. Following Crowder and Hamed (1993), the future spot price is the cash price on the termination day of the futures contract. The frequency of each series will therefore depend on how many contract months there are for each commodity. However, the construction of the second series still needs further information. A contract is often open for many months, and all of the subsequent daily futures prices reflect the changing market expectation of what the spot price will be on the last day of trading. Standard time series techniques require the selection of *one* of these futures prices to be *matched* with the termination spot price. Consider Brent crude oil futures, which have twelve contract months a year, and thus a time series of twelve termination spot prices a year. The matching futures prices should be sampled from a specific day, less than a month from the last day of trading. If matching futures prices are sampled from a day further than a month from the last day of trading, our time series analysis will suffer from autocorrelation problems because of informational overlap (see Hansen & Hodrick, 1980) in the sense that the previous contract will still be trading when the matching price for the next contract is sampled. Autocorrelation in the errors of the usual regression equation for testing efficiency then induces the appearance of inefficiency even in efficient markets. The futures price, therefore, is selected by working backward 28 days from the contract termination date. A common 28-day forecast horizon also allows us to compare across markets with contracts of different durations. The choice of contracts that are less frequently traded than Brent or gasoil would allow a longer forecast duration and later in the paper these two contracts are dropped from the analysis to facilitate the comparison of efficiency over 28- and 56-day time horizons.

The conventional process of testing for efficiency requires first testing for the presence of cointegration and second testing that the futures price at contract purchase is an unbiased predictor of the spot price at contract termination (see Chowdhury, 1991, and Lai & Lai, 1991). The cointegrating regression is conventionally specified in logarithms of the data series (see Fujihara and Mougoué, 1997) as

$$s_t = \beta_0 + \beta_1 f_{t-1} + u_t \tag{1}$$

where f_{t-1} is the logarithm of the lagged futures price (that is, the futures price at contract purchase) and s_t is the logarithm of the spot price that is matched to the settlement date of the futures contract. The unbiasedness hypothesis is that $\beta_0 = 0$, $\beta_1 = 1$ and u_t is white noise. As is well known, models for cointegrated series can be rewritten in an error-correction representation, described in Granger (1986). If the restriction $\beta_1 = 1$ is not rejected, then there is not strong evidence against the hypothesis that, in long-run equilibrium, the spot price is equal to the futures price plus a (possibly zero) constant. Consequently, it is permissible to specify a short run regression relating the change in spot price to the percentage basis $(f_{t-1} - s_{t-1})$ and lagged changes in spot and futures prices:

$$\Delta s_t = \alpha + \rho(f_{t-1} - s_{t-1}) + \sum_{i=1}^{k} \lambda_i \Delta s_{t-i} + \sum_{i=1}^{k} \gamma_i \Delta f_{t-i} + \varepsilon_t \quad (2)$$

Model (2) provides the foundation of a test of whether information additional to the basis is optimally used in forecasting changes in the spot rate. The statistical significance of lags in eq. (2) is usually claimed as evidence of inefficiency.

However, the above econometric approach is not directly applicable to the data sets chosen for analysis. The use of a common 28-day forecast horizon means that if the logarithm of the termination spot price is defined as s_t, it is incorrect to label the logarithm of the lagged futures price as f_{t-1}, because, clearly, unless the time between termination spots is 28 days[1], s_{t-1} and f_{t-1} will not be sampled on the same day. Periods that are notationally suggested as separate, $(s_t - s_{t-1})$ and $(f_{t-1} - f_{t-2})$, would actually overlap and interpretation of (2) becomes meaningless.

To clarify the approach taken here, the log of the futures price 28 days from maturity is defined as $f_{t-\tau}$, the τ representing a fraction of the unit of observation between contract months. A notationally correct, cointegrating regression becomes

$$s_t = \beta_0 + \beta_1 f_{t-\tau} + u_t \quad (3)$$

An equation in the spirit of eq. (2) would be a model to explain the change in the spot rate over the forecast horizon. This could be regressed on the basis sampled at the start of the forecast horizon, and the changes in spot and future rates over lagged forecast horizons. In this way there is no informational overlap between variables and the basis is defined correctly.

Two new time series have to be created for each commodity. The logarithm of the spot price sampled on the same day as $f_{t-\tau}$, labeled $s_{t-\tau}$, and the futures price sampled on the same day as s_t, labeled f_t. A useful quasi-ECM to assess efficiency can therefore be defined as

$$\begin{aligned} s_t - s_{t-\tau} = \theta_0 + \theta_1(f_{t-\tau} - s_{t-\tau}) &+ \sum_{i=1}^{k} \lambda_i (s_{t-i} - s_{(t-\tau)-i}) \\ &+ \sum_{i=1}^{k} \gamma_i (f_{t-i} - f_{(t-\tau)-i}) + \varepsilon_t \quad (4) \end{aligned}$$

Summary information on the commodity data analyzed is presented in the Appendix.

[1]The distance between termination spots is greater than our chosen forecast horizon for all markets analyzed. See Table I.

TABLE I

Tests of Cointegration Rank

Commodity	H_0: r	l-max[a]	Trace[a]	Lag Length	Comment
Brent crude	0	177.9[b]	185.3[b]	1	rank 1
	1	7.3	7.3		reject non-cointegration
Live hogs	0	69.0[b]	86.7[b]	1	rank = 2
	1	17.8[b]	17.8[b]		spot and futures I(0)
Live hogs	0	22.7[b]	31.0[b]	6	rank = 1
	1	8.3	8.3		reject non-cointegration
Soybeans	0	104.4[b]	113[b]	1	rank = 1
	1	8.6	8.6		reject non-cointegration
Gasoil	0	175.7[b]	184.3[b]	1	rank = 1
	1	8.6	8.6		reject non-cointegration
Live cattle	0	50.9[b]	58.7[b]	1	rank = 1
	1	7.9	7.9		reject non-cointegration
Dm/$	0	79.5[b]	82.1[b]	1	rank = 1
	1	2.7	2.7		reject non-cointegration
Critical values	0	15.7	20.0		
(0.95)	1	9.2	9.2		

[a]Using small sample degrees of freedom correction (Reimers 1992).
[b]Reject null at 1% level.
[c]Reject null at 5% level.

TESTING FOR MARKET EFFICIENCY

As indicated earlier, the process of testing for efficiency initially requires tests for the order of integration of the individual series; if these series are found to be nonstationary, then it is necessary to test for cointegration as a precondition for market efficiency, and subsequently for unbiasedness. Augmented Dickey-Fuller tests suggested that the null of a unit autoregressive root, that is, integration of order 1, I(1), could not be rejected for any series. With the same order of integration for relevant spot and futures prices, the next stage entailed testing for the presence of cointegration. Testing for cointegration and testing for the joint restriction $\beta_0 = 0$ and $\beta_1 = 1$ in eq. (3) was carried out using the Johansen technique in PcFiml (Doornik & Hendry, 1997).

Table I presents the cointegration results from the application of the Johansen method of reduced rank regression to the 28-day forecast horizon data. This technique specifies the vector error correction model (VECM) of the m-variable VAR for a time series vector X_t as

$$\Delta X_t = \delta_0 + \delta_1 \Delta X_{t-1} + \ldots\ldots + \delta_{k-1}\Delta X_{t-k+1} + \Pi X_{t-k} + v_t \quad (5)$$

where k is sufficiently large that v_t is vector white noise. This technique then tests for the rank of Π, the $m \times m$ parameter matrix attached to the vector of the (lagged) levels of the variables. The lag length k was chosen using the Schwarz Information Criterion. Assuming that X_t is a vector of I(1) variables, then ΠX_{t-k} has to be stationary for v_t to be stationary. The absence of cointegration implies that there are no linear combinations of the X_t that are stationary I(0) and that the rank (r) of Π is zero.

Our case specifies $X_t = (s_t, f_{t-\tau})$ so that Π has a maximum rank of 2. Using the Johansen λ-max (maximal-eigenvalue) and trace statistics[2], the technique sequentially tests the null hypotheses $r = 0$ and $r = 1$. For all commodities, the null of no-cointegration ($r = 0$) is rejected at the 1% significance level. For all commodities except live hogs, the null of reduced rank, $r = 1$, cannot be rejected, and hence cointegration is implied. The rejection of reduced rank for live hogs would imply that the data series for this commodity are stationary, despite the earlier conclusions drawn from the Dickey-Fuller tests. However, the result is sensitive to the lag length chosen for the VECM; As shown in Table I, with a lag length of 6, rank $r = 1$ is not rejected. Although the Schwarz criterion consistently estimates lag length in a correctly specified model, it does, on occasion, produce an overly parsimonious approximation to the underlying data generation process. In this case the analysis based on lag length 6 is preferred, in line with the conclusions of Dickey-Fuller tests on individual series.

Table II presents the test results for unbiasedness. The joint restrictions $\beta_0 = 0$ and $\beta_1 = 1$ imposed on the cointegrating regression (3) are assessed through a likelihood ratio test. The test statistic follows a χ^2 distribution with two degrees of freedom under the null hypothesis. The null is not rejected for live cattle and Dm/$. For all commodities except live hogs, the unit slope restriction on the cointegrating vector is not rejected. Consequently, there is evidence to suggest that efficiency and unbiasedness hold in the long run equilibrium for all but one of the commodities tested. These findings show some consistency with previous tests of commodity market efficiency; Garcia et al. (1988) cite evidence for semistrong form efficiency in relation to live cattle while Beck (1994) suggests that a range of commodity markets, including cattle and corn, are sometimes efficient according to forecast horizon. Ma (1989) also cites evidence for a degree of efficiency in oil futures, although she does note that composite forecasting models can outperform the futures price as a predictor of future spot price. Interestingly, Leuthold et al. (1989),

[2]The trace statistic is a likelihood ratio statistic with a nonstandard null distribution.

TABLE II

Test: $\beta_0 = 0$ and $\beta_1 = 1$ in Cointegrating Regression

Commodity	χ^2 (2)	*p-value*
Brent crude	8.52	0.01
Gasoil	21.8	0.00
Soybeans	14.3	0.00
Live hogs (6 lags)	12.0	0.00
Live cattle	5.72	0.06
$/Dm	4.62	0.10

0.95 critical value χ^2 (2) = 5.99

Test: $\beta_1 = 1$ in Cointegrating Regression

Commodity	χ^2 *(1)*	*p-value*
Brent crude	3.26	0.07
Gasoil	0.59	0.44
Soyabeans	1.22	0.27
Live hogs (6 lags)	8.28	0.00

0.95 critical value χ^2 (1) = 3.84.

Beck (1994), and Schroeder and Goodwin (1991) identify problems with efficiency testing in relation to live hogs; the current analysis also encounters problems as indicated earlier; but for consistency the series will be retained for the short-run analysis. Of course, long-run efficiency is not sufficient for short-run forecast efficiency, the analysis of which requires fitting eq. (4).

Table III presents the results of the short run OLS regressions (4). The lag lengths were selected through general-to-specific testing by initially setting $k = 10$ and dropping lags that were insignificant at the 10% level while preserving symmetry on lag length for $(s_{t-i} - s_{(t-\tau)-i})$ and $(f_{t-i} - f_{(t-\tau)-i})$. This procedure removes all evidence of residual serial correlation. Table IV provides test statistics for the null hypothesis that the coefficients of all lagged variables are zero. For purposes of comparison, Table V presents results for regressions omitting all lagged changes in spot and futures prices. From a comparison of Tables III and V, it is apparent that the estimates of the coefficient of the basis θ_1 are quite sensitive to model specification as regards the inclusion of lagged differences of spot and futures prices. These results suggest that, for all commodities apart from soybeans, information additional to the basis predicts short-run movements in the spot price, although as indicated by Table IV

TABLE III

OLS Regressions for

$$s_t - s_{t-\tau} = \theta_0 + \theta_1 (f_{t-\tau} - s_{t-\tau}) + \sum_{i=1}^{k} \lambda_i (s_{t-i} - s_{(t-\tau)-i}) + \sum_{i=1}^{k} \gamma_i (f_{t-i} - f_{(t-\tau)-i}) + \varepsilon_t$$

	Brent Crude	*Gasoil*	*Soybeans*	*Live Hogs*	*Live Cattle*	*Dm/$*
θ_0	0.01	0.02	−0.02	−0.01	−0.01	−0.01
	(0.82)	(1.67)	(−2.62)	(−0.60)	(−2.16)	(−1.30)
θ_1	−0.76	1.18	1.25	0.49	0.22	2.54
	(−1.48)	(3.47)	(4.36)	(2.75)	(1.92)	(3.40)
λ_1	0.39	−0.29	—	−0.28	0.06	−0.30
	(0.85)	(−0.78)		(−2.13)	(0.32)	(−0.58)
λ_2	0.27	0.15	—	—	0.04	0.37
	(0.60)	(0.41)			(0.24)	(0.69)
λ_3	−0.23	−0.16	—	—	−0.07	0.51
	(−0.49)	(−0.46)			(−0.47)	(0.94)
λ_4	0.12	0.11	—	—	−0.41	−0.45
	(0.27)	(0.33)			(−0.89)	(−0.85)
λ_5	−0.59	0.28	—	—	−0.01	−0.10
	(−1.29)	(0.81)			(−0.09)	(−0.19)
λ_6	−0.46	0.08	—	—	0.04	0.23
	(−0.96)	(0.24)			(0.27)	(0.44)
λ_7	−0.13	0.44	—	—	−0.28	0.91
	(−0.29)	(1.30)			(−1.84)	(1.92)
λ_8	0.72	0.87	—	—	—	—
	(1.61)	(2.66)				
γ_1	−0.50	0.40	—	0.21	−0.03	0.26
	(−1.13)	(0.93)		(1.76)	(−0.25)	(0.47)
γ_2	−0.39	−0.22	—	—	−0.26	−0.65
	(−0.88)	(−0.52)			(−2.32)	(−1.11)
γ_3	0.26	0.24	—	—	−0.08	−0.58
	(0.59)	(0.58)			(−0.73)	(−0.99)
γ_4	−0.26	−0.21	—	—	0.01	0.46
	(−0.59)	(−0.51)			(0.13)	(0.78)
γ_5	0.42	−0.32	—	—	−0.03	0.03
	(0.94)	(−0.79)			(−0.31)	(0.05)
γ_6	0.30	−0.00	—	—	0.03	−0.26
	(0.63)	(−0.00)			(0.30)	(−0.46)
γ_7	−0.13	−0.45	—	—	0.04	−0.93
	(−0.28)	(−1.10)			(0.39)	(−1.83)
γ_8	−0.78	−1.14	—	—	—	—
	(−1.66)	(−2.84)				

Note: *t* statistics in parentheses.

TABLE IV

Joint Test of Zero Restrictions on the Coefficients of Lagged Variables in Short-Run Regressions

Commodity	*F*	*p-value*
Brent crude	F(16,58) = 1.26	0.26
Gasoil	F(16,57) = 0.90	0.58
Soybeans	—	—
Live hogs	F(2,82) = 2.40	0.10
Live cattle	F(14,64) = 1.72	0.07
Dm/$	F(14,59) = 1.21	0.30

TABLE V

OLS Regression for $s_t - s_{t-\tau} = \theta_0 + \theta_1 (f_{i-\tau} - s_{t-\tau}) + \varepsilon_t$

Commodity	θ_0	θ_1	*p-value*
Brent crude	0.01 (0.74)	0.26 (0.54)	0.12
Gasoil	0.03 (2.73)	1.54 (5.84)	0.04
Soybeans	−0.02 (−2.62)	1.25 (4.36)	0.39
Live hogs	−0.01 (−1.24)	0.64 (3.96)	0.03
Live cattle	−0.01 (−1.50)	0.29 (3.03)	0.00
Dm/$	0.00 (0.05)	1.36 (2.43)	0.52

Note: *t* statistics in parenthesis: *p* value is for the test of the hypothesis $\theta_1 = 1$.

there is not very strong evidence to support the inclusion of the set of lagged variables except for cattle and hogs.

MEASURING SHORT RUN INEFFICIENCY

The estimates reported in Table III suggest, to a greater or lesser extent, the possibility of some short-run inefficiency in five of the six markets. However, such tests cannot indicate degrees of efficiency or give an indication of how well or badly a futures market is functioning. Yet an indicator of degree of efficiency is potentially of considerable value—given that the cost of hedging rises as markets become less efficient (Krehbiel & Adkins, 1993), and given that there is an association between inefficiency and the social costs associated with futures trading (Stein, 1986). From the viewpoint of a time series econometrician, an appropriate criterion of inefficiency is based upon the *relative* ability of the futures price to forecast the subsequent spot price. The fitted eq. (4) provides the relevant comparative forecast of spot price. The error variance of that

TABLE VI

Efficiency Measures for 28 Days Forecast Horizon

Commodity	*Brent Crude*	*Gasoil*	*Soybeans*	*Live Hogs*	*Live Cattle*	*Dm/\$*
ϕ_c	0.88	0.99	1.00	0.93	0.53	0.96
Degree of inefficiency	0.12	0.01	0.00	0.07	0.47	0.04
Deviation of θ_1 from unity	3.50	0.53	0.86	2.83	7.10	2.10
$\bar{R}_1^2$	0.04	0.28	0.15	0.17	0.18	0.11
$\bar{R}_2^2$	−0.09	0.28	0.15	0.11	−0.56	0.07

θ_1 is the estimated coefficient on the basis, and its absolute deviation from unity is measured in terms of numbers of standard errors.

forecast is then the variance of ε_t, for which an unbiased estimator is available from the fitted regression. On the other hand, futures market efficiency would imply a forecast $f_{t-\tau} + E[(S_t - f_{t-\tau})]$, allowing for the possibility of a systematic discount or premium in futures prices. The error variance of this predictor can be estimated through the sample variance of $(s_t - f_{t-\tau})$. The ratio of these two forecast error variances then provides an estimate of the efficiency of the (mean-corrected) futures price as a predictor of the future spot price. Thus our short-run efficiency measure is

$$\phi_c = \frac{(n - 2k - 2)^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2}{(n - 1)^{-1} \sum_{t=1}^{n} [(s_t - f_{t-\tau}) - \overline{(s_t - f_{t-\tau})}]^2} \tag{6}$$

where n is the number of observations used in estimating eq. (4), and $(2k + 2)$ is the number of estimated parameters in that equation. The numerator of eq. (6) is the estimated error variance of the short-run OLS regression, whereas the denominator is the sample variance of the forecast error, based on the futures price, corrected for degrees of freedom. A ratio of one for the quantity that is estimated by eq. (6) would indicate efficiency, zero complete inefficiency, and values between 0 and 1 varying degrees of inefficiency. Values of ϕ_c for the six markets analyzed[3] are given in the first row of Table VI. Rows 2 and 3 of this table present respectively the degree of inefficiency $(1 - \phi_c)$ and the number of standard errors by which the estimated coefficient on the basis differs from unity in the fit

[3]For completeness, given the results of Table II, the short run analysis was repeated for live hogs removing the restriction of $\beta_1 = 1$ imposed on (4). This made little appreciable difference to the estimate of ϕ_c, the value changing from 0.93 to 0.95.

of eq. (4). Based on eq. (4), inefficiency arises because the futures price is a poor predictor of the future spot price (θ_1 differs from unity), and because other lagged variables provide additional information beyond that incorporated in the futures price. Measuring the extent to which θ_1 differs from unity thus provides an indicator of the ability of the futures price to predict future spot price. The simple correlation between the measure of efficiency and the extent to which θ_1 differs from unity is 0.96: This suggests that the main source of inefficiency is the magnitude (in statistical terms) of the deviation of the coefficient on the basis from unity. Conversely, the inclusion of lagged variables in the short-run regression would appear to explain only a small degree of the inefficiency of the futures price as a predictor of the subsequent spot price in the markets studied.

The inefficiency criterion indicates that the soybean market is efficient (which is consistent with the results obtained by Serletis and Scowcroft, 1991) and, as might be expected, very little inefficiency in the foreign exchange market. By contrast, the live cattle market is very inefficient (consistent with the results reported by Beck, 1994), but the hogs market is much less inefficient. It is also interesting to note the differing degrees of inefficiency in the two oil markets, with evidence of far more inefficiency for Brent crude than for gasoil.

Our efficiency measure ϕ_c is based on the least squares estimation of the regression (4); that is, on the standard approach to *testing* for market efficiency. It therefore provides a useful companion measure associated with conventional tests. Any such test is necessarily carried out at a more or less arbitrary significance level. On one hand, failure to reject a null hypothesis (of market efficiency) does not necessarily imply strong evidence of support for that hypothesis. Failure to reject will necessarily be more likely to occur, all other things equal, the lower the level of significance chosen for the test, the lower the amount of any inefficiency actually present, and the smaller the sample size for any given degree of inefficiency present. On the other hand, given a sufficiently large sample, even a very small degree of inefficiency will lead to rejection of the null hypothesis at conventional levels, so that "real world" significance does not correspond to *statistical significance*. For example, from the perspective of a hedger it may be of little value to know that a particular market is inefficient per se, but rather more useful to know the degree of inefficiency, because it is the latter which might be expected to have a bearing on overall hedging effectiveness. Thus the efficiency measure ϕ_c provides a reasonable *estimate* of degree of inefficiency without relying on pretesting. This calculation is therefore a useful companion to the conventional efficiency testing methodology. However, it is important to stress that both the conventional testing and our efficiency measure in effect assess the *relative* merits of two predictors—the futures price and a forecast based on the regression (4)—rather than the *absolute* quality of forecasts. It is entirely possible by such criteria that a futures market could be assessed as "efficient" even if the futures price was of negligible value in predicting the spot price. If this were the case, then the usefulness of such a market for hedging purposes might be limited.

It is useful to address the issue of absolute predictor quality through the calculation of two adjusted coefficient of determination-like measures

$$\bar{R}_1^2 = 1 - \frac{(n - 2k - 2)^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2}{(n - 1)^{-1} \sum_{t=1}^{n} [(s_t - s_{t-\tau}) - (\overline{s_t - s_{t-\tau}})]^2} \tag{7}$$

and

$$\bar{R}_2^2 = 1 - \frac{(n - 1)^{-1} \sum_{t=1}^{n} [(s_t - f_{t-\tau}) - (\overline{s_t - f_{t-\tau}})]^2}{(n - 1)^{-1} \sum_{t=1}^{n} [(s_t - s_{t-\tau}) - (\overline{s_t - s_{t-\tau}})]^2} \tag{8}$$

These compare respectively forecasts based on the regression (4) and the futures price as a predictor with prediction from the last available spot price. Clearly, from eq. (6), the efficiency measure ϕ_c is related to these two quantities through

$$\phi_c = \frac{1 - \bar{R}_1^2}{1 - \bar{R}_2^2} \tag{9}$$

Values of $\bar{R}_1^2$ and $\bar{R}_2^2$ are given for the six markets in the final two rows of Table VI. These results shed interesting light on the concept of efficiency. For example, although the DM/$ futures market is assessed as 96% "efficient," prediction through the futures price alone explains only 7% of the variability in spot price changes. Judged by our criteria, the performance of the gasoil futures price is impressive. In addition to the efficiency estimate $\phi_c = 0.99$, note that each of the $\bar{R}^2$ measures is a good deal higher for this market than any other. It is therefore possible to conclude that the futures price is a relatively successful predictor of the spot price, and that use of additional information fails to improve that

TABLE VII

Efficiency Measures for 56 Days Forecast Horizon

Commodity	*Soybeans*	*Live Hogs*	*Live Cattle*	*Dm/$*
ϕ_c	0.87	0.99	0.77	1.00
Degree of inefficiency	0.13	0.01	0.23	0.00
Deviation of θ_1 from unity	1.55	1.72	4.33	0.15
$\bar{R}_1^2$	0.26	0.33	0.23	−0.09
$\bar{R}_2^2$	0.15	0.32	0.00	−0.09

predictor noticeably. The case of the live cattle market is particularly striking. A large negative value for $\bar{R}_2^2$ implies that the futures price is substantially inferior to the most recent spot price as a predictor. Nevertheless, prediction gains can be achieved from futures price information in this market through the regression (4). Indeed, $\bar{R}_1^2$ is a little higher for live cattle than for any market other than gasoil. Of course, it is the large discrepancy between the two $\bar{R}^2$ measures that accounts for the very low efficiency value ϕ_c for live cattle futures.[4]

The analytical tools developed in this paper can in principle be applied to any forecast horizon. Of course, any empirical results derived will necessarily be specific to the particular horizon analyzed, and it is entirely possible that different conclusions may be derived from analyses based on shorter or longer horizons. For the four commodities with contract periods of at least two months, the previous analysis was repeated using a forecast horizon of 56 days. Final summary results are given in Table VII which has the same format as Table VI. Comparison of these tables certainly emphasizes the specificity of efficiency measure calculations to the forecast horizon considered. The ϕ_c measure for soybeans is lower at 56 days than at 28 days, whereas the measures in the other three markets are all higher at the longer horizon. Nevertheless, the live cattle market is still by some amount the least efficient. Turning to the $\bar{R}^2$ measures, notice first the small negative values for DM/$. This illustrates a point made earlier. Although a futures market may be judged as entirely efficient through either the standard tests or the associated ϕ_c measure, the futures price may be of no value whatever in the prediction of the spot price. By contrast with this currency market, the $\bar{R}^2$ measures for soy-

[4]This observation sheds light on what at first sight might seem a surprising result—that, in spite of cointegration between spot and futures prices we find such low efficiency in cattle futures. Clearly this is due to short-term dynamics. Once these are properly accounted for, predictability in the cattle market is at about the same level as in other markets. Such an issue can of course arise in any short-term forecasting model irrespective of a long-run relation implied by cointegration.

beans, live cattle and live hogs are all higher for a 56-day forecast horizon than for a 28-days horizon. Of course, this does *not* mean that spot price changes over 56 days are better predicted than over 28 days—the variance of the former is much larger than that of the latter. The implication is that the futures prices are of *relatively* more value in prediction over the longer horizon than over the shorter. In this sense, the performance of the live hogs futures price is particularly notable, mirroring that of gasoil futures over the 28-days horizon in Table VI. The efficiency measure for live hogs is $\phi_c = 0.99$, and use of the futures price alone in prediction yields $\bar{R}_2^2 = 0.32$, comfortably the highest value found.

HEDGING, SPECULATING, AND SIMULATED MARKET TRADING

A question of interpretation of the efficiency measure ϕ_c still remains. The live cattle futures market with a forecast horizon of 56 days has $\phi_c = 0.77$. Although it is clear what "23% inefficient" means in a mathematical sense, what is the economic value of this information? Consider the implications of this degree of inefficiency for an individual hedger. Quantitatively, the methodology developed restricts consideration to a hedge placed 56 days from termination, which is then wound up on that last day of trading. The proportionate cost of hedging for a given commodity is $(S_t - F_{t-\tau})/F_{t-\tau} \approx s_t - f_{t-\tau}$. Hence the proportionate cost corresponds to the forecast error in our analysis. Thus the ratio of forecast error variances can be viewed as the ratio of proportionate cost of hedging variances. In this sense, 23% inefficiency in the live cattle market can be interpreted as implying that the variance of proportionate hedging costs would be 23% lower if the futures price was an efficient predictor of the spot price.

The economic value of ϕ_c can also be looked at from the viewpoint of a speculator. Garcia et al. (1988) note that the finding of inefficiency in futures prices as forecasts is of little value unless one can systematically profit from the information. Does information contained in the efficiency measure ϕ_c allow economic agents to generate such cost adjusted profits?

In Leuthold et al. (1989) out-of-sample forecasts from an econometric model, an ARIMA model and a composite forecasting model are used to generate profits in the live hog futures market. A similar methodology is used here to simulate trading in the live cattle futures market, using a 56-day forecast horizon. From Table VII it can be seen that the relevant ϕ_c is 0.77, suggesting the live cattle market is 23% inefficient. To conclude that the efficiency measure ϕ_c is of practical value, it must

TABLE VIII

Results of Simulated Market Activities for the Live Cattle Futures Market, 1991–96

Period	*Average Return (\$/cwt)*[a]	*Range of Returns (\$/cwt)*	*Variance of Returns*	*Variance of Termination Futures Price*	*MSE*	*N*
Dec '91–Oct '96	0.10	(−5.83, 6.55)	12.70	37.33	24.12	30
Dec '91–Apr '94	1.63	(−2.50, 6.55)	9.70	12.50	11.91	15
Jun '94–Oct '94	−1.54	(−5.83, 5.85)	11.32	17.77	36.32	15

[a]Average return per contract can be found by multiplying by 400. N represents the number of contracts used in the simulations under the trading rules outlined in the text.

therefore be possible to generate a positive average return. The strategy is to buy a futures contract if the forecast exceeds the futures price, and sell if the forecast is below the futures price. As in Leuthold et al. (1989), if the forecast for the next termination date is more than \$0.15 per hundredweight (the approximate commission costs) above (below) the relevant futures price, a futures contract is purchased (sold) at the closing price of the next trading day and held until termination. No trade occurs if the forecast is within \$0.15 of the futures price. The profit or loss of each trade is then recorded after deducting a charge of \$0.15 for commission.

Forecasts of the termination futures price are generated by the short-run OLS regression (4). The data series for live cattle have 87 observations from May 1983 to October 1996. The final 30 were used out-of-sample and the model was reestimated each period to allow incorporation of new data. Following Leuthold et al. (1989), the first observation in the data set is discarded as each new observation is added, thus keeping constant the number of observations in each estimation. According to Harvey (1981), this permits the forecasts of the model to respond more quickly to fundamental structural changes than when old observations are retained. Direct comparison of the forecast and the futures price is complicated by a positive maturity basis. It therefore makes sense to adjust the forecast by an expected basis. Again, as Leuthold et al. (1989), the methodology follows that of Holt and Brandt (1985). The expected basis for each contract is simply the average of the actual maturity basis for the most recent three years.

Table VIII shows the results of the simulated market activities over the period December 1991 to October 1996. The first result to note is that the average return from the trading strategy, over the full 30 forecasts, is positive (0.10 \$/cwt). This indicates that positive profits can be achieved in the live cattle futures market. However, the small size of the average return and the large variance suggest a substantial risk–return trade-off. Decomposing the forecast sample into two periods sheds further light on the result. From December 1991 to April 1994 a much larger positive average return (1.63 \$/cwt), accompanied by a smaller variance, exists than over the full sample. However, from June 1994 to October 1996, the reverse is true, with the simulation generating losses on average and increased variance of return. These results reflect the higher mean squared error (MSE) of the forecasting model over the latter period, which is itself caused by greater price volatility. June 1994 onward is clearly characterized by increased fluctuations in the futures price, as seen from the increased variance of the futures price in Table VIII.

The simulation results express the inability of the forecasting model to anticipate sudden shifts in price. Garcia et al. (1988) draw a similar conclusion when examining the live cattle futures market from 1983 to 1985. They comment that it is difficult to imagine a risk-averse trader using their models for trading signals. However, the positive average return confirms that the efficiency measure ϕ_c is of some economic value and, as the dichotomy in the simulation shows, this value becomes relatively more important to speculators in relatively low periods of price volatility.

CONCLUDING REMARKS

This article has examined futures markets for a number of widely traded commodities in terms of their efficiency, using a carefully selected and standardized data set. The evidence suggests that futures and spot prices are cointegrated with a long-run slope coefficient of unity; that is, the long-run equilibrium condition holds. However, in the short run there is evidence of inefficiencies in most of the markets studied. The simple observation that a market is not efficient provides limited information given the range of possible degrees of inefficiency. The costs of inefficiency (for both individuals and for society) will be greater with greater degrees of inefficiency; accordingly, the ability to measure the degree of inefficiency may be of considerable value and more informative than efficiency tests alone.

This paper proposes that the degree of inefficiency in a market may be measured in terms of the ability of the futures price to forecast the subsequent spot price relative to the forecast produced by the best fitting quasi-ECM. When this criterion of inefficiency is applied to a selection

APPENDIX

Contract Details

Commodity	*Spot Exchange*	*Futures Exchange*	*Contract Period*	f_{t-r} mean (st.dev.)	s_t mean (st.dev)	s_{t-r} (st. dev)	f_t mean (st.dev)
Brent crude	International Petroleum Exchange	International Petroleum Exchange	Monthly (12/89 to 12/96)	18.9$/bl (3.8)	19.2$/bl (4.2)	19.1$/bl (4.1)	19.2$/bl (4.1)
Gasoil	International Petroleum Exchange	International Petroleum Exchange	Monthly (01/90 to 12/96)	175.9 $/metric t. (34.8)	179.8 $/metric t. (38.2)	180.6 $/metric t. (39.5)	175.5 $/metric t. (34.4)
Soybeans	Chicago Board of Trade	Chicago Board of Trade	Two monthly (Jan, Mar, May, Jul, Sept, Nov) 12/79 to 11/96	637.2c/bu (104.1)	627.0c/bu (98.8)	624.8c/bu (100.9)	642.2c/bu (102.1)
Live hogs	Omaha, Nebraska	Chicago Mercantile Exchange	Two monthly (Feb, Apr, Jun, Aug, Oct, Dec) 05/82 to 10/96	48.7c/lb (6.5)	47.9c/lb (6.7)	47.9c/lb (7.3)	48.2c/lb (6.1)
Live cattle	Omaha, Nebraska	Chicago Mercantile Exchange	Two monthly (Feb, Apr, Jun, Aug, Oct, Dec) 05/82 to 10/96	68.7c/lb (6.9)	68.0c/lb (7.0)	68.2c/lb (7.2)	68.0c/lb (6.6)
Deutsch mark	New York	International Monetary Market	Quarterly (Mar, Jun, Sept, Dec) 05/76 to 12/96	2.01Dm/$ (0.44)	2.01Dm/$ (0.46)	2.01Dm/$ (0.45)	2.00Dm/$ (0.45)

of futures markets, with forecast horizons of 28 days, the findings suggest the soybean market is efficient, the gasoil market is 1% inefficient, the DM/$ market is 4% inefficient, the hogs market is 7% inefficient, the Brent crude market is 12% inefficient, and the cattle market is 47% inefficient. The extent of inefficiency is strongly correlated with the divergence, in statistical terms, of the estimated coefficient on the basis from unity in the quasi-ECM. These results are necessarily specific to the length of the forecast horizon. The analysis was repeated for four markets, extending this horizon to 56 days. The DM/$ market was then found to be efficient, the live hogs market 1% inefficient, the soybeans market 13% inefficient, and the live cattle market 23% inefficient. Our findings represent an extension of the work of Fama and French (1987); specifically, rather than simply testing whether the coefficient on the basis is equal to unity, the current research has explicitly examined the forecasting ability of the futures price and linked forecast power (and hence market efficiency) to the information contained in the basis. Noting that Fama and French define the basis as the sum of the expected change in spot price and the expected risk premium, it then seems likely that the forecast inefficiencies found in the markets studied can be ascribed largely to time-varying risk premiums. However, there is also some evidence, especially for cattle and hog markets, that lagged futures and spot prices influence current spot prices. Thus it may be that agents are unable to exploit arbitrage opportunities fully while they are learning about changes in market fundamentals.

BIBLIOGRAPHY

Beck, S. E. (1994). Cointegration and market efficiency in commodity futures markets. Applied Economics, 26, 249–257.

Chowdhury, A. R. (1991). Futures market efficiency: Evidence from cointegration tests. The Journal of Futures Markets 11, 577–589.

Crowder, W. J., & Hamed, A. (1993). A cointegration test for oil futures market efficiency. The Journal of Futures Markets, 13, 933–941.

Doornik, J. A., & Hendry, D. (1997). Modeling dynamic systems using PcFiml for Windows, London: ITP.

Fama, E., & French, K. (1987). Commodity futures prices: Some evidence on forecast Power, premiums and the theory of storage. Journal of Business, 60, 55–73.

Fortenbery, T. R., & Zapata, H. O. (1993). An examination of cointegration relations between futures and local grain markets. The Journal of Futures Markets, 13, 921–932.

Fujihara, R. A., & Mougoué, M. (1997). Linear dependence, nonlinear dependence and petroleum futures market efficiency. The Journal of Futures Markets, 17, 75–99.

Garcia, P., Leuthold, R. M., Fortenbery, T. R., & Sarassoro, G. F. (1988). Pricing efficiency in the live cattle futures market: Further interpretation and measurement. American Journal of Agricultural Economics, 70, 162–169.

Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. Oxford Bulletin of Economics and Statistics, 48, 213–228.

Hansen, L. P., & Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of futures spot rates. Journal of Political Economy, 88, 829–853.

Harvey, A. C. (1981). Time series models, Oxford: Philip Alan Publishers.

Holt, M. T., & Brandt, J. A. (1985). Combining price forecasting with hedging of hogs: An evaluation using alternative measures of risk. The Journal of Futures Markets, 5, 297–309.

Kaminsky, G., & Kumar, M. S. (1990). Efficiency in commodity futures markets. International Monetary Fund Staff Papers, 37, 670–699.

Krehbiel, T., & Adkins, L. C. (1993). Cointegration tests of the unbiasedness expectations hypothesis in metals markets. The Journal of Futures Markets, 13, 753–763.

Lai, K. S., & Lai, M. (1991). A cointegration test for market efficiency. The Journal of Futures Markets, 11, 567–575.

Leuthold, R. M., Garcia, P., Adam, B. D., & Park, W. I. (1989). An examination of the necessary and sufficient conditions for market efficiency: The case of hogs. Applied Economics, 21, 193–204.

Ma, C. W. (1989). Forecasting efficiency of energy futures prices. The Journal of Futures Markets, 9, 393–419.

Reimers, H. E. (1992). Comparison of tests for multivariate cointegration. Statistical Papers, 33, 335–359.

Serletis, A., & Scowcroft, D. (1991). Informational efficiency of commodity futures prices. Applied Financial Economics, 1, 185–192.

Schroeder, T. C., & Goodwin, B. K. (1991). Price discovery and cointegration for live hogs. The Journal of Futures Markets, 11, 685–696.

Stein, J. S. (1986). The economics of futures markets, Oxford: Basil Blackwell.

Shen, C. H., & Wang, L. R. (1990). Examining the validity of a test of futures market efficiency: A comment. The Journal of Futures Markets, 10, 195–196.

Taylor, M. P. (1995). The economics of exchange rates. Journal of Economic Literature, 33, 13–47.