Pricing in Futures and Forward Markets for Non-Ferrous Metals: An Empirical Critique

Clinton Watkins

Department of Economics
University of Western Australia

Abstract The analysis of commodity futures markets is an important field of empirical research in finance. A wealth of theoretical and empirical research has seen the development of models to analyse the properties of non-ferrous metals futures and forward prices. Several models use recent advances in the analysis of non-stationary data. Particular attention has been paid to efficiency, risk premia and price behaviour of contracts on the major futures and forward exchanges for non-ferrous metals. Using results based on both futures and forward price data, the performance of models specifically developed for analysing the pricing of futures contracts or forward contracts are compared. The purpose of the paper is to evaluate the significance of the empirical models published in leading futures market journals since 1990. Published empirical research is evaluated in the light of the type of contract examined, frequency of data used, choice of both dependent and explanatory variables, use of proxy variables, type of model chosen, economic hypotheses tested, methods of estimation and calculation of standard errors for inference, reported descriptive statistics, use of diagnostic tests of auxiliary assumptions, use of nested and non-nested tests, use of information criteria, and empirical implications for non-ferrous metals.

1. INTRODUCTION

Non-ferrous metals are traded in several futures and forward markets for speculative and hedging purposes. Commonly traded metals include: (i) the primary industrially-used metals: aluminium, copper, lead, nickel, tin, and zinc; and (ii) precious metals: gold, platinum, palladium, and silver. The major markets for metals futures contracts are the Commodity Exchange of New York (COMEX), Chicago Board of Trade (CBOT), and the New York Mercantile Exchange (NYMEX). The London Metal Exchange (LME) is the world's largest market for forward contracts in non-ferrous metals, and is also an exchange for spot transactions where physical delivery takes place.

Non-ferrous metals have several properties that are important in relation to modelling futures and forward prices. Metals are storable commodities, and are not subject to seasonal production. However, Fama and French [1988] find metals spot and forward prices have a strong business cycle component. Precious metals, in particular, are considered a store of wealth, and demand increases in periods of instability, anticipated high inflation and currency depreciation. Some authors differentiate their treatment of the precious and the main industrially-used metals, considering that fundamentally different forces affect the markets. However, there is no recent comprehensive analysis of this issue.

1.1 Economic Hypotheses Analysed in Recent Empirical Research

During the 1990's, most empirical work on metals forward and futures markets has examined issues related to market efficiency and unbiasedness. The models analysed follow from the empirical work of the 1980's, where evidence of market efficiency is mixed. Several definitions of market efficiency existing in the literature can lead to confusion.

Frances and Kofman [1991] test for flow parity relationships between forward prices for aluminium, copper, lead, nickel and zinc on the LME. One cointegrating relationship exists between the five metals, so that a long run relationship exists between the forward price series. If efficiency is defined such that a random walk is the best forecasting scheme, the LME is inefficient. Similarly, Agbeyegbe [1992] tests for common stochastic trends among copper, lead and zinc spot prices on the LME, and finds one relationship between the three metals, and a bivariate relationship between copper and lead. However, the author does not interpret cointegration as evidence of inefficiency.

Hsieh and Kulatilaka [1982] argue that markets for metals traded in the same location, such as the LME, warrant a wider view of efficiency. Empirical models may establish whether forecast errors across metal markets aid in predicting current forecast errors in other markets on the exchange. Using a framework for non-stationary data, Chowdhury [1991] interprets cointegration between spot and futures prices in the LME copper market as evidence of efficiency, as the variables will remain close together and not drift apart without bound. However, bivariate cointegration found between pairs of spot prices or pairs of futures prices means that the markets for copper, lead, tin and zinc are inefficient because information from one market will be useful in forecasting prices for the other.

Several recent papers estimate models based on the risk premium hypothesis. Hall [1991] analyses the risk premium in the LME copper, lead, tin and zinc markets using GARCH models, and provides a model supporting the existence of time varying risk premia. Melvin and Sultan [1990] examine COMEX gold futures prices. A time varying risk premium is estimated in terms of oil prices and political unrest in South Africa, two major factors commonly believed to affect the spot price of gold. In this analysis, the risk premium hypothesis is preferred over the cost-of-carry hypothesis since inventory convenience yields will be small for gold, and the costs of storage are relatively low compared to the value of the commodity. However, no empirical testing is undertaken to support this proposition.

A risk premium commensurate with the systematic risk of each contract is identified by Chang et al. [1990] for copper, platinum and silver futures prices using a model based on the Sharpe and Lintner Capital Asset Pricing Model (CAPM). The results are interpreted to support Keynesian normal backwardation in the context of the CAPM. Under normal backwardation, the forward price will, on average, be less than the expected spot price to generate a return to speculative activity.

Speculative efficiency, or the unbiased expectations hypothesis, implies the price of a futures contract is equal to the expected spot market price on the contract delivery date. Under the joint hypothesis of risk neutrality and rational expectations, an empirically testable form of speculative efficiency can be defined so that the futures price is an unbiased predictor of the realised future spot price. The definition has the advantage of not requiring a proxy variable or generated regressor for the expected spot market price. Moore and Cullen [1995] specify their cointegration model according to this definition and speculative efficiency cannot be rejected for the LME aluminium, copper, lead and zinc markets, and is rejected for the nickel market. The hypotheses of unbiased expectations and absence of a risk premium in COMEX copper, gold, and silver, and NYMEX platinum futures markets, are tested by Krehbiel and Adkins [1993] using a cointegration model. Absence of a risk premium is rejected for copper only, while the unbiased expectations hypothesis is rejected for every model except platinum. A joint test for both hypotheses is rejected for all markets, except copper.

Hill et al. [1991] test the efficiency of metals futures markets in relation to the "cold fusion" announcement of 1989. For a futures market to be generally price efficient with respect to all information, it must also prove efficient with respect to any given information set. Platinum and palladium futures markets are shown to be efficient with respect to the "cold fusion" announcement.

Sephton and Cochrane [1990, 1991] examine efficiency of the LME aluminium, copper, lead, nickel, tin, and zinc markets in terms of forecast errors, and in terms of the joint hypothesis of risk neutrality and rational expectations. Tests involving forecast errors aim to evaluate whether lagged forecast errors aid in predicting current forecast errors. The empirical analysis of forecast errors in single market models rejects efficiency for aluminium, lead, tin and zinc, while the copper and nickel markets are efficient. Multiple market models do not reject efficiency for aluminium and lead only. Using a methodology attributed to Fama [1984a, b], Sephton and Cochrane [1991] produce results contradicting previous published empirical papers using the same methodology. Efficiency is rejected for the copper and tin markets, where risk premia cannot be rejected. Stability tests show the results for all models, except for tin, are questionable. An alternative methodology, such as one based on cointegration, is required for valid inference regarding market efficiency.

Speculation is said to have a number of desirable features, including stabilisation of prices by elimination of arbitrage opportunities. Alternatively, the sophisticated speculator can exploit the naive forecasting rules of less sophisticated agents, thereby destabilising prices. Increased speculation in futures markets does not have a stabilising effect on spot prices, according to tests by Kocagil [1997] using COMEX aluminium, copper, gold, and silver contract data. Speculative trade, relative to non-speculative trade, cannot be observed, so a critical condition in terms of observable variables derived from a theoretical futures market model is tested. However, a futures market is, by its nature, a speculative market, and it is difficult to isolate some part of futures trade as speculative trade. It may be more sensible to consider the effects of increased or decreased trading volume on the stability of spot prices.

International linkages between the Shanghai Metal Exchange (SHME), operating under strict Chinese Government controls, and the LME are investigated by Shyy and Butcher [1994]. Spot and forward prices for copper on the SHME are cointegrated with the respective copper spot and forward prices on the LME, and it is claimed that the SHME prices coincide with those of the world market. Although not acknowledged by the authors, the analysis of markets between exchanges involves several problems that complicate the analysis, including exchange rate conversion and trading day differences. Trading on the SHME starts well before that of the LME, and one would expect that if the SHME is important with respect to world metal prices, information from the SHME trade would be accommodated by the participants of the LME exchange.

2. EMPIRICAL ISSUES

Thirteen papers published between 1990 and 1997 in futures market journals can be divided into 116 empirical econometric models relating to futures and forward markets for non-ferrous metals. The papers are by Agbeyegbe [1992], Chang et al. [1990], Chowdhury [1991], Franses and Kofman [1991], Hall [1991], Hill et al. [1991], Kocagil [1997], Krehbiel and Adkins [1993], Melvin and Sultan [1990], Moore and Cullen [1995], Sephton and Cochrane [1990, 1991], and Shyy and Butcher [1994].

2.1 Sample Data

Much of the econometric and statistical modelling has been conducted using monthly data. Table 1 shows the sampling frequency of the data used. Of the 116 models, 100 used monthly data, six used weekly data, and daily data were used for five models. Only one model used quarterly data, and annual data were used on four occasions. Small samples were most commonly used in the empirical models considered. Table 2 provides a breakdown of the sample sizes used.

All models use seasonally unadjusted data. The data series may contain seasonal fluctuations of a deterministic or stochastic nature, but no investigation of seasonality or seasonal unit roots has been conducted. If a futures or forward price series contain a seasonal pattern, modelling of seasonality will more accurately reflect the nature of the variable in the model. Determining the existence of a seasonal unit root in a futures price series is important for understanding how a shock will effect the series. For example, a shock to a futures price series with a seasonal unit root will have a permanent effect on the seasonal pattern of the series, rather than a permanent effect on the level of the series, as is expected if a zero frequency unit root is present.

Overlapping data are frequently encountered in the analysis of forward and futures markets, and occur when the sampling frequency is higher than the futures or forward contract period. The regression model will have serially correlated errors if data are overlapping, which can lead to biased and inconsistent estimates. Four approaches to the overlapping data problem are evident in the literature: select a sampling frequency that will avoid overlapping data, use a modelling procedure that accommodates serial correlation, use averaged data, or ignore the problem altogether. The last two approaches generate inadequate models. Gilbert [1986] shows that using averaged data does not avoid serial correlation. Averaged data are used by Chowdhury [1991] and Kocagil [1997]. Krehbiel and

Adkins [1993] choose to take equally-spaced non-overlapping observations of spot and futures prices to avoid problems of overlapping data. However, this choice of data set places limits on the analysis of futures and forward markets. For example, for 3-month futures contracts, the sampling frequency must be no more than quarterly to have non-overlapping data, but much higher frequency data are often available.

Care must be taken when compiling a data set from sources commonly used for empirical research in finance, such as newspapers. Sephton and Cochrane [1990, 1991] show that discrepancies in the reporting of forward and spot prices in these sources have resulted in inaccurate empirical analysis. In addition, Sephton and Cochrane [1990] demonstrate several analyses of LME market efficiency have not correctly matched forward and prompt prices. The prompt price may be defined as the spot price on the delivery date associated with the sampled forward price.

Of the 116 empirical models considered, 65 models relate to forward markets and 51 analyse futures markets (see Table 3). The theoretical difference between the pricing of futures and forward contracts is illustrated by Cox et al. [1981] using an arbitrage-based model. According to their model, the essential difference between futures and forward prices is related to the difference between holding a long-term bond and rolling over a series of one-day bonds, respectively. Thus futures and forward price models will not be equivalent unless the interest rate is non-stochastic.

LME forward contracts are frequently treated in the empirical literature in an identical manner to futures contracts. Moore and Cullen [1985], and Goss [1986], argue that, although the LME contracts are called forward contracts, they possess many of the properties of futures contracts. The LME contracts are standardised with regard to size, metal purity, and delivery location. There are arrangements for initial margins and margin calls, and there has been an organised clearing house (guaranteeing the contract) since 1987. In fact, Goss [1986] states that LME contracts are futures contracts in the sense in which the term is usually applied in the literature. Some points where the LME contracts differ significantly from futures contracts are that LME 3-month contracts are available on a daily basis, and not quoted for a limited set of dates per year, as is the case with futures contracts. For data prior to 1987, there was no clearing house and the LME was a principals market. Delivery frequently occurs under LME contracts, which is not the case for most futures contracts. In his exchange with Goss [1986] regarding the validity of previous analyses of LME efficiency, Gilbert [1986] highlights important differences between forward and futures contracts. Forward contracts nominate a day of delivery, while futures contracts state a delivery month. LME forward contracts are not routinely marked-to-market, so that the pricing of LME contracts will differ from futures market contracts with similar delivery dates.

2.2 Dependent Variable

Dependent variables used in futures and forward price models are listed in Table 4. Different measures of the spot price figured prominently among the dependent variables, with the natural logarithm of spot prices used in 39 models, the spot price in 11 models, and the change in spot price used six times. The forward price was used in 18 models, the natural logarithm of the forward price appeared as the dependent variable in 6 models, of which 4 used a lagged value. The futures price, detrended to remove either a stochastic or deterministic trend, appeared in 4 models. In 22 models, the risk premium was used as the dependent variable. Average contract return and daily return on futures price were the dependent variable for 7 models and one model, respectively. One model used the forecast error, defined as the difference between the forward price and the prompt price.

2.3 Explanatory Variables

The explanatory variables most frequently occurring in the models are measures of the natural logarithm of the futures or forward price used 19 times, the forecast error for 16 models, the forecast premium in 12 models, and the forward price on 10 occasions (see Table 5). Seven models contain the spot price, while the logarithm of the spot price is in 1 model, and also appears in an additional 4 models as a proxy variable. The difference between the current futures price and the current spot price, the futures basis, is an explanatory variable in 4 models. A measure of risk is generated from a GARCH model, and is used as a generated regressor in 1 model. For 2 models, the oil price is an explanatory variable.

Pricing models for futures and forward contracts frequently use proxy variables because variables specified by theoretical models are often unobservable. Variables used as proxies for unmeasurable variables are listed in Table 6. Market model regressions require a measure of excess return on the market, for which a market return index discounted by the US Treasury Bill rate was used in 39 models. The logarithm of realised spot price is used in 4 models as a proxy variable for the expected future spot price. Deaths, demonstrations and arrests are proxy variables for the degree of political unrest in South Africa. Other proxy variables include an adjusted lagged futures price in 4 models, the spot price discounted by the Treasury Bill rate, and an index of returns on futures contracts in 2 cases each.

A proxy variable in a regression necessarily implies the presence of measurement error since the correct variable is not used. Measurement error in an explanatory variable causes serious problems for ordinary least squares estimation, yielding biased and inconsistent estimators. One solution it to use instrumental variable estimation, which yields consistent but inefficient estimators. Frequently, a suitable instrument correlated with the proxy variable and uncorrelated with the error term is difficult to obtain. The implications of measurement error for estimation and inference do not appear to have been considered in the use of any of the proxy variables in Table 6.

2.4 Model Specification

Table 7 lists the frequency of each type of model used. The simple linear regression model is used 41 times, 12 times with MA(2) errors, and the multiple regression model is specified on 5 occasions. ARMA processes are used in 16 models, and 4 different types of GARCH models are specified a total of 12 times. Cointegration methods are used 27 times for bivariate models, and 2 times for multivariate models.

GARCH models are frequently applied to the modelling of (time-invariant) risk premia in financial data. Hall [1991] compares GARCH-M, SGARCH-M, and SGARCH-M with MA(2) errors in modelling the risk premium in metals forward prices on the LME. GARCH-M models provide a framework in which a time-varying risk premium can be estimated and tested. The standard ARCH or GARCH model is assumed to be non-stochastic. Hall [1991] argues that this is an extreme assumption in the context of empirical applications, and suggests the SGARCH-M model may be more appropriate. The stochastic generalisation of the GARCH-M model specifies the GARCH process as stochastic.

2.5 Methods of Estimation

Table 8 shows the methods of estimation for the 116 models. Ordinary least squares was used 41 times, and 4 times with Newey-West standard errors. Two stage least squares was used on 1 occasion to estimate a model with instruments. The method of estimation was not stated for 29 models, including ARMA, linear regression with MA(2), and GARCH, although the presumption is that maximum likelihood was used. The quasi-maximum likelihood method was used on 12 occasions. Maximum likelihood estimators of the SGARCH-M model are inconsistent, so Hall [1991] uses quasi-maximum likelihood estimation which provides consistent, but not fully efficient estimates.

The estimation methods applied to cointegration models were the Johansen Maximum Likelihood method for 6 models, the Engle-Granger method for 18 models, and Phillips-Hansen Fully Modified Ordinary Least Squares for 5 models.

The Johansen estimation procedure is not appropriate for the case of overlapping data [Moore and Cullen, 1995]. Overlapping spot and futures or forward price data generate moving average errors. When the errors in the cointegrating relationship are characterised by a non-invertible moving average process, the Granger Representation Theorem does not hold. The Phillips-Hansen fully modified OLS estimation procedure can deal with a wider class of serial correlation, which is an advantage for modelling with overlapping data. The Phillips-Hansen technique allows only one cointegrating vector at most. However, as Moore and Cullen [1995] deal with bivariate models only, this does not present a limitation. Estimates of parameters and standard errors are asymptotically equivalent to those produced by maximum likelihood estimation.

2.6 Descriptive Statistics

Descriptive statistics, as reported by the authors, are summarised in Table 9. The coefficient of multiple determination (R^2) was the most frequently reported descriptive statistic, and was used both to indicate the statistical adequacy of a model and to discriminate between models. Goodness of fit measures such as the coefficient of multiple determination and information (or discrimination) criteria, assess how well different models fit the data, with appropriate adjustments for parsimony. The philosophy in using discrimination criteria to choose between models is that the best predicting model is the closest approximation to the "true" specification. Each model is evaluated only in terms of its own performance, which is the principal disadvantage of discriminating between models on the basis of goodness of fit measures. One model will always be chosen, regardless of whether it can predict the consequences of non-nested alternatives.

Schwarz's Bayesian Information Criteria is used by Krehbiel and Adkins (1993) to select the optimal lag length required to conduct the Johansen test for the number of cointegrating relationships. The standard error of the regression, and standard errors due to White, or Newey and West, were also frequently provided. The value of the log-likelihood function is quoted 13 times for models estimated by maximum likelihood. For 24 models, no descriptive statistics were reported at all.

2.7 Diagnostic Testing

For two models, there were no diagnostic tests reported. As can be seen from Table 10, most of the 116 models reported few tests. Serial correlation was the most frequently tested auxiliary assumption for the empirical models. The Durbin-Watson test was conducted for 43 models. However, the test is somewhat limited, typically detecting only first-order serial correlation.

Chang et al. [1990] use OLS to estimate their 39 models for copper, platinum and silver, over the full data set and two sub-samples. A problem with first-order serial correlation for 10 silver contracts models is indicated by the Durbin-Watson statistic. The authors re-estimate these models using the Cochrane-Orcutt iterative process, which results in estimates that are not qualitatively different from their OLS estimates. For this reason, they present the original OLS models, ignoring the Durbin-Watson statistic. If higher-order serial correlation is present in a model, the Durbin-Watson test will identify the first-order component only. Therefore, the problem with the models in Chang et al. [1990] was possibly higher-order serial correlation, and hence it might be expected that the Cochrane-Orcutt method will not produce substantially different results. The Box-Pierce Q-statistic is used in Sephton and Cochrane [1990] to test for autocorrelation. Their model of market efficiency precludes third- or higher-order autocorrelation in the forecast error series for metal forward prices on the LME. A potential problem with the Box-Pierce Q-statistic is that it has very poor power in small- to medium-sized samples. The Ljung-Box test also has poor power, but is a superior test for smaller samples, and is used to test for serial correlation in 12 models. White's hetroscedasticity test and the Bera-Jarque test for normality are each conducted for 12 models, and the ARCH test is conducted for 2 models.

In general, spot, forward and futures prices for metals are found to be non-stationary and integrated of order one by the 35 models that test for the presence of unit roots. It should be noted that in the remaining 71 models, stationarity of the variables is simply assumed. Given the evidence for unit roots in metals spot, forward and futures price series, the relationships described by those 71 models may well be spurious.

No tests of stochastic seasonality are conducted, nor do any models include dummy variables for constant seasonal fluctuations. Thus, the potential problem of seasonality is ignored. There are also no tests for functional form mis-specification. There are 2 cases where Granger-Causality tests are used. For 6 models, joint F-tests for significance of the coefficients are presented. Four likelihood ratio tests for linear trends are reported.

Analysis of expected structural breaks is avoided by almost all authors. Few papers use methods to test for and analyse permanent or transitory structural change in the series. It is surprising that no papers attempt to model structural change in futures, and particularly the forward markets of the LME. It can be argued that if the behaviour of a variable during a period of structural change is not understood, it is not possible to understand the variable's behaviour before and after the structural shift. Thus, it is important to model explicitly structural breaks to determine whether the change is permanent or transitory, whether it is a one-off jump in the series, and whether the order of integration is changed.

Tests for structural change are conducted for 4 models, and 17 models test the stability of parameters. CUSUM of squares tests presented in Sephton and Cochrane [1991] provide evidence of structural change in the LME nickel, copper, aluminium, lead, and zinc markets, while a structural change could not be detected in the tin market.

During 1979-80, the Hunt Brothers attempted to manipulate prices in the silver market. Krehbiel and Adkins [1993] test whether their results for the silver market are sensitive to suspected structural change caused by the Hunt Brother's episode. Using the Perron test, the presence of a unit root is rejected for silver futures prices, so that the cointegration model may not be appropriate for analysis of the silver market.

The tin market collapsed in late-1985, and tin contracts were suspended on the LME from October 1985 to June 1989. An inter-governmental agency, the International Tin Council, dominated the tin market and traded in tin to stabilise prices. In trying to support the tin price, the International Tin Council's reserves were exhausted and, on 24 October 1985, the Council ceased operations with debts of over £900 million. Sephton and Cochrane [1990] delete tin and zinc from their third model since "consistent' series are unavailable for both metals on the LME after 1985. Moore and Cullen [1995] perform unit root tests on a sample from the tin market taken after the resumption of trading, and find that both the forward and spot prices are stationary. They attribute this to the structural change caused by the default of the International Tin Council. The nature of the market had changed, trading was thin, and the market was in cotango for most of the period since trading resumed.

Until the end of 1985, zinc contracts on the LME were classified as either zinc standard or zinc high grade. Subsequently, only high grade contracts were traded. Several empirical models, including those of Agbeyegbe [1992] and Hall [1991], use data for the Zinc market covering the period of the change without testing for structural change, despite using empirical methods whose results are sensitive to structural change. Sephton and Cochrane [1990] are aware of the change in contracts, and specify their sample up to September 1985, thereby avoiding the period of change. Some researchers appear to be unaware of changes in the contract specification for some data used in their empirical analysis, while others approach this problem by selecting a data set pertaining to one specification of the contract only.

2.8 Nested and Non-Nested Tests

Table 10 also shows the frequency of use of nested and non-nested tests relative to the 116 models examined. Four likelihood ratio tests of nested models are undertaken. No non-nested tests between competing separate alternatives are conducted.

3. CONCLUSION

Published empirical research has been evaluated in the light of the type of contract examined, frequency of data used, choice of both dependent and explanatory variables, use of proxy variables, type of model chosen, economic hypotheses tested, methods of estimation and calculation of standard errors for inference, reported descriptive statistics, use of diagnostic tests of auxiliary assumptions, use of nested and non-nested tests, use of information criteria, and empirical implications for non-ferrous metals.

Several conflicting empirical results with regard to futures and forward market models are apparent in the literature. Important empirical issues such as overlapping data, structural change, measurement error, correct use of proxy variables, and non-stationarity of spot futures and forward price series, have frequently been ignored. Diagnostic testing of the auxiliary assumptions is seldomly undertaken on a systematic basis, leaving open to question the statistical adequacy of the models presented. Most research does not consider nested or non-nested testing of competing models. Seen in this light, the empirical conclusions of the analysed research should be interpreted with caution.

4. ACKNOWLEDGMENTS

I would like to thank Michael McAleer for invaluable advice, numerous helpful comments and discussions.

5. REFERENCES

Agbeyegbe, T., Common stochastic trends: Evidence from the London Metal Exchange, *Bulletin of Economic Research*, 44, 141-151, 1992.

Chang, E. C., C. Chen, and S. N. Chen, Risk and return in copper, platinum, and silver futures, *Journal of Futures Markets*, 10, 29-39, 1990.

Chowdhury, A., Futures market efficiency: Evidence from cointegration tests, *Journal of Futures Markets*, 11, 577-589, 1991.

Cox, J., J. Ingersoll, and S. Ross, The relation between forward prices and futures prices, *Journal of Financial Economics*, 9, 321-346, 1981.

Fama, E., Forward and spot exchange rates, *Journal of Monetary Economics*, 14, 319-338, 1984a.

Fama, E., The information in the term structure, *Journal of Financial Economics*, 13, 509-528, 1984b.

Fama, E. and K. R. French, Business cycles and the behaviour of metals prices, *Journal of Finance*, 43, 1075-1093, 1988.

Franses, Phillip Hans and Paul Kofman, An empirical test for parities between metal prices at the LME, *Journal of Futures Markets*, 11, 729-736, 1991.

Gilbert, C. L., Testing the efficient markets hypothesis on averaged data, *Applied Economics*, 18, 1149-1166, 1986.

Goss, B. A., Rejection of unbiasedness is not rejection of market efficiency: Reply to Gilbert, *Applied Economics*, 18, 1167-1178, 1986.

Hall, S. G., An application of the stochastic GARCH-in-mean model to risk premia in the London Metal Exchange, *The Manchester School*, 59, 57-71, 1991.

Hill, S. R., N. H. Moore, and S. W. Pruitt, Cold fusion - hot metal: An analysis of the metals futures market reactions to the cold fusion announcement. *Journal of Futures Markets*, 11, 385-397, 1991.

Hsieh, D. A. and N. Kulatilaka, Rational expectations and risk premia in forward markets: Primary metals at the London Metal Exchange, *Journal of Finance*, 37, 1199-1207, 1982.

Kocagil, A. E., Does futures speculation stabilize spot prices? Evidence from metals markets, *Applied Financial Economics*, 7, 115-125, 1997.

Krehbiel, T. and L. C. Adkins, Cointegration tests of the unbiased expectations hypothesis in metals markets, *Journal of Futures Markets*, 13, 753-763, 1993.

MacDonald, R. and M. P. Taylor, Metal prices, efficiency and cointegration: Some evidence from the London Metal Exchange, *Bulletin of Economic Research*, 40, 235-239, 1988.

Melvin, M., and J. Sultan, South African political unrest, oil prices, and the time varying risk premium in the gold futures market, *Journal of Futures Markets*, 10, 103-111, 1990.

Moore, M. J., and U. Cullen, Speculative efficiency on the London Metal Exchange, *The Manchester School*, 63, 235-256, 1995.

Sephton, P. S., and D. K. Cochrane, A note on the efficiency of the London Metal Exchange, *Economics Letters*, 33, 341-345, 1990.

Sephton, P. S., and D. K. Cochrane, The efficiency of the London Metal Exchange: Another look at the evidence, *Applied Economics*, 23, 669-674, 1991.

Shyy, G. and B. Butcher, Price equilibrium and transmission in a controlled economy: A case study of the metal exchange in China, *Journal of Futures Markets*, 14, 877-890, 1994.

Table 1: Sampling Frequency of Data

Sampling Frequency	Frequency of Use
Daily	5
Weekly	6
Monthly	100
Quarterly	1
Annual	4

Table 2: Sample Sizes Used

Number of Observations	Frequency of Use
50 - 100	42
101 - 150	34
151 - 200	25
201 - 250	10
251 - 300	1
301 - 400	1
401 - 500	1
501 - 600	2

Table 4: Dependent Variables

Dependent Variable[1]	Frequency of use
Log of spot price	39
Risk premium	22
Forward price	18
Spot price	11
Average contract return	7
Change in spot price	6
Log of lagged forward price	4
Detrended futures price	4
Log of forward price	2
Daily return on futures price	1
Forecast Error	1

[1] Franses and Kofman [1991] do not specify a dependent variable in their analysis.

Table 9: Reported Descriptive Statistics

Regression Descriptive Statistics	Reporting Incidence
R^2	75
SE / White's SE / Newey-West SE	66
Log-Likelihood	13
Information Criteria	4
No Descriptive Statistics Reported	24

Table 3: Type of Market Analysed

Type of Market	Frequency of Analysis
Forward	65
Futures	51

Table 5: Choice of Explanatory Variable

Type of Explanatory Variable	Frequency
Log of forward / futures price	19
Forecast error	16
Forward premium	12
Forward price	10
Spot price	7
Futures basis	4
Oil Price	2
Log of spot price	1[1]
Risk Variable	1[2]

[1] Also appears in 4 models as a proxy variable.
[2] There is 1 case of a generated regressor.

Table 6: Use of Proxy Variables

Type of Proxy Variable	Frequency
Market return index discounted by T-bill rate	39
Adjusted lagged futures price	4
Log of spot price	4[1]
Spot price discounted by T-bill rate	2
Index of return on futures contracts	2
Deaths (in South Africa)	2
Demonstrations (in South Africa)	2
Arrests (in South Africa)	2

[1] Also appears in 1 model as an explanatory variable (non-proxy).

Table 7: Model Specification

Model Specification	Frequency
Simple Linear Regression	41
Simple Linear Regression with MA(2) Errors	12
Multiple Linear Regression	5
ARMA	16
GARCH	13
Bivariate Cointegration	27
Multivariate Cointegration	2

Table 8: Methods of Estimation

Methods of Estimation	Number of Models Estimated
OLS	41[1]
OLS with Newey-West SE	4
2SLS	1
Presume ML	29
Quasi-ML	12
Johansen ML Method	6
Engle-Granger Method	18
Phillips-Hansen FM-OLS	5

[1] Cochrane-Orcutt iterative procedure used to verify parameter estimates in 10 cases.

Table 10: Reported Diagnostics, and Nested and Non-Nested Tests.

Diagnostics and Nested / Non-Nested Tests	Reporting Incidence
No Diagnostics Reported	2
Serial Correlation: Durbin-Watson	43
Serial Correlation: Box-Pierce Q	16
Serial Correlation: Ljung-Box	12
Unit Root	35
Cointegration	24
Structural Change	4
Parameter Stability	17
Linear Trend	4
Seasonality	0
Mis-specification	0
Normality	12
Heteroscedasticity	12
ARCH	2
Causality	2
Joint Significance of Coefficients	6
Nested Tests	4
Non-Nested Tests	0