THE GOLD–SILVER SPREAD: INTEGRATION, COINTEGRATION, PREDICTABILITY, AND EX-ANTE ARBITRAGE

MAHMOUD WAHAB
RICHARD COHN
MALEK LASHGARI

INTRODUCTION

A central implication of the Efficient Markets Hypothesis [Fama (1970)] as it applies to Speculative Futures Markets is that individual futures prices, or a linear combination thereof, should evolve over time as a simple martingale process, if agents are risk neutral and markets are perfect. Alternatively, for small percentage price changes, the logarithmic first differences will be serially uncorrelated; with zero

We appreciate the helpful comments from two anonymous referees. Any remaining errors are our own.

■ *Mahmoud Wahab is an Assistant Professor of Finance at the University of Hartford.*

■ *Richard Cohn is the George Ansley Professor of Finance at the University of Hartford.*

■ *Malek Lashgari is a Professor of Finance at the University of Hartford.*

The Journal of Futures Markets, Vol. 14, No. 6, 709–756 (1994)

CCC 0270-7314/94/060709-48

drift.[1] As is well known, efficient markets tests presume a set of information that conditions expectations of subsequent prices. Many of these tests have been questioned on the ground that only a subset of the relevant information that market participants use to generate expectations is included in an econometric model. Accordingly, conclusive interpretation of anomalous evidence that seems to contradict the basic tenets of market efficiency has always been a controversial issue, for it may reflect (i) a flawed model, (ii) market inefficiency in which prices take temporary swings from their fundamental values, (iii) or markets are efficient, but the predictable variation may be explained by time-varying equilibrium returns. In the absence of an equilibrium pricing model that captures the inter-temporal behavior of the distributional moments, it is difficult to distinguish between these alternative competing hypotheses.

Market participants could efficiently employ at least two sets of information: (i) information contained in the history of each time series of prices (properly applicable to a weak-form test of market efficiency), and (ii) information that is contained in the history of each, as well as in the history of several other relevant series (suitable for testing a stronger-form of rationality of price expectations). Because information gathering and processing costs are non-negligible, economically rational agents would form optimal conditional forecasts (in the sense of conditional unbiasedness and efficiency) by employing broader information sets and more sophisticated information processing techniques only when the enhanced prediction benefits outweigh the costs.

In line with the above argument, the present study attempts to examine the relationship between gold and silver cash and futures markets using daily prices, with specific emphasis on the intertemporal behavior and predictability of the so-called gold–silver spread. A spread position involving the two metals is defined as a simultaneous long (cash or futures) position in one metal and a short (cash or futures) position of the same maturity in the other metal. A spreader will then accrue risk-unadjusted profits if the difference in prices changes in a favorable

[1]Under the stronger assumption that the first difference of log prices is serially independent and identically distributed, futures prices should obey a random walk process [Fama (1970)]. Furthermore, with risk aversion, positive trading cost, and a systematic component to speculation risk, futures prices should follow a submartingale (random walk with drift) process so that the change in log futures prices will be "approximately" uncorrelated, but with a positive mean drift. If expected returns to speculation (risk premium) are time varying, changes in log futures prices may be serially correlated due to serial correlation in expected returns. The induced serial correlation in realized returns due to possible time variation in the risk premium is, however, likely to be negligible if the variance of the risk premium is small compared to the variance of realized returns [French and Roll (1986)].

direction; that is, if profits on the long position accrue at a faster rate than on the short position, or that the net profit on the combined position is positive.

This article experiments with three alternative definitions of the gold–silver spread. In the first, a historical time series of the spread is generated from a log–log regression of gold on silver prices, and is referred to as the "unconstrained" gold–silver spread. In the second, the simple ratio of historical gold-to-silver prices is used as the spread series, and is referred to in this article as the "constrained" gold–silver spread. This ratio has traditionally been named the gold–silver parity [Ma (1985)], and is unlikely to be the best measure of the "true" price relation between the two precious metals, unless one is willing to assume, a priori, that the two metals are perfectly substitutable according to a factor of proportionality. In essence, defined as such, the parity, which is taken to be a proxy measure of the spread, would have a basic property, that of equi-proportionality between gold and silver prices (that is, that the long-run slope in a log–log regression is unity, and the intercept is zero). Because statistical inferences about these parameters are problematic if the levels representation of the parity contains a unit root, it is extremely difficult to verify whether those parameter restrictions are, indeed, valid. More specifically, a standard t-test is inappropriate because the unconditional variance of the parity is infinite under the null of a unit root. Therefore, it is readily apparent that the unconstrained gold–silver spread and the (constrained) gold–silver parity differ in that the latter is a special case of the former. Because of skepticism about the value of imposing such assumptions, alternative definitions of the spread are employed. Third, because a gold–silver spread series obtained from historical log–log regressions of gold on silver prices would not necessarily be identical to the reciprocal of a silver–gold spread series obtained from a reverse log–log regression of silver on gold prices, a third spread series is constructed and is referred to as the "unconstrained" silver–gold spread.

At issue in this article is whether an econometric model that is consistent with the empirical regularities in the data can be simulated forward on a periodic basis to generate predictions upon which a trading position in the precious metals spread can be established to secure an arbitrage profit. In principle, a test for arbitrage profits should be carried out ex ante. The reason is that what appears ex post as arbitrage profit opportunities is not necessarily a guaranteed ex-ante exploitable profit opportunity, for prices at the next available transaction may move unfavorably to the would be arbitrageur.[2] But arbitrage involving the gold–silver spread may involve less risk than speculating (risk arbitraging) on the future price of each metal taken separately. This would follow from the fact that the variance of price changes of a linear combination of long and short positions in the precious metals would be reduced as a result of positive correlation between the price of gold and the price of silver. The stronger the positive correlation, the lower the variance of the combined position.[3]

In this article, a no-arbitrage model of the fair theoretical value estimate of the gold–silver futures spread is presented. This model is a simple extension of the basic no-arbitrage cost-of-carry model for pricing futures (forward) contracts on storable commodities within a partial equilibrium framework. The model connects the futures level of the spread (or changes therein) with the cash market level (or changes therein). It implies that, in perfectly functioning markets, there should be complete simultaneity between the futures and cash spreads so that new information disseminating into the market place will be immediately reflected in the cash and futures prices of each metal (and accordingly into the cash and futures spreads). However, owing to lower costs of transacting in futures markets than cash markets, increased financial leverage opportunities, and ease of short selling, futures market participants may be situated to respond to news affecting precious metals prices in a more dynamic fashion. Because of this imperfect substitutability of cash and futures markets, trading by information (and noise) can be undertaken more frequently in the futures market than is possible in the cash market. In turn, futures prices may move first, followed by cash market prices. Indeed, these are some of the arguments provided to justify the importance of the price discovery role served by futures markets. This differential speed of response of the two markets would in turn give rise to a lead–lag structure that is predominantly influenced by the different institutional characteristics of both markets; or, alternatively, by the extent of differential market imperfections in both the cash and futures markets. A stronger lead from cash-to-futures is not, however, inconceivable, for cash prices (and their recent changes) represent part of the information set used by futures traders. Changes

[2]Of course, the joint nature of the maintained hypothesis of no arbitrage and a correctly specified model of the ex-ante spread should be always kept in mind.

[3]Although gold and silver may not be perfectly substitutable, with a pairwise price change correlation of +1, one would expect that they are, indeed, positively correlated. The estimates here indicate a pairwise correlation of about 0.45.

in cash prices may subsequently induce changes in futures prices, thus giving rise to a tendency of futures price changes to lag price changes in the cash market. These timing issues are fundamentally empirical.

A central feature of this study's attempt to model the gold–silver cash and futures spreads is a presumption that the conditioning information set used by economically rational traders to form optimal conditional forecasts of the alternatively defined gold–silver cash and futures spreads includes not only historical changes in the futures spread but also historical changes in the cash market spread. Furthermore, if the precious metals cash and futures markets are inter-linked by the forces of arbitrage, as they should be in a reasonably efficient market, any divergence between the actual and the theoretically fair value of the spread should exhibit mean-reverting behavior, which is additional information that is useful for forecasting purposes. This study should therefore provide results that extend the previously reported evidence on the potential for daily profitable arbitrage in the precious metals markets.

This study uses relatively new statistical techniques to investigate the intertemporal behavior and predictability of the gold–silver cash and futures spreads. Several methodological issues are addressed. The first issue concerns the integration properties of the data. More specifically, stationarity tests are conducted for each of the three spread series. Stationarity tests are of particular importance for at least two reasons: (i) the statistical theory underlying modern time series analysis relies on the assumption of covariance stationarity, an assumption that is, more often than not, found to be violated in many financial and economic time series; and (ii) stationarity is a requirement with important modeling implications when solving for expected values under rational expectations. A failure to account for stochastic trends in the data risks misspecifying the model, can seriously affect the estimated dynamics, and has critical implications for forecasting.[4]

A second issue, which is closely related to the first, is whether the gold–silver cash and futures spreads are cointegrated, which is addressed using techniques that are now familiar from the theory of cointegrated processes of Engle and Granger (1987). This is important because, if the two spreads are cointegrated, there is a statistical payoff from employing a broader information set that includes not

[4]Furthermore, failure to account for unit roots, if they are present in the data, invalidates conventional statistical inferences and hypotheses tests because Granger and Newbold's (1974) well documented spurious regression phenomenon is obtained.

only historical realizations of the cash and futures spreads but also, potentially, the historical mean-reverting tendencies created by the forces of arbitrage in the precious metals cash and futures markets. Although each series taken individually may be a martingale sequence, any pair of cointegrated series may be so related that the component series will not drift infinitely apart over time, and a linear combination may prove to be forecastable. Consequently, a finding of cointegration among the cash and futures spreads would have important consequences for the proper specification and statistical testing procedures to employ when modeling the cash and futures spread series. When a forecasting model is used to project such time series, a vector autoregressive model in differences is inappropriate because this model will suffer misspecification, any potentially important long-run relation between the variables will be obscured, and forecasts of the component series will diverge. If cointegration is admitted, the so-called vector error correction formulation can be estimated for the cash and futures spread and iterated forward on a periodic basis to generate multistep forecasts of the changes in the spreads, using all past data prior to each forecast origin.

A third issue pertains to the temporal behavior of the forecasting model's parameters. Since forecasts are made conditional on the estimated parameters, the precision of the forecasts will be governed in part by the extent to which the postulated behavior of the forecasting model's parameters coincides with the observed movements in these parameters. Since there is no theoretical rationale to presume that market participants' reaction functions are temporally fixed, a forecasting model with time varying parameters (TVP) is also estimated, and the accuracy of its forecasts is compared with that of forecasts generated by a fixed coefficient (FC) forecast model. The performance of the arbitrage strategy is then compared for both models.

It is found that trading results are likely to be sensitive to the particular definition of the spread. Results from a variety of statistical tests suggest treating the logarithm of each spread series as the sum of a permanent component and a stationary mean reverting transitory predictable component within an error correction framework. Trading profits that are positive on a pre-transactions cost basis become losses on a post-transactions cost basis if the assumed level of transactions costs in the cash and futures markets surpasses a breakeven level of about \$7 on the combined gold and silver positions (corresponding to an average percentage cost of around 0.08% of the in-sample mean of gold and silver prices). Trading results are obtained from iteratively reestimating appropriately specified error-correction models over a hold-out trading sample of about four years in length. To facilitate reasonable comparability with earlier studies, a variety of simple moving average trading rules are also introduced. Moving average trading results indicate that if the moving average rules are predicated on the "levels" of the spread series (constrained and unconstrained, and with and without a neutral band), extreme losses are realized over the hold-out sample, a finding that is hardly surprising given the non-stationary character of the empirical distributions of the levels of the spread. Additionally, if the moving average rules are applied to the first differenced series of the constrained spread, and regardless of the width of the neutral band, extreme losses are still realized, a result which reflects on the repercussions of imposing potentially spurious coefficient restrictions on the bivariate distribution of gold and silver (cash and futures) prices. On the other hand, employing the moving average rule on the differenced series of the unconstrained spread yields profits for the wider- but not the narrower-band alternatives of the trading rule. Furthermore, these profits appear to be "excess" profits in the sense that they do not merely represent compensation for systematic risk inherent in the spread position. In fact, any risk seems to be totally diversifiable within a stock index portfolio context. However, transactions costs should be carefully considered before such strategies can be implemented.

The remainder of the article is organized as follows. First, a brief review of related research is presented, followed by a model linking the futures and cash gold–silver spreads under the assumptions of no transactions costs and no taxes. Then the article briefly reviews the integration and cointegration tests used in the study, and pursues the formulation of an appropriate forecasting model, and then describes the data, the forecasting strategy, and the simulated trading strategy. Finally, the results are presented, and a summary of the major findings.

REVIEW OF PREVIOUS STUDIES

Several earlier studies have examined the profitability of arbitrage trading involving the precious metals. Ma (1985) examined the potential for arbitrage profits from spreading between gold and silver cash prices between 1972 and 1984. Using daily London afternoon fixing price quotes for gold and silver, the author assumed a stationary autoregressive process of order (m) for the gold–silver cash parity (or constrained spread) which was inverted to its moving average representation. Using alternative lag specifications of a moving average model, an ex-ante equilibrium parity was generated and compared to the current actual parity level. As long as the actual parity fell within some postulated confidence bounds around the predicted parity, no trading action was triggered. If the current parity penetrated either the upper or the lower bound, an appropriate trading action that involved buying the underpriced metal and selling the overpriced metal was triggered. By perturbing the width of the confidence interval (number of standard deviations around the ex-ante parity) and the moving average order, Ma detected ex-ante arbitrage profit opportunities, which were then interpreted as evidence of weak-form inefficiency. Ma and Soenen (1988) replicated the earlier study using daily COMEX futures prices for gold and silver from 1976–1986. Significant arbitrage profits were found from spreading the gold–silver futures parity on a daily basis.

Monroe and Cohn (1986) examined the pricing efficiency of the gold futures market relative to the Treasury bill futures market by formulating an intracommodity spread in gold futures coupled with a position in T-bill futures. To the extent that prices for different delivery dates on the gold contract should reflect some interest rate that is closely related to the T-bill rate (as a measure of the opportunity cost of storing gold), gold futures prices should exhibit a well-defined relation with T-bill futures prices in equilibrium. The authors demonstrated that potential arbitrage profits could be made by using trading rules that exploit observed disequilibria in the gold and T-bill futures markets.

A related study, although it does not directly address the gold–silver spread trading issue, is that of Chan and Mountain (1988). The authors examined the relationship between the gold and silver cash markets using weekly prices for gold and silver quoted in Toronto from 1980–1983. Assuming an adaptive expectations framework, they estimated an unconstrained vector autoregressive model to identify causality and predictability of gold and silver cash prices. Their results indicate not only a simultaneous relation between price changes in the gold and silver cash markets but also that a feedback causal relationship reliably existed, implying some forecastability of subsequent precious metals prices.

A study that examined the temporal behavior of price movements in the spot and futures markets for gold and silver, as well as other storable commodities, is that of Garbade and Silber (1982). They proposed a simultaneous dynamic model in the spirit of a bivariate vector autoregressive model in which the interrelationship between spot and futures prices depends on the elasticity of supply of arbitrage capital. The supply elasticity of arbitrage capital was in turn presumed to be constrained by market imperfections. When the authors compared COMEX futures opening prices for gold and silver with their cash market counterparts, the cash markets appeared to be the locus of price discovery. On the other hand, when COMEX closing prices were compared with cash market prices the reverse was observed.

THE MODEL

The Theoretical Price Relation

In the absence of transactions costs and taxes, a no-arbitrage relation linking futures to cash prices of storable commodities is the familiar cost of carry model.[5] This model, when customized to price futures contracts on gold and silver, yields:

$$F_{g,t} \leq C_{g,t} \exp[g(T - t)] \quad \text{and} \quad F_{s,t} \leq C_{s,t} \exp[s(T - t)]$$

where $F_{g,t}$ and $C_{g,t}$ are, respectively, the futures and cash prices for gold quoted at t, and $F_{s,t}$ and $C_{s,t}$ are, similarly, futures and cash prices for silver quoted at t. The exponents g and s are, respectively, continuously compounded net marginal carrying costs of storing gold and silver from the present, t, to some future delivery date, T, so that $T - t$ is the time remaining to expiration of each futures contract. The net marginal costs of carry, g and s, consist of the following components under the assumption that each component is paid out at a known and constant continuously compounded rate: (i) a marginal outlay for storage and insurance, (ii) a marginal interest cost reflecting the opportunity cost of owning a non–interest-bearing stock of gold or silver rather than alternative interest bearing assets, and (iii) a marginal convenience yield representing the monetary value accruing from, for example, avoiding costly changes in production schedules and potential loss of sales arising from possible shortages of the precious metal on hand.[6] In such frictionless markets, arbitrage profits arise whenever

[5]As argued by Garbade and Silber (1982), the cost of carry model of equilibrium (no arbitrage) futures prices is much more appropriate for pricing deferred futures contracts on precious metals, in comparison to other storable commodities, for a variety of reasons. First, transactions and storage costs for precious metals are relatively low when compared to other commodities. Second, precious metals can be easily sold short. Third, there is much less heterogeneity in the deliverable grade that can be used to satisfy a short futures position.

[6]It is generally assumed for precious metals that the marginal cost of storage and insurance and the marginal convenience yield are offsetting, so that their net impact would be negligible compared to the marginal interest cost. Furthermore, foregone interest costs on the initial margin deposit may be assumed negligible because initial margin may be posted in the form of Treasury bills or other high grade marginable securities instead of cash, if the value of the deposited securities exceeds a certain minimum.

actual futures prices diverge from their theoretical fair value estimates. Thus, there is a tendency for the two prices to stay close together. With transactions costs, however, futures prices will fluctuate within a band around the theoretical value without representing a potential profit opportunity.[7]

Taking the natural logarithm of the random variables in the above equations yields:

$$\ln F_{g,t} \leq \ln C_{g,t} + g(T - t) \tag{1}$$

and

$$\ln F_{s,t} \leq \ln C_{s,t} + s(T - t) \tag{2}$$

Define a new variable $P_{f,t} = \ln F_{g,t} - \theta_0 - \theta_1 \ln F_{s,t}$ where $P_{f,t}$ might be termed the gold–silver futures spread (or futures parity), where θ_0 represents the long run average value of gold futures prices, and θ_1 is some proportionality coefficient that is not necessarily constrained to equal unity (so that the two metals need not be regarded as perfectly substitutable). The spread variable, which is written as an unconstrained linear combination of the two futures prices, is obtained by subtracting $\theta_0 + \theta_1 \ln F_{s,t}$ from both sides of eq. (1) so that,

$$\begin{aligned} P_{f,t} = \ln F_{g,t} - (\theta_0 + \theta_1 \ln F_{s,t}) &\leq \ln C_{g,t} - (\theta_0 + \theta_1 \ln F_{s,t}) \\ &+ g(T - t) \end{aligned} \tag{3}$$

Substituting eq. (2) into the RHS of eq. (3), one obtains:

$$\begin{aligned} P_{f,t} &\leq \ln C_{g,t} - \{\theta_0 + \theta_1(\ln C_{s,t} + s(T - t)) + g(T - t)\} \\ &\leq \ln C_{g,t} - (\theta_0 + \theta_1 \ln C_{s,t}) + (g - \theta_1 s)(T - t) \end{aligned} \tag{4}$$

or

$$P_{f,t} \leq P_{c,t} + (g - \theta_1 s)(T - t) \tag{4a}$$

so that $P_{f,t} - P_{c,t} - (g - \theta_1 s)(T - t) \leq 0$ must hold to prevent arbitrage. Eq. (4a) may be regarded as a simple no-arbitrage model linking the futures gold–silver spread ($P_{f,t}$) to the cash spread ($P_{c,t}$). It values the futures spread as a forward spread, thereby ignoring the possible

[7]The width of the band derives from round-trip commissions in the cash and futures markets and also from any impact of the transaction on market prices.

pricing consequences of daily settlement risk on the futures price in a stochastic interest rate setting. The link in eq. (4a) will be enforced by the actions of arbitrageurs. For, if the actual futures spread deviates from the theoretical futures spread by an amount exceeding transactions costs (in the absence of any risk premium), arbitrage will occur.

The Theoretical Change-in-the-Spread Relation

A clear implication of the theoretical relation in eq. (4a) is that causation is not indicated, and that only contemporaneous price adjustments in both markets are possible. In other words, conditional on eq. (4a) being correct, exclusion restrictions on lead and lagged values of the cash and futures spreads should hold at stringent significance levels when an empirical version of eq. (4a) is estimated.

The theoretical change in the spread can be similarly derived by lagging eq. (4a) one period to yield eq. (5), then subtracting eq. (5) from eq. (4a), to yield eq. (6):

$$P_{f,t-1} = P_{c,t-1} + (g - \theta_1 s)(T - t + 1) \tag{5}$$

and

$$\Delta P_{f,t} = \Delta P_{c,t} - (g - \theta_1 s) \tag{6}$$

where Δ refers to the first difference; $P_{f,t}$ and $P_{c,t}$ are, respectively, the futures and cash gold–silver spreads; and the last term denotes the differential cost of carry weighted by the proportionality scalar θ. Some of the implications of eq. (6) are: (i) the change in the futures spread is contemporaneously perfectly positively correlated with the change in the cash spread; (ii) the expected rate of change in the cash spread is equal to the expected rate of change in the futures spread plus the weighted differential cost of carry; and (iii) the standard deviation of the change in the futures spread equals that of the change in the cash spread (under a constant differential cost of carry assumption). Given market imperfections, however, the observed spread-change relation will be noisy. Furthermore, if there are cost incentives for economic agents to prefer transacting in one market over the other, a lead–lag relation between changes in the futures-cash spreads is likely.

INTEGRATION AND COINTEGRATION TESTS

It is fairly well established that the presence of a unit root in the levels representation of, for example, the gold–silver cash (or futures) spread series constitutes a necessary (but not sufficient) condition for the unpredictability of either spread series. If the additional condition that increments of the spread are serially uncorrelated with zero mean and constant variance is verified, the change in the spread is unforecastable, so that either spread series is a martingale sequence (or a sub-martingale in the presence of a positive drift). On the other hand, serially correlated variations in the spread may imply predictability, and this predictable variation may be modeled by many candidate stochastic processes, for example, the Box–Jenkins univariate ARIMA class of models, or the more complex multivariate transfer function models. Such models do fairly well in a short- rather than a long-term forecasting context. Therefore, long- and short-term forecasts would be accommodated usually by constructing different models covering different forecast horizons. Although there would be a single data generating process, the long- and short-term models can be thought of as approximating different parts of this generating process.

An alternative modeling procedure is to resort to the theory of cointegrated processes; employing an error correction framework that combines the short- and long-term models by incorporating, simultaneously, short- and long-term dynamics. Therefore, forecasts that impose the cointegrating restriction should outperform forecasts that ignore this information. Since a necessary (but not sufficient) condition for two series to be cointegrated is that they be integrated of the same order, cointegration tests proceed after tests for a common order of integration. Furthermore, because a natural cointegrating partner to the cash (futures) gold–silver spread is the futures (cash) spread, predictability of the cash and futures spreads may be enhanced by endogenizing both series within an error-correcting system if cointegration is admitted.

From a purely empirical view, observed high correlations among the levels of financial or economic time series are characteristic of data that potentially could be cointegrated, even if each of the series is integrated of order 1, denoted $I(1)$. This discussion is based on Engle and Granger (1987) and Granger (1988).

Two sequences of random variables are said to be cointegrated if: (i) each is non-stationary in the level representation (i.e., each contains a unit root or a stochastic trend), (ii) each is stationary in first differences so that each series is $I(1)$, and (iii) there exits a linear combination of the two variables that is stationary or integrated of order zero. The economic interpretation of such a relationship is that the two variables tend to wander over time without returning to a constant mean, but economic forces do not allow the series to wander apart permanently.

Usual tests for cointegration proceed in two steps. First, the individual series are tested for a common order of integration. If the series are integrated of the same order, pairwise linear combinations are then tested for stationarity (cointegration).

Univariate Tests for Unit Roots

This section is concerned with the presence of a unit root in the logarithmic levels representation of each of the gold–silver cash and futures spreads. This discussion is motivated by considering the following autoregressive model of the gold–silver cash (futures) spread:

$$P_t^i - \mu = \rho(P_{t-1}^i - \mu) + e_t \qquad i = 1, 2 \tag{7}$$

with 1 denoting the cash spread, and 2 denoting the futures spread. Both spreads are expressed in natural logarithms at time t; μ is the long-run value of the spread, and $e_t \sim N(0, \sigma^2)$ is the error term and is independent of past values of P_t^i. In this model, the long-run behavior of the spread is critically dependent on the value of the autoregressive coefficient ρ. If $0 < \rho < 1$, the spread is stable so that any deviation $(P_t^i - \mu)$ from the long-run value would shrink over subsequent periods.

On the other hand, in the presence of a unit root in the logarithmic level representation of the spread, the autoregressive coefficient ρ equals unity, and deviations therefore have no tendency to shrink so that, in effect, a stable gold–silver spread does not exist.[8]

To avoid imposing a priori spurious coefficient restrictions when estimating the level of the spread, the following log–log regressions are estimated:[9]

$$\ln F_{g,t} = \alpha + \beta \ln F_{s,t} + P_{f,t} \quad \text{and}$$

$$\ln C_{g,t} = \alpha' + \beta' \ln C_{s,t} + P_{c,t} \tag{8}$$

so that $P_{f,t} = \ln F_{g,t} - (\alpha + \beta \ln F_{s,t})$ and $P_{c,t} = \ln C_{g,t} - (\alpha' + \beta' \ln C_{s,t})$ where $P_{f,t}$ and $P_{c,t}$ denote, respectively, the historical futures and cash levels of the unconstrained gold–silver spread, and

[8]Also note that a direct implication of a unit root in the level representation of the spread is that gold and silver price sequences are noncointegrated.

[9]It is readily seen that the set of equations in (8) provides the logarithmic levels of futures (and cash) spreads, respectively.

other variables are as defined earlier. This definition of the spread is to be contrasted with the alternative "parity" definition, which constrains the intercept and proportionality coefficient between the two metals prices. Finally, because gold and silver cash (and futures) prices are treated as being jointly endogenous, and because a least squares fit of a reverse regression will not simply give the reciprocal of the coefficient in a forward regression, the unconstrained cash and futures series of the silver–gold spread are estimated by reversing the LHS–RHS variables designations and reestimating the set of eqs. (8) in log–log form. Another way to put it is that if gold and silver cash (or futures) prices do not exhibit similar unconditional volatilities, the slope coefficients in the forward and reverse regressions will not equalize.

Although the integration and cointegration properties of the forward and reverse definitions of the cash (and futures) spreads should remain unchanged, outcomes of the trading strategy may not. Because a published series of the spread is not readily available to market participants, a gold–silver spread is not uniquely defined.[10] Furthermore, constraining the long-run proportionality coefficient on relative prices to one, when the true value is not one, is a classical problem of measurement error that jeopardizes forecast quality, in the sense of conditional unbiasedness and efficiency. Therefore, another way to view the unconstrained definition of the gold–silver cash (and futures) spreads given in the set of eqs. (8) is that least squares estimation, as a data transformation procedure, is applied to the gold and silver cash (and futures) price series. Defined as such, changes in the value of the gold–silver spread are nothing but changes in the price of gold net of scaled changes in the price of silver. Since the trading strategy

[10]Consider, for example, that the current gold cash (or futures) price is at $600/oz., while the current silver price is at $6/oz. If one denotes the constrained spread as the value of the gold–silver spread under a joint constraint imposed on the intercept and the proportionality coefficient in the set of eqs. (8), while the unconstrained spread is denoted as the value of the spread in the absence of such a constraint; then, for example, the following two logarithmic futures spreads are possible: (i) constrained $(\alpha, \beta) = (0, 1)$: $P = \ln F_{g,t} - (\alpha + \beta \ln F_{s,t}) = 4.605$ and (ii) unconstrained $(\alpha, \beta) = (2, 0.2)$: $P = \ln F_g - (\alpha + \beta \ln F_{s,t}) = 4.038$ using actual rather than a-priori constrained parameter values. Taking antilogs, the respective values for the constrained and unconstrained spreads are 100 and 56.74. The β proportionality coefficient in this case may be referred to as a cross-hedge elasticity coefficient. Now, assume that next period's gold price is $605/oz., while next period's silver price is $6.10/oz. (reflecting a percentage appreciation for gold of 0.83%, and an appreciation rate for silver of 1.67%, and knowing that the rate of price change for silver is usually more volatile than that for gold). If the same parameters are imposed at time 1 as those at time 0, the following two new values for the constrained and unconstrained spreads are obtained: (i) constrained spread (antilog) is 99.18, for a percentage change of −0.82%; and (ii) unconstrained spread (antilog) is 57.03, for a percentage change of 0.51%. Because this trading strategy is predicated on predictable changes in the spread, alternative definitions of the spread could potentially trigger different buy/sell signals for the precious metals.

adopted in this article is based entirely on "net" predictable changes in the spread, and because there is no assurance that the alternative definitions of the gold–silver spread would yield identical buy–sell signals on a day-to-day basis, there is a need to examine the potential for arbitrage profits predicated on each of the three spread series.

The null hypothesis tested in eq. (7) is that $\rho = 1$, which is not testable using conventional statistical inference techniques.[11] Thus, unit root tests based upon the methods of Fuller (1976) and Dickey and Fuller (1979) are conducted. The Augmented Dickey Fuller (ADF) test is used on the following regression:[12]

$$\Delta P_t^i = \alpha_0 + \phi t - \beta P_{t-1}^i + \sum_{j=1}^{m} \alpha_j \Delta P_{t-j}^i + \eta_t \tag{9}$$

where $\Delta P_t^i = P_t^i - P_{t-1}^i$ and $\Delta P_{t-1}^i = P_{t-1}^i - P_{t-j-1}^i$, m is the number of lags of ΔP_t^i which is arbitrarily set to a maximum order of 4, and t is a linear time trend. The empirical percentiles of the statistic for testing H_0: $\beta = 0$ are tabulated in Fuller [(1976), Table 8.5.2, p. 343], and are improved upon by MacKinnon (1991) to yield more precise critical values for any finite sample size. MacKinnon's critical values are used in this study.[13]

Tests for Cointegration

To test for cointegration of the cash and futures gold–silver or silver–gold spreads, the so-called cointegrating regression linking the levels of the spreads is estimated as:

$$P_{f,t} = a + bP_{c,t} + z_{f,t} \tag{10}$$

$$P_{c,t} = a' + b'P_{f,t} + z'_{c,t} \tag{10a}$$

[11] It is well known that the usual t-test is inappropriate here because under the null hypothesis of a unit root, the variance of the spread is infinite.

[12] The Augmented Dickey Fuller's (ADF) test is used to account for possible temporally dependent and heterogeneously distributed errors. This procedure is asymptotically equivalent to the Phillips and Perron (1988) test, which also accounts for the non-independent and identically distributed (niid) processes. Furthermore, a trend term is included in the model to allow the alternative hypothesis to be trend-stationarity. Note that if m (lag order) is constrained to zero, this is simply the Dickey–Fuller test.

[13] Fuller's tabulated critical values are based on at most 10,000 replications of Monte Carlo simulations, and they are provided for a limited number of finite sample sizes. On the other hand, MacKinnon's critical values are based on 25,000 replications using alternative sample sizes. Asymptotic critical values corresponding to any finite sample size can be computed easily using the tabulated coefficients of the estimated response surface regressions at the 1%, 5% and 10% levels. The estimated critical values are given by: $\beta_\infty + \beta_1 T^{-1} + \beta_2 T^{-2}$ where β_∞ denotes estimated asymptotic critical value, and β_1 and β_2 are coefficients in the response surface regression which adjust for sample size.

and the residuals $z_{f,t}$ and $z'_{c,t}$ are recovered and are subjected to the usual unit root tests using the ADF-testing equation as shown in eq. (9). Cointegrating regressions represent the long-term equilibrium relationship between the variables in the regressions while the cointegrating error terms $z_{f,t}$ and $z'_{c,t}$ are measures of disequilibrium.[14] Rejecting the null hypothesis of a unit root in these residuals implies that the variables in the regression are cointegrated. If the residuals in eqs. (10) and (10a) are stationary, then by Granger's representation theorem [Engle and Granger (1987)] the following error correction representations are estimated; with a constant included to capture any secular drift in either series:

$$\Delta P_{f,t} = \gamma_0 + \gamma_1(z_{f,t-1}) + \sum_{i=1}^{m} \beta_{i,f}\Delta P_{f,t-i} + \sum_{j=0}^{m'} \beta'_{j,c}\Delta P_{c,t-j} + e_{f,t} \tag{11}$$

$$\Delta P_{c,t} = \gamma'_0 + \gamma'_1(z_{f,t-1}) + \sum_{j=1}^{n} \alpha_{j,c}\Delta P_{c,t-j} + \sum_{i=0}^{n'} \alpha'_{i,f}\Delta P_{f,t-i} + e_{c,t} \tag{11a}$$

$$\text{where } z_{f,t} = P_{f,t} - (a + bP_{c,t}) \tag{11b}$$

and

$$\Delta P_{f,t} = \alpha_0 + \gamma_0(z'_{c,t-1}) + \sum_{i=1}^{p} \lambda_{i,f}\Delta P_{f,t-i} + \sum_{j=0}^{p'} \lambda'_{j,c}\Delta P_{c,t-j} + e'_{f,t} \tag{12}$$

$$\Delta P_{c,t} = \alpha'_0 + \gamma'_1(z'_{c,t-1}) + \sum_{j=1}^{q} \delta_{j,c}\Delta P_{c,t-j} + \sum_{i=0}^{q'} \delta'_{i,f}\Delta P_{f,t-i} + e'_{c,t} \tag{12a}$$

$$\text{where } z'_{c,t} = P_{c,t} - (a' + b'P_{f,t}) \tag{12b}$$

[14]The need to estimate two cointegrating regressions instead of one stems from the non-uniqueness of the parameters. Essentially, a least squares fit of a reverse cointegrating regression will not simply give the reciprocal of the coefficient in a forward cointegrating regression. This is so even though a least squares fit provides a consistent estimate of the cointegrating parameter [see Engle and Granger (1987)].

The error correction mechanism implies a long-run linear relationship between the futures and the cash spread levels. Past deviations from the long-run equilibrium values of the respective spreads lead to current partial corrections, which are useful in generating ex-ante predicted changes in the spreads. Furthermore, because historical values of the cash (and futures) spreads appear in the error correction models together with past equilibrium error terms, the predicting error correction models employ a broader information set that encompasses not only historical cash prices of both metals, but also historical futures prices. Therefore, improved forecast performance in the form of lower mean square prediction errors of the dependent variable is likely. Finally, the unconstrained definition of the spread imposes the asymptotically "true" proportionality coefficient on the predicting error correction equation so that the final error correcting models are based on consistent estimates of this coefficient.

DATA, FORECAST STRATEGY, AND SIMULATED TRADING STRATEGY

The data are (i) daily gold and silver cash prices covering the period between January 1982, and July 1992, as supplied by Handy and Harman in New York; and (ii) daily gold and silver futures settlement prices of nearby-expiration gold and silver futures contracts traded on the COMEX, as reported in the *Wall Street Journal* over the same period.[15]

To examine the potential for profitable ex-ante arbitrage opportunities in the cash and futures gold–silver spreads, conditional forecasts of the spreads are required which raises several issues of concern. The first is the determination of a suitable forecast horizon, which is taken to be one week long. A decision needs to be made about how often the forecasting model should be reestimated to update the parameters. A daily forecast horizon seems unduly restrictive, computationally burdensome, and too expensive to typify the practice of a trader who is likely to estimate a forecasting model only once over a suitably defined forecast period. On the other hand, a longer forecast horizon, such as one month, may be too optimistic about the stability of a trader's reaction function to changes in the variables viewed as influencing the spread.

Second, one must decide whether to use multistep- or one step-ahead forecasts. This article uses a one week multistep-ahead forecast because it is simpler and less computationally demanding, as

[15]Gold futures contracts started trading on the COMEX in 1978, while silver futures contracts have been traded on the same exchange since 1963.

well as not being necessarily inferior to one step-ahead forecasts [see Swamy and Schinasi (1989)]. Therefore, the forecasting strategy involves estimating the parameters of adequately specified error correction models for the gold–silver cash and futures spreads only once from observations in the estimation period, which are then used to form one week-ahead multistep forecasts using actual realizations of all explanatory variables in the forecast period. The model is reestimated every five observations by appending the latest five observations to the estimation period so that the information set used in forecasting is based on an increasing sequence of historical realizations of the spread series. This procedure is utilized repeatedly until the entire data set is exhausted. The first estimation period is taken to be January 1982 through June 1988, and the first forecast horizon spans the first week in July 1988, and so on. Both fixed and time varying parameter versions of the forecasting models are tried.

Because estimated error correcting models are used to generate ex-ante values of the gold–silver spread (cash and futures), it is important to consider a few issues that may jeopardize forecast quality. The first problem arises because the error correction models are specified to include contemporaneous values of the endogenous variables which introduces simultaneity into the estimation process, making parameter estimates biased and inconsistent and potentially inefficient in the presence of cross-equation error correlation. To overcome this problem, pairs of error correction equations are estimated using three-stage least squares instrumental variables estimation, with instruments taken as fitted values generated from adequately specified ARIMA models of the endogenous variables. However, because the implications (forecasting, trading, and arbitrage) drawn from estimating error correction models using the instrumental variable methodology are conditioned by the choice of the particular instrument, the models are respecified to include different instruments. More specifically, expected changes in the cash and futures spreads are generated from three alternative moving averages: a two-week moving average, a one-month moving average, and a two-month moving average (corresponding to the typical expiration cycle of gold futures contracts). While numerous variations of the moving average rule are possible, only a few simple ones are used. A moving average rule was used in previous studies of the gold–silver parity by Ma (1985), and Ma and Soenen (1988). A moving average of historical realizations is used as a proxy for expectations to: (i) capture information about trends of varying lengths in the expected change series and (ii) simplify estimation, thereby avoiding the need to augment the system of error correction models with auxiliary equations describing expectations formation as part of the system. A presumption, needed to justify the use of the moving average representation in price changes, is that the stationarity requirement is more than likely satisfied in first differences (rather than levels). To evaluate the forecasting performance of each alternative instrument, the mean squared forecast error (MSE) criterion is employed, and is computed as follows, given n-pairs of forecasts (F_t) and corresponding actual values (A_t) of changes in the cash and futures spreads:

$$\text{MSE} = \sum_{t=1}^{n}(F_t - A_t)^2/n$$

To establish the dominance of one moving average instrument over the basic instrument (generated from an ARIMA model) under the mean squared error criterion, the difference, D_i, in the squared forecast errors generated when employing the basic instrument versus a moving average instrument, is computed for each of the cash and futures spreads as:

$$D_i = (F_{i1} - A_i)^2 - (F_{i2} - A_i)^2$$

where D_i is the difference for the ith spread ($i = 1, 2$); F_{i1} and F_{i2}, respectively, refer to forecasted values for the ith spread using the basic instrument (1) and a moving average instrument (2). Instrument 1 is judged to outperform instrument 2 if the mean of these differences is negative and statistically significant at the 5% level. The appropriate test is a two-tailed t test.

For further insight into the forecasting performance of each instrument, the MSE corresponding to each instrument is decomposed into three components, each of which represents a particular type of forecast error [Theil (1971)]:

$$\text{MSE}_i = (\tilde{F}_i - \tilde{A}_i) + (\sigma_{F_i} - \sigma_{A_i})^2 + 2(1 - \rho)\sigma_{F_i}\sigma_{A_i}$$

where $(\tilde{F}_i, \tilde{A}_i)$ and $(\sigma_{F_i}, \sigma_{A_i})$ are, respectively, the means and standard deviations of the forecast and actual series, and (ρ) is the correlation coefficient between the forecasted and actual changes in the ith spreads. The first term is called the bias component, the second is a measure of inefficiency, and the third measures the error due to imperfect correlation (random noise).

Each component can be normalized for ease of comparability by dividing through by the MSE:

$$\text{Bias proportion: } U^M = (\tilde{F} - \tilde{A})^2/\text{MSE}$$

$$\text{Inefficiency proportion: } U^S = (\sigma_F - \sigma_A)^2/\text{MSE}$$

$$\text{Covariance proportion: } U^C = 2(1 - \rho)\sigma_F\sigma_A/\text{MSE}$$

The second difficulty concerns potential heteroskedasticity in the error correction model's residuals, which may lessen the model's forecasting ability.[16] Conditional serial-heteroskedasticity is checked using the Ljung–Box chi-square test on the autocorrelation function of squared residuals.

The third problem pertains to potential autocorrelation in the error terms. A Ljung–Box test is applied on the autocorrelation function of the residuals because the Durbin–Watson test cannot be used to test for serially correlated errors due to the presence of lagged endogenous variables on the RHS of each equation; the Durbin h test lacks power; and it is a test only for first-order serial correlation.

The remaining problem concerns possible variation over time in the error correction model parameters. Because forecasts are made conditional on the estimated parameters, failure to incorporate information contained in the temporal behavior of the parameters may increase forecast error variance and jeopardize forecast quality. Therefore, as a first step, a Farley–Hinich test of parameter stability is conducted. It is postulated that those parameters that are suspected of temporal instability follow a simple deterministic linear time trend process. If the null hypothesis of coefficient stability is rejected, this information is incorporated into the forecasting model.[17]

If the forecasting model indicates a positive predicted change in the gold–silver spread, a long position in gold is established coupled with a

[16]Numerous studies have found that many financial time series exhibit conditional heteroskedasticity. This is often interpreted to be due either to the arrival process of news, which exhibits "lumpiness," or to the gradual market response to news. Such autoregressive conditional heteroskedasticity effects, however, are likely to be more pronounced when using higher frequency data, rather than daily.

[17]This is not meant to imply that a deterministic linear time trending behavior is the only possible temporal behavior. Rather, this behavior is postulated because it is the simplest to model. Alternative deterministic or stochastic processes may require a separate study of their own.

short position in silver.[18] Likewise, if the forecasted change in the spread is negative, a reverse strategy is established in each metal; that is, gold is shorted and silver is purchased. Once a position is open, if the forecast for the subsequent change in the spread is non-negative, an open long position is maintained. In the case of a negative predicted change, an open short position would be maintained also. These positions are closed out one day before expiration and a shift occurs in the next nearby contract on the expiration day of the previous contract.

EMPIRICAL RESULTS

Cross-Correlations Between Alternative Measures of the Spread

To get a feel for how close the alternative definitions of the precious metals spreads are to each other, their in-sample cross-correlations are computed. These results are presented in Table I. First, note that the constrained gold–silver cash and futures spreads (parities) are highly cross-correlated with a pairwise correlation coefficient of 0.998. Similarly, the unconstrained gold–silver cash and futures spreads are highly cross-correlated with a correlation coefficient of 0.993. Both results suggest that the cash and futures markets are well linked by arbitrageurs who trade precious metals in both markets. Second, the constrained and unconstrained gold–silver cash spreads are modestly cross-correlated with a coefficient of 0.376. This suggests that a trading strategy based on the gold–silver cash spread would yield results that are sensitive to the way the spread is defined. A similar observation holds for trading strategies predicated on the gold–silver futures spread. With a cross-correlation coefficient of 0.376 between the constrained and unconstrained futures spreads, the performance of the trading strategy is also likely to be affected by the particular constraints imposed on the joint distribution of gold and silver futures prices. Third, the empirical cross-correlation between the unconstrained gold–silver futures spread

[18]Precious metals can be sold short in the cash market. Indeed, there exists a cash market in New York for the borrowing of gold and silver for delivery against short sales [see Garbade and Silber (1982)]. Even if silver is fairly- or under-priced relative to its fair (equilibrium) value, a short silver position would still be recommended. This is so because if the expected parity exceeds the current parity, then the extent of under-pricing of gold is probably stronger than the degree of under-pricing in silver so that speculative risk-unadjusted profits may still be locked-in. Just as an arbitrageur who speculates on subsequent movements in the gold–silver spread is more concerned with the change in the gold–silver differential than with the change in the price of each metal, each considered separately, the issue of the extent of mispricing in each precious metal taken separately should not be of primary concern in this context.

TABLE I
Sample Cross-Correlations between Logarithmic Levels of Constrained and Unconstrained Cash and Futures Gold–Silver and Silver–Gold Spreads
Sample Period: Jan. 1982–Jul. 1992

	P_c^c	P_f^c	P_c^u	P_f^u	P_c^{*u}	P_f^{*u}
P_c^c	1.0000					
	(0.0000)					
P_f^c	0.9980	1.0000				
	(0.0001)	(0.0000)				
P_c^u	0.3761	0.3743	1.0000			
	(0.0001)	(0.0001)	(0.0000)			
P_f^u	0.3764	0.3766	0.9935	1.0000		
	(0.0001)	(0.0001)	(0.0001)	(0.0000)		
P_c^{*c}	−0.9964	−0.9944	−0.4525	−0.4522	1.0000	
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	
P_f^{*u}	−0.9951	−0.9970	−0.4438	−0.4466	0.9979	1.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)

Notes:
Superscript *c* and *u* identify, respectively, the constrained and unconstrained spreads.
Subscripts *c* and *f* identify, respectively, the cash and futures spreads.
P and P^* denote, respectively, the gold–silver and the silver–gold spreads.
Numbers reported in parentheses immediately below each row are marginal significance levels.
Negative cross-correlations occur because of cross-correlating a ratio with an inverse ratio.
Unconstrained silver–gold spreads (P_c^{*u} and P_f^{*u}) are estimated from the set of eq. (8), after reversing the LHS and RHS variables designations.

is −0.452, while the correlation between the unconstrained gold–silver futures spread and the unconstrained silver–gold futures spread is −0.446. Of course, in the constrained version of the spread, the cross-correlation between the gold–silver spread and the silver–gold spread (cash or futures) should be negative unity, because one series is simply the reciprocal of the other. These results are generally supportive of the need to examine alternative definitions of the precious metals spread when searching for ex-ante profitable arbitrage opportunities.

Time Series Properties of the Alternative Spreads

Table II presents the empirical autocorrelation estimates for the alternative spreads for lags 1 through 6, using both the logarithmic level as well as the logarithmic first difference of each spread. The levels of the alternative series are highly autocorrelated. The smallest of the six autocorrelation coefficients in any one series is 0.970. Naturally, the question of whether each series follows a stationary stochastic process depends on the rate of decay of the autocorrelation function, because it is the flatness of the autocorrelations, not just their level, that

TABLE II
Sample Autocorrelation Coefficients of Logarithmic Levels and Logarithmic First Differences of Constrained and Unconstrained Cash and Futures Gold–Silver and Silver–Gold Spreads
Sample Period: Jan. 1982–Jul. 1992

	Autocorrelation Coefficients							*Daily*	
Spread Series	r1	r2	r3	r4	r5	r6	*Ljung–Box Statistic*	*Standard Deviation*	*Sample Size*
$P^c_{c,t}$	0.998	0.996	0.994	0.993	0.991	0.986	9999.9[a]	0.8561	2596
$P^c_{c,t} - P^c_{c,t-1}$	−0.078	0.029	−0.038	−0.021	0.015	−0.032	25.72[a]	0.0139	2595
$P^c_{f,t}$	0.997	0.995	0.993	0.991	0.990	0.988	9999.9[a]	1.1880	2596
$P^c_{f,t} - P^c_{f,t-1}$	−0.263	−0.052	0.013	−0.002	−0.012	0.032	215.03[a]	0.0204	2595
$P^u_{c,t}$	0.995	0.991	0.987	0.983	0.979	0.975	9999.9[a]	0.1035	2596
$P^u_{c,t} - P^u_{c,t-1}$	−0.109	0.024	0.028	−0.039	0.018	−0.006	39.15[a]	0.0102	2595
$P^u_{f,t}$	0.990	0.986	0.982	0.978	0.974	0.970	9999.9[a]	0.1041	2596
$P^u_{f,t} - P^u_{f,t-1}$	−0.231	−0.066	0.028	−0.023	0.001	0.032	155.80[a]	0.0140	2595
$P^{*u}_{c,t}$	0.995	0.991	0.987	0.983	0.979	0.975	9999.9[a]	0.1038	2596
$P^{*u}_{c,t} - P^{*u}_{c,t-1}$	−0.095	0.025	−0.029	−0.034	0.019	−0.027	33.04[a]	0.0138	2595
$P^{*u}_{f,t}$	0.990	0.985	0.982	0.977	0.974	0.970	9999.9[a]	0.1045	2596
$P^{*u}_{f,t} - P^{*u}_{f,t-1}$	−0.282	−0.062	0.020	−0.008	−0.007	0.008	244.72[a]	0.0211	2595

Notes:
See footnotes in Table I for alternative definitions of the spreads.
Ljung–Box Statistic testing the null hypothesis that all serial correlations taken jointly up to the 6th lag are insignificantly different from zero. The test statistic is distributed as a χ^2 variate with degrees of freedom equal to the number of lags.
Asymptotic standard errors for the autocorrelation coefficients taken individually can be approximated as the square root of the reciprocal of the number of observations (e.g., ± 0.0196 for 2595 observations) under the null hypothesis of zero serial correlation at lag k. One observation is lost when taking first differences.
[a]Denotes statistical significance at the 1% level.

signifies the presence of an autoregressive root close to or equal to unity. Also presented in Table II are the first six autocorrelations of the first differences of each series. As observed, the first differences are much less autocorrelated than the levels, and they exhibit rapid decay. All first-difference series demonstrate negative first-order serial correlations that are reliably different from zero at conventional significance levels.

While the autocorrelations in Table II suggest that most of the series contain a unit root, the autocorrelations do not provide a formal test. Therefore, Table III presents the Dickey Fuller (DF) and Augmented Dickey Fuller (ADF) test results of the null hypothesis of a single unit root in the levels representation of each spread series. The specifications of the DF and ADF-tests are given according to eq. (9). Because a time trend term is included in the model, the alternative hypothesis is that these variables follow a stationary process (around a deterministic time

TABLE III

Tests for A Unit Roof in the Logarithmic Levels of Constrained and Unconstrained Cash and Futures Gold–Silver and Silver–Gold Spreads
Sample Period: Jan. 1982–Jul. 1992

Parity	*DF(0)*	*ADF(1)*	*ADF(2)*	*ADF(3)*	*ADF(4)*
$P^{c}_{c,t}$	−2.712	−2.414	−2.448	−2.521	−2.452
$P^{c}_{f,t}$	−3.369	−3.228	−3.043	−2.946	−2.905
$P^{u}_{c,t}$	−3.324	−3.108	−3.170	−3.091	−3.019
$P^{u}_{f,t}$	−3.106	−2.946	−2.581	−2.533	−2.454
$P^{*u}_{c,t}$	−3.346	−3.072	−3.114	−3.054	−2.952
$P^{*u}_{f,t}$	−3.342	−3.124	−3.065	−2.946	−2.876

Notes:
DF and ADF are, respectively, Dickey Fuller and Augmented Dickey Fuller tests for Unit Roots. Critical values are obtained following the method in MacKinnon [1991, Table I]. For the full sample period (2596 observations) critical values are −3.9670 (1%), −3.4141 (5%), and −3.1288 (10%). These critical values apply for both tests.
Numbers appearing in parentheses next to DF and ADF indicate the number of lagged terms included in the testing eq. (9).

trend). As shown, all test statistics are well below the 1%, and even the 5%, critical values so that the null hypothesis of a single unit root is not rejected for any of the series.[19] This suggests that first differencing is adequate to render each series stationary.

Cointegration test results are presented in Table IV for alternative linear combinations of the cash and futures spreads. The null hypothesis of noncointegration is always reliably rejected at better than the 1% significance level, suggesting that linear combinations of the cash and futures spreads examined are indeed cointegrated.

The cointegrating regression parameters are also presented. The OLS estimates (α, β) are generally very close to $(0, 1)$. Unfortunately, the usual t test that the slope is unity is not applicable because of non-stationarity in the levels of each series, which biases and renders inconsistent the estimated standard errors of the parameters. Given that pairs of cash and futures spreads appear to form a cointegrated system, optimally specified error correction models (ECM) are estimated. The optimal specification of the ECM entails a decision about the appro-

[19]As a precautionary step, Dickey–Fuller (and augmented Dickey–Fuller) tests for two unit roots also conducted by regressing the second differences in each series against the lagged first differences (augmented by up to three lagged values of the second differences). The test results suggest rejection of the null hypothesis of two unit roots for both the constrained and unconstrained cash and futures spreads with almost complete confidence. Similarly, Dickey–Pantula (1987) type tests for three then two, then one unit root are run, and the results are also supportive of one unit root being present in the levels of each series. These results are not reported but are available from the authors upon request.

TABLE IV
Cointegration Test Results for Constrained and Unconstrained Pairs of Cash and Futures Spreads
Sample Period: Jan. 1982–Jul. 1992

Spread Series	*DF(0)*	*ADF(1)*	*ADF(2)*	*ADF(3)*	*ADF(4)*
$P^c_{c,t}$ & $P^c_{f,t}$	−37.581[a]	−26.741[a]	−21.931[a]	−19.240[a]	−18.349[a]
$P^c_{f,t}$ & $P^c_{c,t}$	−37.649[a]	−26.803[a]	−21.984[a]	−19.290[a]	−18.402[a]
Cointegrating Regression				*Cointegrating Regression*	
$P^c_{c,t} = 0.027 + 0.933\, P^c_{f,t} + z^c_{c,t}$			$P^c_{f,t} = -0.011 + 1.003\, P^c_{c,t} + z^c_{f,t}$		
$R^2 = 0.996$ DW $= 1.411$ $\rho_1 = 0.294$			$R^2 = 0.996$ DW $= 1.414$ $\rho_1 = 0.293$		
$P^u_{c,t}$ & $P^u_{f,t}$	−40.803[a]	−28.509[a]	−21.309[a]	−19.545[a]	−17.726[a]
$P^u_{f,t}$ & $P^u_{c,t}$	−41.030[a]	−28.738[a]	−21.512[a]	−19.746[a]	−17.918[a]
Cointegrating Regression				*Cointegrating Regression*	
$P^u_{c,t} = -4.2\text{E}{-16} + 0.987\, P^u_{f,t} + z^u_{c,t}$			$P^u_{f,t} = -4.16\text{E}{-16} + 0.999\, P^u_{c,t} + z^u_{f,t}$		
$R^2 = 0.987$ DW $= 1.566$ $\rho_1 = 0.217$			$R^2 = 0.987$ DW $= 1.575$ $\rho_1 = 0.212$		
$P^{*u}_{c,t}$ & $P^{*u}_{f,t}$	−38.662[a]	−27.534[a]	−22.043[a]	−19.361[a]	−18.310[a]
$P^{*u}_{f,t}$ & $P^{*u}_{c,t}$	−38.754[a]	−27.619[a]	−22.118[a]	−19.434[a]	−18.386[a]
Cointegrating Regression				*Cointegrating Regression*	
$P^{*u}_{c,t} = 9.05\text{E}{-16} + 0.992\, (P^{*u}_{f,t}) + z^{u'}_{c,t}$			$P^{*u}_{f,t} = -9.04\text{E}{-16} + 1.003\, (P^{*u}_{c,t}) + z^{u'}_{f,t}$		
$R^2 = 0.996$ DW $= 1.464$ $\rho_1 = 0.268$			$R^2 = 0.996$ DW $= 1.467$ $\rho_1 = 0.267$		

Notes:
DF and ADF are, respectively, Dickey Fuller and Augmented Dickey Fuller tests for Unit Roots. Critical values are obtained following the method in MacKinnon [1991, Table (1)]. For the full sample period (2597 observations) critical values are −3.9670 (1%), −3.4141 (5%), and −3.1288 (10%). These critical values apply for both tests.
Numbers appearing in parentheses next to DF and ADF indicate the number of lagged terms included in the testing eq. (9).
DW is Durbin–Watson Statistic.
Pairs of Spreads appearing under the parity column are arranged such that the first term is the LHS variable while the second term is the RHS variable in a cointegrating regression involving the two spreads.
[a]Denotes statistical significance at the 1% level.

priate lag orders of the LHS variables, which are determined using Akaike's Final Prediction Error (FPE) criterion.[20] The lag distribution

[20]The Final Prediction Error (FPE) is defined as the (asymptotic) mean squared prediction error:

$$\text{FPE of } Y_t = E(Y_t - \hat{Y}_t)^2$$

where $\hat{Y}_t$ is the predictor of Y_t. Akaike (1969) defines the estimate of the FPE by:

$$\text{FPE}_y(m, n) = \frac{T + m + n + 1}{T - m - n - 1} \cdot \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 / T$$

where m and n determine the optimal lag orders in a two-variable equation system. The first term on the RHS of the equality is a measure of estimation error, and the second term is a measure of modeling or specification error. FPE reflects a tradeoff between bias, when a lower lag order model is selected, and inefficiency, when a higher order is chosen, by choosing the smallest FPE specification.

is truncated initially to four lags because there is little reason to believe that longer lags would be warranted when using daily data.[21]

Error Correction Model Estimates

Four optimally specified error correction regressions are performed for each pair of cash and futures spread series. These alternative specifications are based on forward and reverse representations of Engle and Granger's (1987) cointegrating regressions. Each pair of error-correcting equations that use the same cointegrating regression error series is jointly estimated using three stage least squares. To save space, only the results for the constrained and unconstrained gold–silver spreads are presented in Table V for the minimum FPE error correction models.

TABLE V

Three Stage Least Squares Estimates of the Error Correction Models (ECM) for Constrained and Unconstrained Cash and Futures Gold–Silver Spreads
Sample Period: Jan. 1982–Jul. 1992

Panel A: Constrained Gold–Silver Futures and Cash Spreads

1. Lagged Futures Pricing Shocks as Error Correcting Mechanism

ECM1: $$\Delta P^c_{f,t} = \gamma_0 - \gamma_1 z^c_{f,t-1} + \sum_{i=1}^{2} \beta_{if} \Delta P^c_{f,t-i} + \sum_{j=0}^{2} \beta_{jc} \Delta P^c_{c,t-j} + e^c_{f,t}$$

0.0005	−0.486	−0.202	−0.104	0.773	0.153	0.024
(0.958)	(−17.664)[a]	(−7.805)[a]	(−4.911)[a]	(31.431)[a]	(5.186)[a]	(0.933)

ECM2: $$\Delta P^c_{c,t} = \gamma'_0 - \gamma'_1 z^c_{f,t-1} + \alpha_{1c} \Delta P^c_{c,t-1} + \alpha_{13} \Delta P^c_{c,t-3} + \sum_{i=0}^{2} \alpha'_{if} \Delta P^c_{f,t-i} + e^c_{c,t}$$

Where $z^c_{f,t-1} = P^c_{f,t} - (a + bP^c_{c,t})$

0.0006	−0.386	−0.112	−0.008	0.389	0.095	0.042
(1.190)	(−21.518)[a]	(−5.824)[a]	(−0.548)	(32.841)[a]	(5.598)[a]	(3.230)[a]

System-Weighted R^2: 0.7081 System-Weighted MSE: 0.825

2. Lagged Cash Pricing Shocks as Error Correcting Mechanism

ECM3: $$\Delta P^c_{f,t} = \alpha_0 + \gamma_0 z^c_{c,t-1} + \sum_{i=1}^{3} \lambda_{if} \Delta P^c_{f,t-i} + \sum_{j=0}^{2} \lambda'_{jc} \Delta P_{c,t-j} + e^{c'}_{f,t}$$

0.0001	−0.598	−0.233	−0.126	−0.029	1.224	0.205	0.033
(0.247)	(−20.883)[a]	(−8.445)[a]	(−5.294)[a]	(−1.791)	(56.556)[a]	(6.573)[a]	(1.363)

ECM4: $$\Delta P^c_{c,t} = \alpha'_0 + \gamma'_0 z^c_{c,t-1} + \sum_{j=1}^{2} \delta_{1c} \Delta P^c_{c,t-j} + \sum_{i=0}^{2} \delta'_{if} \Delta P^c_{f,t-i} + e^{c'}_{c,t}$$

Where $z^c_{c,t-1} = P^c_{c,t-1} - (a' + b'P^c_{f,t-1})$

0.0002	−0.434	−0.127	−0.001	0.562	0.127	0.058
(0.763)	(−24.270)[a]	(−6.601)[a]	(0.077)	(56.462)[a]	(7.510)[a]	(4.535)[a]

System-Weighted R^2: 0.7090 System-Weighted MSE: 0.779

continued

[21] Because there are two variables on the RHS of each error correction equation, there are 5^2 possible choices for each equation. Therefore, to reduce the computational burden, Hsiao's (1981) sequential estimation procedure is followed. A series of over-fitting F-tests are also conducted to ensure specification adequacy.

TABLE V (Continued)

Panel B: Unconstrained Gold–Silver Futures and Cash Spreads

1. Lagged Futures Pricing Shocks as Error Correcting Mechanism

ECM5: $\Delta P^u_{f,t} = \gamma_0 - \gamma_1 z^u_{f,t-1} + \beta_{if}\Delta P^u_{f,t-1} + \sum_{j=0}^{1} \beta'_{jc}\Delta P^c_{c,t-j} + e^u_{f,t}$

−0.00001	−0.632	−0.094	0.776	0.079
(−0.007)	(−24.214)[a]	(−4.357)[a]	(31.513)[a]	(3.234)[a]

ECM6: $\Delta P^u_{c,t} = \gamma'_0 - \gamma'_1 z^u_{f,t-1} + \alpha_{1c}\Delta P^u_{c,t-1} + \sum_{i=0}^{1} \alpha'_{if}\Delta P^u_{f,t-i} + e^u_{c,t}$

Where $z^u_{f,t-1} = P^u_{f,t-1} - (a + bP^u_{c,t-1})$

0.0000	−0.461	−0.101	0.577	0.078
(0.0000)	(−26.236)	(−6.129)[a]	(32.282)[a]	(5.367)[a]

System-Weighted R^2: 0.785 System-Weighted MSE: 0.707

2. Lagged Cash Pricing Shocks as Error Correcting Mechanism

ECM7: $\Delta P^c_{f,t} = \alpha'_0 + \gamma'_0 z^u_{c,t-1} + \sum_{i=1}^{3} \delta_{if}\Delta P^u_{f,t-i} + \sum_{j=0}^{2} \delta_{jc}\Delta P^u_{f,t-j} + e^u_{f,t}$

0.0000	−0.871	−0.276	−0.216	−0.078	0.780	0.270	0.211
(0.0001)	(−34.476)[a]	(−9.049)[a]	(−8.057)[a]	(−3.774)[a]	(32.181)[a]	(8.073)[a]	(7.216)[a]

ECM8: $\Delta P^c_{c,t} = \alpha'_0 + \gamma'_0 z^u_{c,t-1} + \delta_{1c}\Delta P^u_{c,t-1} + \sum_{i=0}^{2} \delta'_{if}\Delta P^c_{f,t-i} + e_{c,t}$

Where $z^u_{c,t-1} = P^u_{c,t-1} - (a' + b'P^u_{f,t-1})$

0.0000	−0.430	−0.128	0.660	0.108	0.039
(0.0001)	(−21.501)[a]	(−6.458)[a]	(31.821)[a]	(5.901)[a]	(2.783)[a]

Notes:
Superscripts c and u identify, respectively, the constrained and unconstrained spreads.
Subscripts c and f identify, respectively, the cash and futures spreads.
P and P^* denote, respectively, gold–silver and silver–gold spreads.
[a]Denotes significance at the 1% level.

Panel A contains the results for the constrained gold–silver cash and futures spreads (parities). The models ECM1 and ECM2 are estimated jointly, using lagged futures equilibrium pricing errors as error correction terms, while ECM3 and ECM4, also jointly estimated, use the cash equilibrium pricing errors as the error correction mechanism. The error correction coefficients serve two purposes: to identify the direction of the causal relation between changes in the spot and futures spreads and to indicate the speed with which disequilibria are corrected by short-run adjustments in the LHS variables. For example, ECM1 has the interpretation that changes in the constrained futures spread, $\Delta P^c_{f,t}$, are due to both short-run effects from lagged values of $\Delta P^c_{f,t}$ and $\Delta P^c_{c,t}$, and to last period's futures cointegrating errors (from the cointegrating regression) that represent adjustments to long-run equilibrium. If the error correction coefficient is small or statistically insignificant, then $\Delta P^c_{f,t}$ has little tendency to adjust to correct a disequilibrium situation;

thus most of the adjustments may be accomplished through subsequent adjustments in the cash spread $\Delta P^c_{c,t}$ in ECM2.

An examination of the estimates of error correction models ECM1 through ECM4 reveals several interesting results. First, deviations of gold and silver cash prices from their equilibrium values trigger stronger subsequent adjustments in futures prices than do deviations of futures prices with respect to cash prices. More specifically, the coefficient γ_0 in ECM3 of -0.598 is almost twice the magnitude of the coefficient γ_1' of -0.386 in ECM2. This suggests that about 60% of the gap or deviation from the equilibrium value of the constrained gold–silver cash spread is amended each day through adjustments in gold–silver futures markets. On the other hand, about 40% of the deviations from the equilibrium constrained gold–silver futures spread are corrected within a day through subsequent adjustments in the gold–silver cash spread. These results are consistent with the transactions cost advantage of futures over the cash market and are also suggestive of a potential causal ordering between the two markets. Pricing shocks in the cash market appear to be relatively more influential than pricing shocks in the futures markets.

Second, the cash and futures spreads exhibit subsequent adjustments in response to their own pricing disequilibria. The coefficient γ_1 in ECM1 is -0.486, and the coefficient γ_0' in ECM4 is -0.434, which can also be interpreted as evidence of mean reversion in the futures market that is more pronounced than in the cash market. Third, the contemporaneous coefficients in ECM1–ECM4 are close to (but statistically different from) unity, indicating that, although the cash and futures markets are interrelated on a predominantly instantaneous basis, some delayed adjustments also reliably prevail. Indeed, the coefficients on the one-day lagged changes in the cash and futures spreads across the four models are always statistically significant at the 1% level. In some cases the coefficients on the two-day lagged changes in the cash and futures spreads are also statistically significant. Lagged futures-to-cash and cash-to-futures interactions are generally of similar magnitudes, suggesting that feedback exists between the two markets. Furthermore, linear restrictions (using an F-test) on one-and two-day lagged coefficients are rejected with almost complete certainty, providing supporting evidence of the statistical significance of lagged interactions. The in-sample fit of the models is impressive, with a goodness-of-fit R^2 measure of about 0.71. Panel B presents the parameter estimates of the four error correction models for daily changes in the unconstrained gold–silver cash and futures spreads. The results imply conclusions very similar to those obtained from panel A. The error correction models ECM5 though ECM8 for the unconstrained gold–silver cash and futures spreads demonstrate better in-sample fit, with an R_2 statistic of 0.78 for ECM5 and ECM6, and 0.801 for ECM7 and ECM8. To ensure proper specification of the error-correction models, a variety of standard diagnostic tests are applied to the residuals and their squared values (results not reported). The evidence is broadly consistent with model adequacy.

To investigate the extent to which the instrument choice affects the forecast performance of the estimated error correction models, Table VI reports forecast performance results, together with a decomposition of the MSE statistics associated with using three moving average instrumental variables. Results are shown for error correction models ECM1 through ECM8, corresponding to the constrained and unconstrained cash and futures gold–silver spreads. Several interesting findings are revealed. First, an ARIMA instrument appears to have consistently produced lower MSE across all ECMs when compared to other moving average instruments. Evidence about the statistical significance of this differential forecast performance is shown in panel C. Indeed, fitted values from an ARIMA(1, 1, 1) model for each of the cash and futures constrained and unconstrained spreads outperform a two-month moving average rule. This superior performance is statistically significant at the 0.01% level. On the other hand, from a statistical significance viewpoint, an ARIMA instrument appears to produce forecasts of the same quality as those produced by a two-week or a one-month moving average rule. None of the differences in squared residuals are statistically significant at any reasonable level, save one case.

To shed more light on the performance of each instrument, Panels A and B present evidence about the decomposition of the MSE statistics into their three components. The bias proportion, U^M, is an indication of systematic error; it is hoped that U^M is close to zero. The variance (inefficiency) proportion, U^S, indicates the ability of the instrument to replicate the degree of variability in the variable of interest; it is hoped that U^S is close to zero also. Finally, the covariance proportion, U^C, measures the random or unsystematic error component of the total forecast error. Ideally, this proportion would be close to 1. Once more, the results suggest that an ARIMA instrument does just as well, if not better than the alternative moving average instruments. Given these findings, all reported results are based on estimating and simulating forward ECMs with ARIMA instruments. The question then naturally arises: Can incorporating an

TABLE VI

Forecast Performance and Decomposition of the MSE Statistics for Alternative Instruments

	Panel A: Constrained Cash and Futures Gold–Silver Spreads															
	ECM1				*ECM2*				*ECM3*				*ECM4*			
Instrument	*MSE*	U^M	U^S	U^C	*MSE*	U^M	U^S	U^C	*MSE*	U^M	U^S	U^C	*MSE*	U^M	U^S	U^C
ARIMA	0.164	2.453	0.202	0.794	0.082	6.931	0.217	0.782	0.081	2.995	0.040	0.959	0.039	0.011	0.088	0.912
MA(10)	0.165	3.865	0.300	0.699	0.086	7.352	0.362	0.636	0.087	2.099	0.159	0.848	0.045	0.045	0.464	0.535
MA(22)	0.218	1.537	0.319	0.680	0.176	8.109	0.365	0.634	0.097	2.864	0.175	0.824	0.051	0.084	0.466	0.533
MA(44)	0.327	1.507	0.310	0.687	1.021	9.327	0.301	0.697	0.719	0.003	0.492	0.507	0.129	0.197	0.054	0.945
	Panel B: Unconstrained Cash and Futures Gold–Silver Spreads															
	ECM5				*ECM6*				*ECM7*				*ECM8*			
	MSE	U^M	U^S	U^C	*MSE*	U^M	U^S	U^C	*MSE*	U^M	U^S	U^C	*MSE*	U^M	U^S	U^C
ARIMA	0.080	0.399	0.040	0.959	0.039	0.011	0.088	0.911	0.081	0.195	0.039	0.960	0.039	0.023	0.061	0.938
MA(10)	0.087	0.409	0.151	0.848	0.045	0.045	0.464	0.535	0.087	0.225	0.170	0.829	0.045	0.196	0.160	0.839
MA(22)	0.087	1.864	0.174	0.824	0.050	1.000	0.466	0.533	0.088	0.812	0.213	0.785	0.074	0.273	0.457	0.542
MA(44)	0.081	0.003	0.492	0.507	0.120	1.970	0.453	0.545	0.089	0.146	0.862	0.137	0.125	0.386	0.388	0.610

continued

TABLE VI (Continued)

Instrument Pair	Panel C: *t-statistics for the difference* (D_j) *in Squared Errors Associated with Alternative Instruments*							
	ECM1	*ECM2*	*ECM3*	*ECM4*	*ECM5*	*ECM6*	*ECM7*	*ECM8*
[ARIMA,	−0.066	−0.528	0.138	−0.642	0.138	−0.642	0.160	−0.764
MA(10)]	(0.947)	(0.597)	(0.890)	(0.521)	(0.891)	(0.521)	(0.872)	(0.444)
[ARIMA,	−1.563	−0.197	0.162	−1.178	0.161	−1.179	−0.750	−2.195
MA(22)]	(0.118)	(0.844)	(0.871)	(0.239)	(0.870)	(0.241)	(0.453)	(0.028)[a]
[ARIMA,	−3.878	−5.871	−7.682	−3.693	−7.683	−3.690	−8.026	−3.656
MA(44)]	(0.0001)[b]	(0.0001)[b]	(0.0001)[b]	(0.0002)[b]	(0.0001)[b]	(0.0002)[b]	(0.0001)[b]	(0.0003)[b]

Notes:
MSE denotes mean square error. MSE numbers are multiplied by 1000.
U^M, U^S, U^C denote, respectively, bias, inefficiency, and covariance proportions. Bias proportions are multiplied by 1000. The three proportions add up to unity.
MA(q) denotes a moving average of the qth order, used as an instrument. The original instrument used in generating the ECM estimates reported in Table (V) is generated from an ARIMA(1, 1, 1) model.
Numbers reported in parentheses below the standard t-statistics, in panel C, are marginal significance levels.
[a]Denotes statistical significance at better than the 5% level.
[b]Indicates statistical significance at better than the 1% level.

error correction mechanism together with the observed lagged interactions be exploited to generate ex-ante arbitrage profits?

Test of Parameter Stability

The purpose of this study is to see whether a statistical model that is consistent with the empirical regularities in the data could profitably exploit the serial and cross-dependencies in the series. The study does not seek to identify an "optimal" forecasting model or trading strategy that is capable of maximizing profits at any risk level. An additional complexity is introduced into the analysis, that is, recognition of possible movements in the error correction models' underlying parameters.

There are several reasons why the coefficients may change over time. First, even if the explanatory variables capture all information used by traders, there is no reason to believe that information is used in the same way over all time periods. Second, because there is wide diversity of participants in the precious metals markets (individual traders, and public and private institutions), it is difficult to argue that each participant would react to developments that affect gold and silver prices according to a temporally stable fixed coefficient reaction function, even if one is willing to assume that participants' reaction functions are cross-sectionally identical at any point in time. Third, because forecasts of changes in the spread are made conditional on the estimated parameters, forecast quality may be improved by relaxing the assumption of fixed coefficients.

Therefore, for example, the first error correction model ECM1 is extended as follows to allow for linearly and deterministically trending parameters:

$$\begin{aligned}\Delta P^c_{f,t} = \gamma_0 - \gamma_1 z^c_{f,t-1} + \gamma'_1 T z^c_{f,t-1} + \sum_{i=1}^{2} \beta_{i,f} \Delta P^c_{f,t-1} \\ + \sum_{i=1}^{2} \beta'_{i,f} T \Delta P^c_{f,t-i} + \sum_{j=0}^{2} \beta'_{j,c} \Delta P^c_{c,t-j} \\ + \sum_{j=1}^{2} \beta''_{j,c} \Delta P^c_{c,t-j} T \Delta P^c_{c,t-j} + e^c_{f,t}\end{aligned}$$

If γ'_1 is statistically significantly different from zero, the error correction coefficient is declared time varying. Likewise, if the coefficients $\beta'_{1,f}, \beta'_{2,f}, \beta''_{1,c}, \beta''_{2,c}$ are statistically significant, then each exhibits a deterministic linear trend.

Test results are presented in Table VII where only the estimates of the newly created coefficients (for example, in ECM1, these are $\gamma_1^{\prime}$, $\beta_{i,f}^{\prime}$'s and $\beta_{j,c}^{\prime\prime}$'s) and their t-statistics are reported.[22] Panel A presents results for the constrained gold–silver spread (parity) while panel B presents the results for the unconstrained gold–silver spread. Results for the unconstrained silver–gold spread are not reported in the interest of brevity, but reflect qualitatively the same conclusions. As is apparent, almost all models exhibit trending behavior in the error correction coefficients, albeit with different signs. Some coefficients are trending upwards over time while some are trending downwards. An upward trend shows that correction responses of a particular market (cash or futures) are strengthening over time, whereas a downward trend reflects a weakening response. Some of the coefficients representing lagged interactions in "own" market as well as across the two markets (cash and futures) exhibit similar time-trending tendencies. Having detected coefficient instability over time, this additional information is incorporated into the forecasting models.

TABLE VII

Farley–Hinich Test For Parameter Stability

Variable	ECM1	ECM2	ECM3	ECM4
	Panel A: Constrained Cash and Futures Gold–Silver Spreads (Parities)			
$T \cdot z_{t-1}^{c}$	0.0079(3.451)[a]	−0.00043(−2.858)[a]	−0.00075(−3.351)[a]	0.00004(2.789)[a]
$T \cdot \Delta P_{ft-1}^{c}$	−0.0003(−1.407)	0.00038(2.535)[b]	−0.00044(−1.929)	0.00038(2.439)[b]
$T \cdot \Delta P_{f,t-2}^{c}$	0.0000(0.111)	0.00001(0.425)	−0.00001(−0.138)	0.0004(0.398)
$T \cdot \Delta P_{c,t-1}^{c}$	−0.0000(−0.167)	−0.00017(−1.012)	0.00001(0.434)	−0.00002(−0.976)
$T \cdot \Delta P_{c,t-2}^{c}$	−0.0000(−1.112)	−0.00003(−0.613)	−0.00001(−0.877)	−0.00001(−0.559)
	Panel B: Unconstrained Cash and Futures Gold–Silver Spreads			
	ECM5	*ECM6*	*ECM7*	*ECM8*
$T \cdot z_{t-1}^{u}$	0.00094(4.094)[a]	−0.00006(−4.269)[a]	−0.00010(−4.037)[a]	0.00007(4.231)[a]
$T \cdot \Delta P_{ft-1}^{u}$	−0.00067(−3.080)[a]	0.00005(3.576)[a]	−0.00006(0.203)	0.00005(3.520)[a]
$T \cdot \Delta P_{f,t-2}^{u}$	0.00002(0.142)	0.00001(0.701)	0.00001(−3.013)[a]	0.00001(0.644)
$T \cdot \Delta P_{c,t-1}^{u}$	−0.00004(−2.302)[b]	−0.00002(−1.316)	−0.00004(−2.292)[b]	−0.00002(−1.341)
$T \cdot \Delta P_{c,t-2}^{u}$	0.00001(0.242)	0.00001(0.239)	0.00001(0.224)	−0.00001(−0.717)

Notes:
[a] and [b]Denote, respectively, statistical significance at the 1% and 5% levels.

[22]Because the results and conclusions from estimating error correction models without time trending coefficients (Table V) are essentially unchanged when appending time trend terms (insofar as error corrections and contemporaneous and lagged interactions are concerned), only new information concerning the estimates of the newly created coefficients are reported.

Out-of-Sample Fit and Forecast Evaluation

As mentioned earlier, the forecast strategy calls for generating one week multistep forecasts of the spread. This results in 207 weekly forecast periods. To account for the fact that some of the parameters are not constant over time, the minimum FPE error correction models are extended to include time trend terms, as previously demonstrated in the Farley–Hinich test. Presumably, the predictable parameter behavior provides useful information for updating the forecasts.

In judging the accuracy of forecasts generated by fixed versus time-varying parameter versions of the competing minimum FPE error correction models, three error statistics are used: mean forecast error (ME), mean absolute forecast error (MAE), and root mean square error (RMSE), as follows:

$$\mathrm{ME}_j^i = (1/5N)\sum_{n=1}^{N}\sum_{m=1}^{5}(\Delta\hat{P}_{j,t+m}^i - \Delta P_{j,t+m}^i)$$

$$\mathrm{MAE}_j^i = (1/5N)\sum_{n=1}^{N}\sum_{m=1}^{5}\left|(\Delta\hat{P}_{j,t+m}^i - \Delta P_{j,t+m}^i)\right|$$

$$\mathrm{RMSE}_j^i = \left\{(1/5N)\sum_{n=1}^{N}\sum_{m=1}^{5}(\Delta\hat{P}_{j,t+m}^i - \Delta P_{j,t+m}^i)^2\right\}^{1/2}$$

where $\Delta\hat{P}_{j,t+m}^i$ is the daily ($m = 1, 2, \ldots, 5$) forecasted change in the jth spread ($j = 1, 2, 3$; with 1 denoting the constrained gold–silver spread, 2 denoting the unconstrained gold–silver spread, and 3 denoting the unconstrained silver–gold spread) for the ith market ($i = 1, 2$; with 1 denoting the cash market, and 2 denoting the futures market). N is the number of weekly forecast periods, and thus $5N$ is the number of daily forecasts. For space considerations, Table VIII contains the forecast error statistics for only the constrained and unconstrained gold–silver spreads.

Several observations are warranted. First, there is a tendency for a time-varying parameter version to outperform a fixed coefficient version of the forecasting model. This is hardly surprising since few time varying parameters are detected, and this additional information is incorporated into the forecasting procedure.[23] Second, a predictive error correction model of, for example, the change in the constrained gold–silver futures

[23]Obviously, other stochastically varying parameter models (of which a random coefficients model is a special case) may be capable of outperforming a simple deterministic linear trend model—this area is left for future research.

TABLE VIII
Out-of-Sample Forecast Error Statistics

	Panel A: Constrained Cash and Futures Gold–Silver Spreads (Parities)							
	ECM1		*ECM2*		*ECM3*		*ECM4*	
Statistic	*FC*	*TVP*	*FC*	*TVP*	*FC*	*TVP*	*FC*	*TVP*
ME	0.057	0.053	−0.036	−0.031	−0.030	−0.036	0.024	0.030
MAE	0.800	0.730	0.600	0.520	0.800	0.684	0.600	0.600
RMSE	1.280	1.060	0.900	0.800	1.280	0.960	0.903	0.905
	Panel B: Unconstrained Cash and Futures Gold–Silver Spread							
	ECM5		*ECM6*		*ECM7*		*ECM8*	
	FC	*TVP*	*FC*	*TVP*	*FC*	*TVP*	*FC*	*TVP*
ME	0.050	0.050	−0.024	−0.022	0.023	0.019	−0.019	−0.017
MAE	0.539	0.530	0.414	0.410	0.531	0.439	0.404	0.402
RMSE	0.901	0.894	0.629	0.601	0.900	0.805	0.628	0.594

Notes:
All numbers are multiplied by 100.
FC and TVP denote, respectively, fixed coefficients model and time varying parameter model.
ME denotes the mean error, MAE is the mean absolute error, and RMSE is root mean square error.
All error measures are computed from the changes in logarithms of the spreads.

spread that uses lagged equilibrium pricing errors in the cash spread as the error correcting mechanism (ECM3) does better in predictions than one that uses "own" lagged equilibrium pricing errors (i.e., in the futures spread) as the error correcting mechanism (ECM1) if parameters are allowed to vary. By the same token, a predictive model of changes in the cash spread that conditions such prediction on lagged equilibrium pricing errors in the futures spread (ECM2) will outperform one that uses "own" lagged pricing errors as the error correction mechanism (ECM4) when parameters are, similarly, allowed to vary. This is a clear manifestation of: (i) the importance of considering the behavioral linkages between the cash and futures markets for precious metals; and (ii) the benefit of incorporating additional information about the intertemporal behavior of the parameters into a forecasting model. Similar observations can be made based on the results in panel B.

Simulated Trading Results

Table IX reports trading profits generated by acting on out-of-sample predicted changes in the daily precious metal spread for a few selected error correction models exhibiting the best in-sample fit. Trading profits are reported in dollar terms for only the constrained and unconstrained

TABLE IX

Cumulative Dollar Trading Results Based on Selected Error Correction Models for Constrained and Unconstrained Cash and Futures Gold–Silver Spreads
Trading Period: July 1, 1988–July 31, 1992

		Constrained Gold–Silver Spread				*Unconstrained Gold–Silver Spread*			
		ECM2		*ECM3*		*ECM6*		*ECM7*	
Contract		*FC*	*TVP*	*FC*	*TVP*	*FC*	*TVP*	*FC*	*TVP*
August	1988	$(1186)	$(916)	$42	$104	$298	$(855)	$74	$170
October	1988	1145	450	(126)	(16)	(895)	(345)	(222)	(510)
December	1988	(1381)	(356)	281	(207)	(1961)	(705)	494	1135
February	1989	(685)	(559)	(345)	22	167	285	(605)	(1390)
April	1989	112	202	19	107	259	315	33	75
June	1989	1275	1487	394	(51)	487	(670)	692	1590
August	1989	(1018)	(459)	(281)	(192)	(459)	185	(494)	(1135)
October	1989	(555)	(496)	168	(18)	(17)	410	296	680
December	1989	1845	651	631	280	(2372)	(1820)	(1029)	1545
February	1990	(1741)	321	270	115	(1044)	(400)	474	1090
April	1990	1261	476	(586)	(61)	385	645	1108	(1365)
June	1990	(1744)	(1243)	(336)	(159)	1360	905	(590)	1290
August	1990	1355	1597	1312	578	936	795	2303	1005
October	1990	188	429	497	162	1034	1105	873	(365)
December	1990	(1205)	(1078)	(90)	34	270	410	(159)	(460)
February	1991	(1587)	1160	(610)	(228)	1160	1665	243	10
April	1991	(191)	422	(245)	237	479	970	(431)	95
June	1991	(191)	(505)	23	(75)	(505)	(360)	(1071)	560
August	1991	737	51	139	41	(11)	(25)	41	620
October	1991	(424)	692	154	123	692	90	(500)	(1150)
December	1991	(1042)	(1656)	(285)	182	1826	945	(367)	(845)
February	1992	309	72	(209)	54	363	(285)	270	390
April	1992	832	241	97	(15)	59	(195)	170	(1355)
June	1992	(746)	(74)	382	(3)	100	150	670	1540
Total		$621	$1369	$1296	$1014	$2611	$3215	$2273	$3220
Std. Dev.		$103	$103	$207	$158	$143	$140	$218	$193

Notes:
FC and TVP denote, respectively, fixed coefficient and time-varying parameter error correction models.
Std. Dev. is the standard deviation of daily variation margin flows.

gold–silver cash and futures spreads based on trading, always, in the nearby futures contract for liquidity considerations. Details of trading results for the unconstrained silver–gold cash and futures spreads are not reported but are available upon request.

Before reporting these results, two additional observations are worth noting. First, although gold futures contracts traded on the COMEX expire regularly on a bimonthly basis, silver futures contracts traded on the same exchange exhibit an irregular expiration cycle. For example, between July 1988 and the end of July 1992, 24 nearby gold futures contracts were traded, while only 20 nearby futures contracts on silver were traded over the same forecast period. During this time period the expiration cycle of silver futures varied between quarterly and bimonthly expiration frequencies. Therefore, since gold futures expire more frequently relative to silver futures, and since gold and silver positions have to be established simultaneously, the trading strategy is anchored to the gold futures expiration cycle. Second, because the profitability of trading the precious metal spread in the cash market is also investigated, these results are also reported on a bimonthly basis corresponding to the gold futures expiration cycle so as to make the trading results fairly comparable. Because a COMEX futures contract on gold calls for delivery of 100 troy ounces of gold while a silver futures contract traded on the same exchange calls for delivery of 5000 ounces of silver, daily dollar profits (or losses) on the combined position, for the case of positive predicted changes in the spread, are computed as follows:

$$\$\text{Profit} = \{100 \times (G_t - G_{t-1}) + 5000(S_{t-1} - S_t)\}$$

where G_t and S_t refer to gold and silver prices at time t, respectively. On the other hand, for a negative predicted change in the spread, the daily change in wealth over period t is computed as follows:

$$\$\text{Profit} = \{100 \times (G_{t-1} - G_t) + 5000(S_t - S_{t-1})\}$$

The results are presented in Table IX on a contract-by-contract basis for selection error correction models based on the fixed coefficient (FC) and time-varying parameter (TVP) versions of these models. Additionally, flows are presented on a pre-tax and pre-transactions cost basis. As shown in Table IX, gross profits and losses occur from one contract to the next, and vary depending on the spread that is being traded and the underlying assumptions about the intertemporal behavior of the parameters. These figures assume only one contract of each maturity is traded in each transaction for either gold or silver.

Trading based on the constrained gold–silver spread generated the lowest gross profits, albeit with the lowest risk, as measured by the standard deviation of daily changes in wealth. For example, trading based on error correction model ECM2, which generates predicted changes in the constrained gold–silver cash spread, resulted in a gross profit across the entire four-year period of \$621 on the combined gold and silver positions. It is also readily apparent that allowing for coefficients to vary over time does improve forecasting performance; profits doubled with no increment to the daily standard deviation of changes in the trader's wealth. On the other hand, the error correction model ECM3, which generates predicted changes in the constrained gold–silver futures spread, indicates gross trading profits over the entire period of \$1296 on the combined position. The fixed coefficient version of the model suggests higher trading profits can be made, in comparison to the time-varying parameter version, albeit with a higher level of risk in daily variation margin flows.[24]

An examination of the results generated from a trading strategy predicated on error correction models ECM6 and ECM7 (used in predicting changes in the unconstrained gold–silver cash and futures spreads, respectively) reveals a more encouraging scenario. Three comments are in order. First, the models do better in forecasting the unconstrained gold–silver cash and futures spreads, rather than their constrained counterparts, in an out-of-sample context. Gross trading profits have more than doubled, without doubling the risk level involved in trading the unconstrained spread. Second, a simple allowance for temporal variation in the parameters further improves the trading performance of the models. Third, trading the unconstrained gold–silver spread in the cash market (ECM6) yields similar results to trading the spread in the futures market (ECM7), especially when trading is based on the time-varying parameter versions of the models. Given the transactions cost advantage of futures markets, and the ease with which short sales can be established in futures markets, this evidence is indeed encouraging. Furthermore, this result does not come as a surprise given the cross-correlation results reported earlier. Finally, the results from trading the unconstrained silver–gold spread (not reported) suggest that higher profits can be made by trading the cash spread (ECM10) rather than by trading the futures spread (ECM11). Additionally, trading risks in the form of uncertainty about daily variation margin flows are also the highest among all three alternative definitions of the spread.

To shed more light on the economic significance of these results, Table X presents frequency distributions of the alternative trading signals issued by the selected error correction models for the three spread series. Four trading signals are possible when simultaneously considering the implications of the forecast strategy together with the trading strategy. These are: (i) buy gold and short silver (if the predicted change in the gold–silver spread is positive); (ii) continue to hold a long position in gold and a short position in silver (if the subsequent

[24] It is very possible that a time-trending parameter model is not a suitable parameterization of the intertemporal behavior of the parameters of a model that forecasts changes in the constrained gold–silver futures spread. An alternative characterization of this behavior is needed.

predicted change in the gold–silver spread is positive, or if negative, is less than the transactions costs); (iii) sell gold and buy silver, thereby reversing the initial position (if the predicted change in the gold–silver spread is negative); and finally (iv) continue to hold a short position in gold coupled with a long position in silver (if the predicted change in the gold–silver spread is negative, or if positive, is less than the transactions costs incurred in reversing the previous position).

Marginal transactions costs in the futures market are estimated at $15 for gold and silver futures contracts per one-way trip, whereas transactions costs in the cash market are estimated at $30 per transaction involving the same amount of gold and silver. Because the models generate percentage predicted changes in the spread, the dollar transactions cost is transformed to a percentage cost.[25] Since there are two passive signals and two active trading signals, the economic significance of the results presented in Table IX can be assessed by adding up the frequencies of the active trading signals and multiplying the sum by the assumed dollar transactions cost. The resulting figure can then be compared to the gross dollar profits reported in Table IX to obtain net profits after transactions costs.

The frequency distributions for four possible trading signals are presented in Table X for both the fixed coefficient and variable coefficient versions of the forecasting models. The four trading signals entail two "active" buy or sell signals (A and C) and two "passive" hold signals (B and D). For example, using the fixed coefficient version of ECM2, the reported frequencies suggest that over the four-year forecast horizon, a long gold–short silver (i.e., a long spread) trading signal would have been issued 154 times (signal A) while a short gold–long silver (short spread) signal would have been issued by the model 118 times (out of the total of 995 times). On the other hand, a "hold" signal of long gold–short silver would have been issued by the same model 413 times while a "hold" signal of short gold–long silver would have been given 310 times. Because it is the total number of active signals (A and C) that determines transactions costs, one needs to compare the dollar cost of trading based on the total number of these active signals with the profit figures reported in Table IX. The economic implications are clear. The profit results shown in Table IX suggest that on a

[25]This transformation is accomplished as follows:

$$\%\tau = \{(\$\tau/(G_t \times 100)) + (\$\tau/(S_t \times 5000))\}$$

where τ is transactions cost, G_t is the gold price in period t, and S_t is the silver price in the same period.

TABLE X

Frequency Distribution of Trading Signals; Forecast Period: July 1, 1988–July 31, 1992

	Panel A: Constrained Gold–Silver Spread															
	ECM2								*ECM3*							
	FC				*TVP*				*FC*				*TVP*			
Frequency	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
# of cases	154	413	118	310	178	401	126	290	361	133	155	345	371	116	153	355
(%) cases	(15.4)	(41.5)	(11.8)	(31.3)	(17.9)	(40.3)	(12.6)	(29.2)	(36.1)	(13.3)	(15.6)	(35.0)	(37.5)	(11.7)	(15.4)	(35.4)
	Panel B: Unconstrained Gold–Silver Spread															
	ECM6								*ECM7*							
	FC				*TVP*				*FC*				*TVP*			
# of cases	185	334	177	299	180	333	179	303	186	315	186	308	186	317	179	314
(%) cases	(18.6)	(33.5)	(17.8)	(30.1)	(18.1)	(33.5)	(18.0)	(30.4)	(18.7)	(31.6)	(18.7)	(31.0)	(18.7)	(31.8)	(17.9)	(31.6)
	Panel C: Unconstrained Silver–Gold Spread															
	ECM10								*ECM11*							
	FC				*TVP*				*FC*				*TVP*			
# of cases	292	153	128	422	298	143	128	425	340	151	138	366	355	147	122	391
(%) cases	(29.4)	(15.4)	(12.9)	(42.3)	(29.9)	(14.4)	(12.8)	(42.9)	(34.1)	(15.2)	(13.9)	(36.8)	(33.8)	(14.8)	(12.3)	(39.1)

Notes:
FC and TVP denote, respectively, fixed coefficient and time-varying parameter error correction models.
A denotes a signal to buy gold and sell silver.
B denotes a signal to "hold" onto a long gold-short silver position, i.e., no trading action required.
C denotes a signal to sell gold and buy silver.
D denotes a signal to "hold" onto a short gold-long silver position, i.e., no trading action required.
Interpretation of above trading signals is reversed for panel C, when trading the unconstrained silver–gold spread.

pre-transactions cost basis, a trader who uses ECM2 in its fixed coefficient version as the basis for trading the constrained gold–silver spread in the cash market would realize a dollar profit figure of \$621. Now, if one adds up the number of active trading signals (A and C) underlying the fixed coefficient version of ECM2, and multiplies the sum by \$30 per transaction for a futures-equivalent cash market position in gold and silver, a total cost of \$8160 is obtained as a result of implementing the model's trading recommendations. On a post-transactions cost basis, the reported gross dollar profit of \$621 is, therefore, clearly insufficient to cover even a small fraction of the cost of transacting for the ordinary individual trader. Alternatively, one can determine a breakeven level of transactions cost beyond which active trading (based on the model's recommendations) would result in net losses on a post-transactions cost basis. Essentially, the fixed coefficient version of ECM2 suggests that 272 active trades would have been issued by the model to a trader who trades the gold–silver spread in the cash market on a constrained basis (the sum of trading signals A and C). At a total profit level of \$621, the breakeven transactions cost level on the combined gold–silver position should not have exceeded \$2.28 (or less) per transaction to yield zero (or positive) trading profits. On the other hand, the variable coefficient version of ECM2 issued 304 active trading signals (signals A and C) which enabled traders to accrue gross profits of \$1369 per one futures-equivalent position in each of gold and silver. Therefore, as long as the trader's transactions cost on the combined position does not exceed \$4.50 per transaction, the reported gross profits may indeed yield positive dollar profits.

The moral of the story is clear. The reported gross profits are not necessarily net positive profits unless expected transactions costs faced by a trader fall below some breakeven level. At the assumed transactions cost levels, the reported gross profits result in net losses which are conditioned by both the model used to generate predicted changes in the precious metal spread, and the concommitant trading signals.

A notable difference between the results reported under ECM2 and those reported under ECM3 is that the number (and therefore, the percentage) of "active" trading signals issued in the futures market (ECM3) is more than double that issued in the cash market. This comes at the expense of the number of passive "hold" signals (B and D), which is much less when trading the constrained gold–silver spread in the futures market rather than in the cash market. Therefore, a model for trading the precious metals spread in the futures market issues far more active than passive trading signals, whereas a model for trading the spread in the cash market exhibits a reverse signaling pattern.

The evidence in panels B and C confirms a similar picture when considering the gross dollar profits from trading the unconstrained gold–silver spread (panel B) or the unconstrained silver–gold spread (panel C), coupled with the number of active trading signals (signals A and C). Net profits on a post-transactions cost basis are very difficult to realize. For example, cash market trading based on the error correction model (ECM6) of the unconstrained gold–silver spread generates 362 active signals (for the FC version) and 359 active signals (for the TVP version). The respective gross dollar profits (from Table IX) are clearly insufficient to yield positive net profits at the assumed transactions cost level in the cash market. By similar reasoning, trading based on the unconstrained spread in the futures market would result in dollar losses on a post-transactions cost basis, even when considering the lower costs of transacting in the futures markets. It is worth noting that the unconstrained gold–silver spread allows for a higher breakeven level of transactions cost (based on the reported gross profits and resulting "active" trading frequencies), which in turn increases the chance of realizing positive net profits in comparison to trading based on predictable changes in the constrained gold–silver spread. A similar interpretation applies to trading the unconstrained silver–gold spread.

Overall, the simulated trading results suggest that the short-term memory observed within and across the two precious metal price series could probably be exploited to generate positive trading profits on a pre-tax and, more importantly, on a pre-transactions cost basis. On a post-transactions cost basis, however, positive trading profits are not possible for the ordinary market participant. In fact, the average breakeven level of transactions cost turns out to be about $7 per transaction in the cash or futures markets, which involves one gold futures and one silver futures contract if trading is undertaken in the futures market, or their equivalent if trading is undertaken in the cash market. To what extent actual transactions costs faced by traders conform with this average figure needs to be interpreted by each trader contingent on his or her own particular situation.

Furthermore, it should be noted that the trading results are predicated on closing prices when, in actual fact, the forecasted change would have to be based on a price quoted at least few minutes prior to closing on the day of the forecast origin. This would be necessary to allow time for updating the forecast and implementing the desired trade, at least for the first day of each forecast period.

In addition, the gross profit figures are not adjusted for systematic risk (if any) of trading the precious metal spread. In essence, reported profits may not necessarily represent "excess" profits if they are simply a compensation for possible non-diversifiable risk associated with a precious metals position. To investigate this possibility, the following augmented single index market models are estimated for the constrained and unconstrained cash and futures spreads:

$$\Delta P_t^i - R_f = \alpha^i + \beta^i(R_m - R_f) + \epsilon_t^i - \theta_1 \epsilon_{t-1}^i$$

where ΔP_t^i, R_f, and R_m are the logarithmic first differences of the ith spread series, R_f is the riskless return on a one-month T-bill, and R_m is the return on the S&P 500 index, β^i is a measure of the systematic risk of the ith spread, α^i is a measure of excess return, and a first-order moving average term is included to account for short horizon autocorrelations in the spreads. The regression results are (with t-statistics in parentheses):

$$\Delta P_{f,t}^c - R_f = \underset{(9.719)^{**}}{0.00054} + \underset{(0.517)}{0.0062}(R_m - R_f) + \epsilon_{f,t}^c - \underset{(-8.676)^{**}}{0.284}\epsilon_{f,t-1}^c$$

$$\Delta P_{c,t}^c - R_f = \underset{(8.892)^{**}}{0.00048} + \underset{(0.134)}{0.0028}(R_m - R_f) + \epsilon_{c,t}^c - \underset{(-9.707)^{**}}{0.317}\epsilon_{c,t-1}^c$$

$$\Delta P_{c,t}^u - R_f = \underset{(13.774)^{**}}{0.00057} + \underset{(0.338)}{0.0049}(R_m - R_f) + \epsilon_{c,t}^u - \underset{(-2.638)^{*}}{0.086}\epsilon_{c,t-1}^u$$

$$\Delta P_{f,t}^u - R_f = \underset{(14.332)^{**}}{0.00066} + \underset{(0.368)}{0.0054}(R_m - R_f) + \epsilon_{f,t}^u - \underset{(-2.356)^{*}}{0.077}\epsilon_{f,t-1}^u$$

The statistical significance of the coefficients is indicated by * (5%) or ** (1%). The intercept terms imply annualized excess percentage returns of about 11.5% to about 16% (with daily compounding). The β coefficients are all statistically insignificant at any reasonable level, and are of extremely negligible magnitude. Any risk appears to be totally idiosynchratic, and thus entirely diversifiable within a stock index portfolio context. This may not be surprising considering the fact that a spread position entails long and short positions in the precious metals so that even if both metals load positively (or negatively) onto the market index, the net systematic risk of the spread may be negligible. Furthermore, these potential excess returns are certainly of a much smaller order of magnitude compared to excess returns reported in, for example, Ma (1985). Employing a simple moving average rule, and depending on the moving average order, Ma reported possible "excess" returns between 100% to 147% per year, over a 13-year period ending in 1984.

A Simple Trading Rule

This section evaluates the forecasting performance of a simple trading rule: the moving average rule. This rule is examined for two reasons. The first is to provide a comparison between trading results generated from the relatively more complex error correction models and simple moving average models. If similar or superior trading results can be obtained from simple moving average decision rules, then the cost of employing relatively more complex modeling procedures may not be justified. Second, a moving average model was employed by Ma (1985) and Ma and Soenen (1988) who investigated the profitability of spreading between gold and silver (cash and futures) markets, and found that substantial profits could be made over the period extending from the early seventies through 1986.

An infinite number of variations are possible, and the moving average rule is often modified by introducing a band around the moving average, which reduces the number of active trading signals. In its simplest form, the moving average rule is expressed as: buy gold–sell silver if the actual spread is below the moving average, and vice versa if the actual spread is above the moving average. If a neutral band is imposed, the above rules are modified as follows: buy gold–sell silver if the actual spread is below the lower bound, and vice versa if the actual spread is above the upper bound.

The following variations are employed: 5-day, 6-day, 7-day, 28-day, and 44-day moving averages. In addition, the following band-widths are imposed: zero standard deviation (corresponding to a situation of daily trading), one quarter of a standard deviation, one half of a standard deviation, three quarters of a standard deviation, and one standard deviation (all of which correspond to a situation of infrequent trading). Results based on a zero standard deviation band allow for a comparison with results obtained from daily trading based on error-correction model forecasts, while results based on bands of varying widths allow for a reasonable comparison with the moving average decision rules employed by Ma (1985) and Ma and Soennen (1988). Trading results are based on moving averages computed from the logarithmic levels and first-differences of the constrained and unconstrained cash and futures spreads. These results are presented in Table XI and cover the same four-year trading period. To conserve space, results for only the extreme band-widths (zero and one standard deviation bands) are reported.

The results in Table XI are striking. Overwhelmingly, all moving average rules examined lead to cumulative trading losses for the constrained gold–silver spread, and for the most part, the unconstrained gold–silver spread as well. The only exception appears to be a moving average model of the first-differences of the unconstrained spread, and only if a 1% band around the moving average (which limits the number of unnecessary trades) is imposed. Cumulative profits over the four-year period range from \$1750–\$3695 on the spread, depending on the order of the moving average. These figures appear to be comparable to dollar profits reported in Table IX from trading the unconstrained spread, and in some cases, they surpass them. However, one must recall that a direct, clean comparison between these profit figures may be somewhat inappropriate given that figures reported in Table IX correspond to daily (frequent) trading, while profits shown in Table XI result from infrequent trading due to the imposition of a confidence band. Also,

TABLE XI
Cumulative Dollar Trading Results Based on Moving Average Rules for Constrained and Unconstrained Gold–Silver Spreads
Trading Period: July 1, 1988–July 31, 1992

	Panel A: Constrained Gold–Silver Spread							
	(i) Levels				*(ii) First Differences*			
	Cash		*Futures*		*Cash*		*Futures*	
Rule	*0.00SD*	*1.00SD*	*0.00SD*	*1.00SD*	*0.00SD*	*1.00SD*	*0.00SD*	*1.00SD*
MA(5)	−139,105	−7,065	−134,835	−7,145	−210,095	−193,285	−187,525	−174,505
MA(6)	−129,965	−7,465	−125,810	−7,090	−209,695	−187,135	−187,470	−167,390
MA(7)	−118,565	−7,695	−109,805	−7,705	−209,465	−179,605	−186,705	−163,215
MA(28)	−74,615	−6,835	−74,435	−6,765	−199,305	−177,015	−177,245	−155,105
MA(44)	−40,740	−5,750	−8,835	−5,675	−193,600	−162,940	−173,845	−153,825
	Panel B: Unconstrained Gold–Silver Spread							
MA(5)	−177,390	−108,875	−174,505	−105,885	−172,095	3065	−143,965	3145
MA(6)	−169,605	−98,715	−168,215	−95,810	−162,925	3465	−150,450	3490
MA(7)	−165,505	−79,805	−162,940	−85,305	−177,925	3695	−147,745	3705
MA(28)	−113,805	−48,315	−105,885	−37,395	−176,395	2835	−151,335	2765
MA(44)	−65,810	−13,600	−68,905	−13,445	−173,890	1750	−153,085	1675

Notes:
MA denotes moving average. Numbers in parentheses indicate moving average order.
SD is standard deviation. This controls the width of the band (confidence level) around the moving average.

although risk associated with spreading the precious metals appears to be totally diversifiable, it is still of interest to note that the average standard deviation associated with the daily profits (losses) generated by the moving average rules for the unconstrained spread is about $329, which is more than double the comparable figure that results from acting upon the error correction model predictions for the same spread. This suggests that the risk-return tradeoff offered by these simple moving average rules is unattractive. Of course, the higher risk associated with such moving average models stems from the limited (univariate) information set that is utilized. On the other hand, the error correction models utilize a broader (multivariate) information set, potentially resulting in lower mean square prediction errors, and hence lower volatility of daily flows.

SUMMARY AND CONCLUSIONS

This article examines the relationship between the gold and silver cash and futures markets using daily prices with specific interest in investigating the intertemporal dependence and predictability of the so called gold–silver spread. A no arbitrage model linking the cash and futures markets levels of the precious metal spread is developed and the statistical properties of each spread series, taken separately as well as those of a linear combination of the cash and futures spread series, are examined and exploited within the framework of the cointegration-based error correction model to generate out-of-sample predicted changes in the cash and futures markets spreads. Alternative definitions of the precious metals spread together with alternative formulations of the forecast strategy are simulated. The findings are summarized as follows. First, based on in-sample estimates of the cross-correlations between the alternatively defined measures of the precious metals spread, trading results are likely to be sensitive to the way the spread is defined. Second, unit roots are reliably detected in each cash and futures spread series, but a linear combination of the cash and futures spreads is stationary; thus, an error correction representation for each series is appropriate. Third, although the spot and futures markets appear to be mostly simultaneously interrelated, lagged interactions reliably exist, attesting to the predictability of returns from spreading the precious metals. Fourth, trading profits that are positive on a pre-transactions cost basis become losses on a post-transactions cost basis if the assumed levels of transactions costs in the cash and futures markets are representative for the typical spread trader. Fifth, moving average trading rules result in striking losses for the constrained and unconstrained cash and futures spreads. The only exception is when one acts upon rather infrequent trading signals issued by a moving average model of the unconstrained spread. Profits are possible albeit at a high risk level. These results are, of course, conditional on the parameterizations adopted. Furthermore, they are conditional on daily trading. Thus, an interesting avenue for future research is to refine the model's parameterizations and examine the potential for ex-ante profitable arbitrage on an intra-daily basis. Moreover, the results are predicated on the presumption of a constant cross-elasticity coefficient linking the two precious metals price series. It would be of interest to examine the impact of alternative stochastic representations of this coefficient on the intertemporal predictability and ex-ante arbitrage profitability involving precious metal spreads.

BIBLIOGRAPHY

Akaike, H. (1969a): "Statistical Predictor Identification," *Annals of the Institute of Statistical Mathematics,* 21:203–217.

Akaike, H. (1969b): "Fitting Autoregressions for Prediction," *Annals of the Institute of Statistical Mathematics,* 21:243–247.

Chan, M. L., and Mountain, D. C. (1988): "The Interactive and Causal Relationships Involving Precious Metal Price Movements," *Journal of Business and Economic Statistics,* 1:69–77.

DeBondt, W., and Thaler, R. (1985): "Does the Stock Market Overreact?" *Journal of Finance,* 40:793–805.

Dickey, D. A., and Fuller, W. A. (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of American Statistical Association,* 74:427–431.

Dickey, D. A., Fuller, W. A., and Pantula, S. (1987, October): "Determining the Order of Differencing in Autoregressive Processes," *Journal of Business and Economic Statistics,* 4:455–461.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica,* 50:987–1008.

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

Engle, R. (1991): *Long-Run Economic Relationships: Readings in Cointegration,* New York: Oxford University Press.

Engle, R. (1988): "Some Recent Developments in a Concept of Causality," *Journal of Econometrics,* 39:199–211.

Engle, R. (1986): "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics,* 48: 213–228.

Engle, R., Ito, T., and Lin, W. (1990): "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market," *Econometrica,* 58:525–542.

Fama, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance,* 25:383–417.

Farley, J. U., and Hinich, M. J. (1970): "A Test For a Shifting Slope Coefficient in a Linear Model," *Journal of the American Statistical Association,* 65:1320–1329.
Farley, J. U., and McGuire, T. W. (1975): "Some Comparisons of Tests For a Shift in the Slopes of a Multivariate Linear Time Series Model," *Journal of Econometrics,* 3:297–318.
French, K. R., and Roll, R. (1986): "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics,* 17:5–26.
Fuller, W. A. (1976): *Introduction to Statistical Time Series,* New York: Wiley.
Garbade, K. D., and Silber, W. L. (1982): "Price Movements and Price Discovery in Futures and Cash Markets," *Review of Economics and Statistics,* 65:289–297.
Granger, C. (1969): "Investigating Causal Relations by Econometric Models and Cross Spectral Methods," *Econometrica,* 37:424–438.
Granger, C., and Newbold, P. (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics,* 2:111–120.
Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., and Lee, T. C. (1985): *The Theory and Practice of Econometrics,* New York: Wiley.
Ma, C. K. (1985): "Spreading between the Gold and Silver Markets: Is There a Parity?" *The Journal of Futures Markets,* 5:579–594.
Ma, C. K., and Soenen, L. A. (1988): "Arbitrage Opportunities in Metal Futures Markets," *The Journal of Futures Markets,* 8:199–209.
MacKinnon, J. G. (1991): "Critical Values for Cointegration Tests," in *Long Run Economic Relationships: Readings in Cointegration,* Engle, R. F., and Granger, C. W. J., (ed.), Oxford University Press, pp. 267–276.
Monroe, M. A., and Cohn, R. A. (1986): "The Relative Efficiency of the Gold and Treasury Bill Futures Markets," *The Journal of Futures Markets,* 3:477–493.
Ohanian, L. E. (1988): "The Spurious Effects of Unit Roots on Vector Autoregressions: A Monte Carlo Study," *Journal of Econometrics,* 39:251–266.
Phillips, P. C. B., and Perron, P. (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika,* 75:335–346.
Stoll, R. H., and Whaley, R. E. (1990): "The Dynamics of Stock Index and Stock Index Futures Returns," *Journal of Financial and Quantitative Analysis,* 25:441–468.
Summers, L. (1986): "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance,* 41:591–600.
Swamy, P. A., and Schinasi, G. J. (1989): "Should Fixed Coefficients Be Reestimated Every Period for Extrapolation?" *Journal of Forecasting,* 8:1–17.
Theil, H. (1971): *Principles of Econometrics,* New York: Wiley.