Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 600

First version: December 1997

Revised version: September 1998

EVALUATING FORECASTS OF CORRELATION USING OPTION PRICING

Michael S. Gibson and Brian H. Boyer

EVALUATING FORECASTS OF CORRELATION USING OPTION PRICING

Michael S. Gibson and Brian H. Boyer*

Abstract: A forecast of the correlation between two asset prices is required to price or hedge an option whose payoff depends on both asset prices or to measure the risk of a portfolio whose return depends on both asset prices. However, a number of factors make it difficult to evaluate forecasts of correlation. We develop a forecast evaluation methodology based on option pricing, extending a technique that Engle et al (1993) introduced to evaluate volatility forecasts. A forecast of the variance-covariance matrix of joint asset returns is used to generate a trading strategy for a package of simulated options. The most accurate forecast will produce the most profitable trading strategy. The package of simulated options can be chosen to be sensitive to correlation, to volatility, or to any arbitrary combination of the two. In an empirical application, we focus on the ability to forecast the correlation between two stock market indices. We compare the correlation forecasting ability of three more sophisticated models (two GARCH models and a two-state Markov switching model) and two simple moving averages. We find that the more sophisticated models produce better correlation forecasts than the simple moving averages.

Keywords: forecast evaluation, GARCH, Markov switching, weekend effect

* We thank Stephen Figlewski and participants in the First Semiannual Washington Area Finance Conference, the 8th Annual Derivatives Securities Conference, and workshops at the University of Virginia and the Federal Reserve Board for helpful comments and suggestions. Remaining errors are the responsibility of the authors. The first author is a staff economist in the Division of International Finance. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. The authors' email addresses are gibsonm@frb.gov and boyerb@umich.edu.

1 Introduction

A forecast of the correlation between two financial asset prices is required to price or hedge an option whose payoff depends on both asset prices or to measure the risk of a portfolio whose return depends on both asset prices. However, a number of factors make it difficult to evaluate forecasts of correlation. First, correlation, like volatility, is unobservable. Second, even if true correlation could be observed, the appropriate loss function is subject to uncertainty. Forecasts which minimize standard statistical loss functions may not minimize economic loss functions. Third, most models that forecast correlation also forecast volatility. A test of such a model is a joint test, not a test of the ability to forecast correlation. To circumvent these difficulties, we propose a methodology to evaluate forecasts of correlation based on option pricing. Each correlation forecast is used to generate an option trading strategy. The most accurate correlation forecast will produce the most profitable trading strategy.

Considerable attention has been given in the literature to forecasting volatility.[1,2] The most common evaluation technique is to regress squared asset returns on volatility forecasts (e.g. Lamoureux and Lastrapes (1993) and Jorion (1995)). Another technique compares the mean square error of forecasts, with forecast error measured as the difference between squared returns and the volatility forecast (e.g. West and Cho (1994) and Figlewski (1997)). Although both techniques use squared returns as the forecast benchmark and the quadratic loss function to evaluate forecast accuracy, they do not always agree on which forecast of volatility is best. Squared returns are a sensible benchmark since they are an unbiased and consistent estimate of true volatility. However, true volatility is never observed. Further, there is no natural metric to measure the dollar value of volatility forecast errors.

---

[1] We use the term 'volatility' in this paper both to refer to the general variability of a time series and as a synonym for standard deviation.

[2] See Mayhew (1995) and Figlewski (1997) for surveys.

To overcome these difficulties, Engle et al (1993) introduced a new technique to evaluate volatility forecasts based on the models' ability to accurately price options.[3]

In this paper, we extend the methodology of Engle et al (1993) to the multivariate case for the purpose of evaluating forecasts of the entire variance-covariance matrix. One strength of our methodology is that it evaluates forecasts with an economically meaningful loss function: option pricing errors. Another strength of our methodology is its ability to emphasize the accuracy of any combination of volatility and correlation forecasts, e.g. the methodology can be tailored to identify good forecasters of correlation, good forecasters of volatility, or good forecasters of correlation and volatility.

Little attention has been given in the literature to evaluating forecasts of correlation. A small but growing literature compares the forecast accuracy of implied correlation from cross-currency options with correlation forecasts based on historical pairs of returns.[4] However, these papers evaluate forecast accuracy using a statistical loss function, such as mean square error, not an economic loss function such as option pricing errors.

In the next section of the paper, we set out our methodology. We describe the steps that must be followed each day: each forecasting model makes a forecast, each forecast is used to price and trade options, and profits are realized based on the next day's actual asset price movements. We describe how our methodology can be tailored to focus on any desired combination of volatility and correlation forecasts by trading a package of options. In section 3, we explore our methodology's performance on simulated data. In section 4, we describe the five forecasting models we choose to implement in our empirical application: two GARCH models, a two-state Markov switching model,

---

[3]See also Noh, Engle and Kane (1994) and Engle, Kane and Noh (1997).

[4]Campa and Chang (1997) and Walter and Lopez (1997) are two recent examples.

and two moving averages. In section 5, we implement our methodology using data on the S&P 500 and NASDAQ composite equity indices. Section 6 concludes.

2 Evaluating forecasts by their ability to accurately price options

To evaluate forecasts of correlation, we extend the methodology of Engle et al (1993), who based their evaluation of univariate volatility forecasting models on the models' ability to accurately price options. We specify a group of forecasting "agents," each with her own forecasting model for the variance-covariance matrix of joint returns on two assets.[5] The forecasting agents price and trade an option whose payoff depends on the two asset prices (a "rainbow option"). Each forecasting agent is assumed to be risk-neutral.

For each day t in our sample, the following sequence is followed:

1. Each agent forecasts the variance-covariance matrix of joint returns between $t-1$ and t, using past returns through day $t-1$.[6]
2. Each agent uses her forecasts of volatility and correlation to price a 1-day European rainbow option whose payoff depends on the date t returns on the two assets.
3. Agents trade options among themselves. For each pair of agents, the agent whose forecast leads to a higher option price buys one option from the agent whose forecast leads to a lower option price. With n forecasting agents, each agent will make n-1 trades each day. We restrict the trade amount to a single option.[7] Each trade takes place at the mean of the two

[5]The five forecasting models we use in our empirical tests are described in Section 4 below. In the simulations presented in Section 3, the agents use static forecasts that do not require any estimation.

[6]For an analysis of correlation forecasting that uses more information than just past returns, see Karolyi and Stulz (1996).

[7]This restriction is necessary to limit the amount of trading, since two risk-neutral agents who put different values on an option would want to trade an infinite quantity of the option. A more involved alternative would be to specify each agent's risk preferences, have each agent compute the risk of each trade (incorporating the forecast uncertainty), and have agents bargain over an amount and a price for each day's trade. We have chosen the simpler alternative.

agents' prices. The "bank account" of the option-selling agent is credited with the amount of the option premium, and the "bank account" of the option-buying agent is debited the same amount.

4. Each agent hedges her exposure to both of the underlying asset prices by selling short a number of shares of each asset equal to the agent's estimate of the "delta" of her option position with respect to that asset.[8] Hedging is done by buying or selling at date *t-1* prices. Each agent's "bank account" is credited with the amount of the short sales (or debited if the hedge is a long position in the underlying asset).

5. The "bank account" balances earn one day's interest at the risk-free rate. Negative balances are charged interest at the risk-free rate.

6. Using actual date *t* returns, the rainbow option's payoff is calculated. The "bank account" of each agent who bought options is credited with the payoff amount (which is zero if the option expires out of the money). The hedge positions are sold at date *t* prices.

7. Each agent's final "bank account" balance represents her profit from the day's trading.

Forecasting agents with more accurate forecasts should, on average, earn higher profits.

2.1 Choosing an option pricing model

Our methodology requires each agent to price a 1-day rainbow option. In principle, each agent should use an option pricing model specific to the data generating process assumed by that agent. For example, an agent who forecasts assuming a stationary lognormal distribution for returns would use an extension of the Black-Scholes model to price rainbow options, an agent who forecasts assuming returns follow a GARCH process should use a GARCH option pricing model, and so on.

While it would be possible to price a European rainbow option under any assumed data generating process, either by solving a partial differential equation or using Monte Carlo techniques,

[8] The instantaneous delta is used for hedging, and the hedge is not rebalanced before the option expires.

doing so for all of the data generating processes assumed by the various forecasting agents would increase the computational burden of our methodology. For some data generating processes (i.e. those with stochastic volatility), no-arbitrage pricing methods could be inappropriate and an equilibrium model, involving additional assumptions on preferences, could be required. To reduce the computational burden of our methodology, we follow Engle et al (1993) and require all forecasting agents to use a common option pricing model--an extension of the Black-Scholes model to rainbow options, assuming stationary lognormal returns. Under the Black-Scholes assumptions, analytic solutions are available for several rainbow options and simple numerical techniques can provide solutions for other rainbow options.

Although it simplifies computation, forcing all forecasting agents to price options under the Black-Scholes assumptions will introduce some error. We believe the error will be small because all the options we consider in this paper expire in one day and are either at the money or close to the money. In addition, imposing a common option pricing model on all agents removes one of the two ways a forecasting agent could outperform her competitors: by a superior forecast of correlation and/or volatility, or by a superior option pricing model. Consequently, our results will be biased against finding any differences among forecasting agents.

2.2 The relative importance of correlation and volatility

The price of a rainbow option depends on both correlation and volatility. Time series models that produce different correlation forecasts will likely produce different volatility forecasts. This poses a problem for our methodology, if we are interested in isolating the ability to forecast correlation independently of the ability to forecast volatility.

The relative importance of volatility and correlation forecasts will depend on the use to which those forecasts will be put. Each user of a forecast will have his or her own loss function over the deviation of forecast volatilities from true volatilities and the deviation of forecast correlation from true correlation. Algebraically, these loss functions could be written in a general way as

$$L = w_1 g_1(\sigma_1^f - \sigma_1) + w_2 g_2(\sigma_2^f - \sigma_2) + w_3 g_3(\rho^f - \rho) \qquad \textbf{(1)}$$

where a superscript "f" indicates a forecast, the w_i are the weights attached to the volatility and correlation terms, and the g_i(.) are functions that are minimized at zero (*e.g.*, absolute value or quadratic). The best forecasting model for a particular application will minimize *L*, given a set of weights. The choice of weights could influence which forecasting model is identified as the "best." For example, $\{w_1 = w_2 = w_3 = 1\}$ would favor a "generalist" forecasting model that does a good job of forecasting both volatility and correlation, while $\{w_1 = w_2 = 0, w_3 = 1\}$ would favor a "specialist" forecasting model that does a good job forecasting correlation.

One way to tailor the relative importance of volatility and correlation in our methodology would be to vary the choice of rainbow option to be traded by the forecasting agents. Some rainbow options are relatively more sensitive to volatility, some to correlation. If the application of interest is pricing a particular rainbow option, choosing that rainbow option to be traded by the forecasting agents will naturally produce the correct "weights" on volatility and correlation (and would correctly specify the individual loss functions g_i).

We develop a more general way to tailor the relative importance of volatility and correlation, because we expect that the uses of correlation forecasts may be wider than simply pricing various rainbow options. In particular, a situation where volatility risk can be hedged but correlation risk cannot implies a low cost of making an incorrect volatility forecast but a higher cost of making an incorrect correlation forecast. (In (1), this would be the case of $\{w_1 = w_2 = 0, w_3 = 1\}$.) In what follows, and in our empirical work, we focus on this case, because we think it is of particular interest.

However, we stress that our methodology could be extended to handle arbitrary weights on the components of the loss function L.

To tailor the relative importance of volatility and correlation, we replace the single rainbow option with a "package" of seven options: a single 1-day rainbow option and six 1-day European call options (three on each asset). The call options are defined to have strike prices at 100%, 101% and 102% of the asset's current price.[9] We will vary the amounts of the call options contained in the package to tailor the sensitivity of the package's price to the two volatilities. Buying call options will increase the sensitivity of the package price to volatility without affecting its sensitivity to correlation, increasing the relative importance of volatility. Conversely, selling call options will reduce the sensitivity to volatility, reducing the relative importance of volatility. Forecasting agents will price and trade the package of options in place of the single rainbow option. The rest of the methodology as described above remains unchanged.

2.3 Tailoring the methodology to focus on correlation

For our "option package" methodology, we want to find a package of options whose price is sensitive to correlation but insensitive to volatility. Let $V(\sigma_1, \sigma_2, \rho; x)$ represent the price of the package of options, which is written so as to explicitly note its dependence on the volatility of the two assets (σ_1, σ_2), their correlation (ρ), and the amounts of the six call options included in the package (the 6x1 vector x).[10] We adopt the convention that the first three elements of the vector x specify the amounts of the call options on asset 1 with strike prices at 100%, 101%, 102% of the current price, and the last three elements refer to calls on asset 2. We will refer to the call option on asset 1 with

[9]Adding more than three call options on an asset does not improve the ability to tailor the sensitivity of the option package's price to volatility, because the price-volatility profiles of the additional options will be linear combinations (to the accuracy of a computer) of the price-volatility profiles of the first three options. See footnote 13 for a discussion of how the strike prices were chosen.

[10]Of course, the package price will also depend on other variables including the risk-free interest rate, the dividend rates on the two assets, the time to maturity, current asset prices, and the strike prices of the individual options. These variables will not differ across forecasting agents.

strike price equal to the current price as C_1, the call on asset 1 with strike price at 101% of the current price as C_2, and so on. Using this notation, the function V can be written as

$$V(\sigma_1,\sigma_2,\rho;x) = R(\sigma_1,\sigma_2,\rho) + x_1 C_1(\sigma_1) + x_2 C_2(\sigma_1) + x_3 C_3(\sigma_1) + x_4 C_4(\sigma_2) + x_5 C_5(\sigma_2) + x_6 C_6(\sigma_2) \qquad \textbf{(2)}$$

where $R(\sigma_1, \sigma_2, \rho)$ is the price of the rainbow option and $C_j(\sigma)$ is the price of the j^{th} call option.

Our objective to find a package of options that is insensitive to volatility but sensitive to correlation can be expressed as choosing x so that

$$V(\sigma_1^{(i)},\sigma_2^{(j)},\rho^{(k)};x) = V_k \quad \forall i,j,k \in [1,...,n] \qquad \textbf{(3)}$$

where a superscript identifies which agent's forecast of that variable is being used. In words, if forecasting agent k prices the package using her own correlation forecast ($\rho^{(k)}$) and any of the n forecasts of σ_1 and σ_2, she will come up with a constant price V_k. This condition should hold for all n forecasting agents, which implies a system of n^3 equations. Using (3) and the definition of V in (2), we can write the system of equations in matrix form as

$$Ay = b \qquad \textbf{(4)}$$

where

$$A=\begin{bmatrix} C_1(\sigma_1^{(1)}) & \cdots & C_6(\sigma_2^{(1)}) & -1 & 0 & \cdots & 0 \\ C_1(\sigma_1^{(2)}) & \cdots & C_6(\sigma_2^{(1)}) & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_1(\sigma_1^{(n)}) & \cdots & C_6(\sigma_2^{(n)}) & -1 & 0 & \cdots & 0 \\ C_1(\sigma_1^{(1)}) & \cdots & C_6(\sigma_2^{(1)}) & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_1(\sigma_1^{(n)}) & \cdots & C_6(\sigma_2^{(n)}) & 0 & 0 & \cdots & -1 \end{bmatrix}, \quad y=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_6 \\ V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix}, \quad b=\begin{bmatrix} -R(\sigma_1^{(1)},\sigma_2^{(1)},\rho^{(1)}) \\ -R(\sigma_1^{(2)},\sigma_2^{(1)},\rho^{(1)}) \\ \vdots \\ -R(\sigma_1^{(n)},\sigma_2^{(n)},\rho^{(1)}) \\ -R(\sigma_1^{(1)},\sigma_2^{(1)},\rho^{(2)}) \\ \vdots \\ -R(\sigma_1^{(n)},\sigma_2^{(n)},\rho^{(n)}) \end{bmatrix}$$

No exact solution to (4) will exist. The economic logic is as follows. We want to remove the sensitivity of the option package's price to volatility by adding the appropriate amounts of six call options. However, each forecasting agent perceives a different price-volatility profile for the rainbow option, because the price-volatility profile depends on the level of correlation (in other words, $\partial^2 V/\partial\rho\partial\sigma$ is not zero for the rainbow options we consider). We could choose a package of call options whose value would be perceived by one forecasting agent to be insensitive to volatility. However, the other agents would still perceive some sensitivity to volatility, and the results would still be sensitive to volatility.

Since no exact solution will exist, we look for the least-squares solution to (4). The least squares solution will choose call options whose price-volatility profile comes closest to offsetting the individual price-volatility profiles of the *n* forecasting agents. In algebraic terms, the least squares solution solves

$$\min_{x}\left[\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}[V(\sigma_1^{(i)},\sigma_2^{(j)},\rho^{(k)};x) - V_k]^2\right] \qquad \textbf{(6)}$$

For large *n* and sufficiently heterogeneous forecasts, the matrix *A* will be of full rank and the least squares solution to (4) will be

$$y = (A'A)^{-1}A'b$$

In other cases, the least squares solution will not be uniquely determined (the matrix A will not be of full rank). In this situation, we choose the solution with the shortest length, calculated as $y = A^{+}b$, where A^{+} denotes the pseudoinverse (Moore-Penrose inverse).[11]

In the context of our particular application, the criterion of "shortest length" is a useful way for us to choose one of the many least squares solutions. Relative to the "shortest length" solution, other solutions will include larger long and short positions in call options on the two assets. These other solutions involve more buying and selling of call options to get to the same place (in terms of smoothing out the price-volatility profile), magnifying the small option pricing errors made by pricing the options under Black-Scholes assumptions. These other solutions would also make our methodology less numerically stable.

We will focus in our empirical example on the case where we care only about the accuracy of the correlation forecast. The preceding discussion has made it clear that we can never remove all sensitivity of the option package's price to volatility. Some residual sensitivity to volatility will remain, although we expect it will be small relative to the sensitivity to correlation. Rather than simply hope we have done a good job of removing the sensitivity of the option package's price to volatility, we can check this by examining the relationship between the agents' correlation forecasts and their option package prices. The relationship will be monotonic if we have done a good job of

[11]See Strang (1980, section 3.4) for a discussion of the properties of the pseudoinverse.

removing the sensitivity to volatility.[12] We can check this by comparing the Spearman rank correlation between the correlation forecast and the option package price on each day with +1 or -1.[13]

3 Using simulations to evaluate the methodology

We have described a methodology for evaluating forecasts of correlation that is based on the intuitive notion that a correct forecast will produce a more accurate rainbow option price than an incorrect forecast. Although it is based on a simple idea, our methodology is too complex to admit a formal proof of consistency. In place of a proof, this section of the paper demonstrates how our methodology performs on simulated data with simple static forecasts. We also use simulations to examine how the methodology performs with four different rainbow options and choose one to use in the empirical application that follows.

3.1 The setup of the simulations

The simulations will use a data generating process for the two daily asset returns consisting of a bivariate normal distribution with means equal to zero, standard deviations of 14.1 percent (annualized) for both asset returns, and a correlation of 0.6. Unlike our empirical application, the simulations do not distinguish between calendar time and trading time; no weekends or holidays are present in the simulations.[14] Each simulation will be repeated for 100 trials and will include five forecasting agents, each of whom will make a static (constant) forecast throughout each trial. One of

---

[12]Whether it will be monotonically increasing or decreasing depends on whether the price of the particular rainbow option used is increasing or decreasing in correlation.

[13]We chose the strike prices of the call options in the option package by doing a grid search over possible combinations of out-of-the-money, at-the-money, and in-the-money strike prices. The strike prices of 100%, 101%, and 102% of the spot price did the best job of isolating the effect of correlation on the option package price, as measured by the fraction of days for which the rank correlation between the option package prices and the correlation forecasts was ±1 (using the real data from Section 5, not the simulated data from Section 3).

[14]To compensate for the absence of weekends and holidays in the simulations, daily variances are multiplied by 250 (not 365) when annualized.

the five agents will make a correct forecast, one will make a correct correlation forecast but an incorrect volatility forecast, one will make a correct volatility forecast (for both assets) but an incorrect correlation forecast, and two will make incorrect forecasts of both volatility and correlation. We will vary several elements throughout the simulations: the sample size, the particular rainbow option used in the option trading methodology, and whether the incorrect forecasts are "close" or "far" from the truth.[15] In this way we will get an idea (although not a proof) of how our methodology performs along these various dimensions.

When an agent makes an incorrect forecast, it is chosen at random from a predetermined set of incorrect forecasts. We use two sets of incorrect forecast, one "close" to the truth and one "far" from the truth (see Table 1). The incorrect volatility forecasts are not assumed to be independent of one another, since in reality a biased volatility forecasting methodology is likely to be biased in the same direction for both volatilities. To create this sort of dependence in our simulations, we force the bias of the two incorrect volatility forecasts to have the same sign by choosing the volatility of asset 1 at random, then choosing the volatility of asset 2 at random from the subset of incorrect volatility forecasts that have the same direction of bias. Correlation forecast errors are assumed to be independent of volatility forecast errors.

One task of our simulations will be to compare how our methodology performs with various rainbow options. We will examine four rainbow options: a call spread, an outperformance option, a call worst, and a basket. Table 2 shows the payoff function for each. In addition to choosing a particular rainbow option, we must choose a strike price. We choose all strike prices to be "at the money." Strike prices that are not at the money make the price-volatility profile more non-monotonic, which makes it harder for our "option package" technique to remove the sensitivity to volatility.

[15]To remove an undesired source of variation across simulations, we do not vary the sequence of pseudo-random numbers used to generate the simulated return data.

3.2 Comparing the basic methodology and the option package methodology

Table 3 shows the results of a simulation of the basic methodology, with agents trading a single call spread option. Table 4 shows results of a simulation of the option package methodology, with a single call spread option and six call options, designed to remove the sensitivity to volatility but preserve the sensitivity to correlation. Comparing these two will show the effect of tailoring the loss function to focus on the accuracy of correlation forecasts. Each table (and each of the remaining tables in this section) has two panels, corresponding to the cases where the incorrect forecasts are "close" or "far" from the truth. Each panel has a table with five columns, corresponding to the five forecasting agents. Each panel's table has three rows, each corresponding to a different sample size (250 days, 1000 days, and 5000 days). Each cell of the table has three pieces of information: the number of trials (out of 100) in which that forecasting agent had the highest profits, the mean across the 100 trials of mean daily profit, and the mean across the 100 trials of standard deviation of daily profit.

Table 3 shows that the basic methodology can identify good forecasts and that its ability to do so is enhanced by a longer sample and when forecasts are more heterogeneous. In Panel A, the agent with completely correct forecasts has the highest profits per day in 40 trials with N=250, 58 trials when N=1000, and 78 trials when N=5000. In Panel B, with more heterogeneous forecasts, the methodology does better, with the correct forecasts winning in 59, 80, and 85 trials. Notice that the basic methodology, with a single call spread option, rewards the ability to forecast both volatility and correlation accurately. In particular, agent 1 does much better than agent 2, even though both have identical, correct forecasts of correlation.

Table 4 shows that the option package methodology, trading a call spread option plus six call options, does a good job of reducing the sensitivity of the trading profits to volatility, enhancing the ability to identify good correlation forecasts. If the package methodology were able to remove all sensitivity to volatility, agents 1 and 2 should have exactly equal performance, since they have identical correlation forecasts. But, as we discussed above, it is impossible to remove all sensitivity to volatility. As a result, agent 1 still outperforms agent 2, but by a small margin. In Panel B, agent 1 has the highest profits in 52 trials with N=250, compared with 33 for agent 2. At N=5000, the package methodology correctly identifies one of the two agents with the correct correlation forecast in 98 of 100 trials. This compares with 87 out of 100 in Table 3 when trading the single call spread option.

3.3 Comparing four rainbow options

The simulations reported in Table 3 and Table 4 used a call spread option as the rainbow option that was priced and traded by the forecasting agents. Different rainbow options have different relative sensitivities to correlation and volatility, creating some variation in how well the option package methodology can remove the sensitivity to volatility. To choose a particular rainbow option to use in our empirical application, we repeat our simulations of the option package methodology with the three remaining rainbow options from Table 2 and compare the results.

Table 5 presents results using an outperformance option. Table 6 presents results using a call worst option. Table 7 presents results using a basket option (with weights equal to one-half). We compare the results in these three tables with those in Table 4 for the call spread option. The statistic we choose to compare across tables is the number of times one of the two agents with a correct correlation forecast had the highest profits, which will be the sum of the numbers on the first line of the leftmost two columns. We look at the results for a sample size of 5,000 because the dataset we use in our empirical application is of that size. For the Panel A results, where incorrect forecasts are close to the truth, our comparison statistic is 86, 57, 79, 76 for the four rainbow options. For Panel B results, it is 98, 96, 98, 100. When the incorrect forecasts are far from the truth, any of the four rainbow options can distinguish a true correlation forecast from a false one. When the incorrect forecasts are close to the truth, the call spread option does a better job.

3.4 Summing up the simulations

We take the following conclusions away from the simulations. First, our methodology can distinguish accurate correlation forecasts from inaccurate ones. Second, among the four rainbow options we consider, the call spread option allows our option package methodology to remove the most sensitivity to volatility. Third, both the sample size and the degree of heterogeneity among forecasts affect how the methodology performs. With twenty years of data (N=5,000) and very heterogeneous forecasts, the methodology appears to work well. With one year of data (N=250) and not-too-heterogeneous forecasts, the methodology cannot do much.

4 Five Forecasting Models

4.1 Model-based forecasts and option-implied forecasts of volatility and correlation

Two approaches to forecasting volatility exist. One involves the statistical analysis of historical returns to develop out-of-sample forecasting models. The other is based on extracting volatilities implied by some option pricing formula. Comparisons of volatility forecasting techniques are numerous; see Figlewski (1997) for a survey. In general, these studies find that both implied volatility from options prices and volatility forecasts based on historical returns contain (non-overlapping) information about future volatility.

These same two approaches can be used to obtain correlation forecasts. We choose to focus on correlation forecasts based on the analysis of historical returns, not on implied correlation extracted from rainbow option prices. By not using implied correlation forecasts, we do not limit the empirical application of our methodology to the small number of financial time series, primarily exchange rates, for which price data on rainbow options is available. Using our methodology to evaluate implied correlation forecasts would be an interesting topic for future research.

4.2 Summary of the five models

We investigate the forecasting ability of five models, three of which allow both conditional volatilities and conditional covariances to vary independently over time: a diagonal bivariate GARCH(1,1), a bivariate two-state Markov regime-switching model, and an exponentially weighted moving average model. These three allow for time-varying conditional correlations.[16] To compare the performance of a model of correlation stability, we investigate a bivariate GARCH(1,1) with constant conditional correlations. Finally, we also look at the forecasting ability of the naive historical correlation. We allow all five models to use N days of historical data to make a forecast.[17] To make forecasts for day 1, pre-sample data from days -(N-1) to 0 will be used.

All five forecasting models can be considered a special case of the following general model for a vector of continuously compounded returns r_t:

$$r_t = E(r_t | \psi_{t-1}) + \epsilon_t$$
$$\mathrm{Var}(\epsilon_t | \psi_{t-1}) = \Sigma_t \equiv \begin{bmatrix} \sigma^2_{11,t} & \sigma^2_{12,t} \\ \sigma^2_{12,t} & \sigma^2_{22,t} \end{bmatrix} \qquad \mathbf{(8)}$$

where ε_t is a serially uncorrelated random vector with zero means, and ψ_{t-1} represents past returns through time t - 1. The algorithms differ in the distribution they assume for ε_t.

4.3 Naive Historical Forecasts (NAIVE)

Naive historical forecasts implicitly assume conditional and unconditional moments are equal, i.e.,

---

[16]Bera and Kim (1996) and Longin and Solnik (1995) find time-varying correlation to be important empirically.

[17]In the empirical tests below, we set N = 1,000.

$$E(r_t \mid \psi_{t-1}) = E(r_t)$$
$$\mathrm{Var}(\epsilon_t \mid \psi_{t-1}) = \mathrm{Var}(\epsilon_t). \qquad \textbf{(9)}$$

The naive forecast of $E(r_t)$ is the sample mean. Variance and covariance forecasts are obtained by calculating moving averages over an N-day window,

$$\sigma^2_{ii,t+1} = \frac{1}{N}\sum_{j=0}^{N} \epsilon^2_{i,t-j} \qquad \text{for } i = 1,2$$
$$\sigma^2_{12,t+1} = \frac{1}{N}\sum_{j=0}^{N} \epsilon_{1,t-j}\epsilon_{2,t-j} \ . \qquad \textbf{(10)}$$

Note that the bivariate normal distribution is a special case for which (9) holds. If the error terms are truly bivariate normal, the forecasts in (10) will be unbiased, efficient estimates of the unconditional variance and covariance. Otherwise, they need not be efficient.

4.4 Exponentially Weighted Moving Average Forecasts (EWMA)

One approach to model time varying variances and covariances is to express each element of the variance-covariance matrix as an exponentially weighted moving average of past squared errors or cross product of errors. Exponentially weighted moving average forecasts are given by

$$\sigma^2_{ii,t+1} = \lambda\,\sigma^2_{ii,t} + (1-\lambda)\epsilon^2_{i,t} \approx (1-\lambda)\sum_{j=0}^{N} \lambda^j \epsilon^2_{i,t-j} \qquad \text{for } i=1,2$$
$$\sigma^2_{12,t+1} = \lambda\,\sigma^2_{12,t} + (1-\lambda)\epsilon_{1,t}\epsilon_{2,t} \approx (1-\lambda)\sum_{j=0}^{N} \lambda^j \epsilon_{1,t-j}\epsilon_{2,t-j}. \qquad \textbf{(11)}$$

where the sums are truncated after N days. The most recent errors carry the highest weight in the weighted average, thus allowing forecasts to quickly reflect market shocks. Note that both the NAIVE and EWMA forecasts average past squared residuals and cross products of residuals but with different weights. Equation (11) gives no indication of how to identify proper values for λ. Following the

RiskMetrics™ Technical Document (1996), we examine EWMA forecasts made with $\lambda = 0.94$ and assume the unconditional mean is zero.

4.5 GARCH Forecasts

The ARCH model of time-varying variance was originally proposed by Engle (1982) and later generalized under the acronym GARCH by Bollerslev (1986). Due to the well-known dynamic nature of conditional volatilities, the univariate ARCH family of forecasting models has won widespread use in forecasting volatility (e.g., Figlewski (1997), Noh et al. (1994), Engle et al. (1993), Baillie and Bollerslev (1992), Jorion (1995), and West and Cho (1994)).

A general bivariate GARCH model, which encompasses the two models investigated in this paper, is the vec representation,

$$(\epsilon_t | \psi_{t-1}) \sim N(0, \Sigma_t)$$

$$\begin{bmatrix} \sigma^2_{11,t} \\ \sigma^2_{12,t} \\ \sigma^2_{22,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_{12} \\ c_2 \end{bmatrix} + A \begin{bmatrix} \epsilon^2_{1,t-1} \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon^2_{2,t-1} \end{bmatrix} + B \begin{bmatrix} \sigma^2_{11,t-1} \\ \sigma^2_{12,t-1} \\ \sigma^2_{22,t-1} \end{bmatrix} \qquad \textbf{(12)}$$

where A and B are 3×3 matrices of parameters. Engle and Kroner (1995) develop a parameterized version of the vec GARCH model that restricts Σ_t to be positive definite while maintaining the generality of the vec model. As a result, estimation of the most interesting special cases of (12) is made practical. We choose to work with two special cases of (12).

In developing the GARCH models used in this paper, we follow Noh et al. (1994) on two accounts. First, we assume an AR(1) model for continuously compounded returns,

$$r_{i,t} = a_{0,i} + a_{1,i} r_{i,t-1} + \epsilon_{i,t} \qquad \text{for } i=1,2. \tag{13}$$

Second, since traders can forecast weekends and holidays with perfect accuracy, we extend the GARCH models to distinguish between calendar time and trading time. We allow non-trading days to induce only a fraction of the volatility of trading days, as documented in French and Roll (1986).[18] Extending the approach of Noh et al (1994) to the bivariate case, we introduce a parameter δ_1 to measure the average intensity of variance over non-trading days and a parameter δ_2 (in the D-GARCH model) to measure the average intensity of covariance over non-trading days. As the δ parameters increase from zero to one, the intensity of variance and covariance over non-trading days increases from none to a level equal to that over trading days.

4.5.1 Diagonal GARCH (D-GARCH)

A natural restriction of the vec GARCH model is to allow each element of Σ_t to depend on only past values of itself and past values of $\varepsilon_{i,t}\,\varepsilon_{j,t}$ (the (i,j) th element of $\varepsilon_t\,\varepsilon_t'$),

$$\begin{aligned} \sigma^2_{ii,t} &= \omega_i + \alpha_i \epsilon^2_{i,t-1} + \beta_i \sigma^2_{ii,t-1} \qquad \text{for } i=1,2 \\ \sigma^2_{12,t} &= \omega_{12} + \alpha_{12} \epsilon_{1,t-1} \epsilon_{2,t-1} + \beta_{12} \sigma^2_{12,t-1} \end{aligned} \tag{14}$$

This restriction is referred to as the *diagonal representation* (D-GARCH), originally proposed by Bollerslev, Engle, and Wooldridge (1988). The D-GARCH model allows both conditional variances and conditional covariances to vary independently, thus allowing conditional correlations to vary over time.

---

[18]The subscript t continues to index trading time, and we separately keep track of the number of calendar days between trading days t-1 and t.

4.5.2 Constant conditional correlation GARCH (C-GARCH)

Another parameterization of the vec model, suggested by Bollerslev (1990), restricts the conditional correlation to be temporally constant (C-GARCH). The diagonal elements of the variance-covariance matrix are modeled as univariate GARCH processes, and the off-diagonal elements are given by multiplying the constant conditional correlation coefficient, ρ, by the appropriate root cross product of diagonal elements,

$$\sigma^2_{ii,t} = \omega_i + \alpha_{i,1}\epsilon^2_{i,t-1} + \beta_{i,1}\sigma^2_{ii,t-1} \quad \text{for } i=1,2$$
$$\sigma^2_{12,t} = \rho\sqrt{\sigma^2_{11,t}\sigma^2_{22,t}}. \tag{15}$$

4.5.3 GARCH models with the weekend effect

After incorporating the "weekend effect," the specification of the D-GARCH model used in this paper is

$$\frac{\sigma^2_{ii,t}}{n_t^{\delta_1}} = \omega_i + \alpha_i\frac{\epsilon^2_{i,t-1}}{n_{t-1}^{\delta_1}} + \beta_i\frac{\sigma^2_{ii,t-1}}{n_{t-1}^{\delta_1}} \quad \text{for } i=1,2$$
$$\frac{\sigma^2_{12,t}}{n_t^{\delta_2}} = \omega_{12} + \alpha_{12}\frac{\epsilon_{1,t-1}\epsilon_{2,t-1}}{n_{t-1}^{\delta_2}} + \beta_{12}\frac{\sigma^2_{12,t-1}}{n_{t-1}^{\delta_2}} \tag{16}$$

and the C-GARCH model takes the form:

$$\frac{\sigma^2_{ii,t}}{n_t^\delta} = \omega_i + \alpha_i \frac{\epsilon^2_{i,t-1}}{n_{t-1}^\delta} + \beta_i \frac{\sigma^2_{ii,t-1}}{n_{t-1}^\delta} \quad \text{for } i=1,2$$

$$\sigma^2_{12,t} = \rho\sqrt{(\sigma^2_{11,t}\sigma^2_{22,t})}. \tag{17}$$

where n_t is a positive integer equal to the number of calendar days between observations t-1 and t.[19] In both instances, the estimation assumes the conditional distribution of $\varepsilon_{i,t}$ is normal with zero mean.

4.6 Markov Forecasts (MARKOV)

Finally, we investigate the forecasting ability of a bivariate two-state Markov switching model.[20] Sola and Timmermann (1994) show that a univariate Markov switching model, like univariate ARCH and GARCH models, can successfully mimic important features of financial time series such as time-varying volatility, persistence in volatility, and excess unconditional kurtosis ("fat tails"). The Markov model assumes returns follow an AR(1) as in (13), with the error term distributed as

$$\epsilon_t | s_t = j \sim N(0,\Sigma_j) \quad \text{for } j=1,2 \tag{18}$$

The unobserved variable s_t defines the regime the model is in at time t and is assumed to follow a two-state Markov process:

$$P(s_t = j \,|\, \psi_{t-1}) = P(s_t = j \,|\, s_{t-1} = i) = p_{ij} \tag{19}$$

We also extend the Markov model to incorporate the weekend effect. Each day's variances and covariances in a particular regime are assumed to be the product of two terms: a term reflecting the variances and covariances associated with the regime and a term reflecting the number of calendar

[19]In our notation, $\sigma^2_{11,t}$ represents the variance of asset 1 between trading days t-1 and t.

[20]See Hamilton (1994) for a description of a univariate two-state Markov switching model.

days since the previous trading day. Specifically, the variance-covariance matrix of returns from t-1 to t in regime j is modeled as

$$\Sigma_{j,t} = \begin{bmatrix} \sigma^2_{11,j,t} & \sigma^2_{12,j,t} \\ \sigma^2_{12,j,t} & \sigma^2_{22,j,t} \end{bmatrix} = \begin{bmatrix} n_t^{\delta_1} h^2_{11,j} & n_t^{\delta_2} h^2_{12,j} \\ n_t^{\delta_2} h^2_{12,j} & n_t^{\delta_1} h^2_{22,j} \end{bmatrix}. \tag{20}$$

This parameterization of the weekend effect obtains the desired result of forecasting higher volatilities and covariances on days when $n_t > 1$. We will avoid the mistake of concluding the data generating process is in the high variance regime on days when volatilities and covariances are high only because of the weekend effect. Forecasts based on the Markov model are expressed as

$$\Sigma_{t+1} = P(s_{t+1} = 1 \mid \psi_t)\Sigma_{1,t+1} + P(s_{t+1} = 2 \mid \psi_t)\Sigma_{2,t+1}. \tag{21}$$

5 Empirical application

5.1 Data and estimation strategy

To apply our methodology, we require data on two correlated financial time series. In order to accurately measure the correlation, we must have data that are recorded at the same time each day.[21] For this reason we choose to apply our methodology to daily, continuously compounded returns for the S&P 500 and the NASDAQ composite index. The data set begins on December 14, 1972, and ends on May 14, 1997. Data though the end of 1993 were obtained from the CRSP tape. All other data was obtained from the US FAME database maintained at the Board of Governors of the Federal

[21]See RiskMetrics™ Technical Document (1996), p. 184, for a discussion of the problem of nonsynchronous data collection and correlation estimation.

Reserve System.[22] The set of observations for which there was a missing value for either series was removed before log differencing, and the set of observations which occurred around the time of the October 1987 crash (October 10 - October 30), was removed after log differencing.[23] Table 8 shows summary statistics of the data.

Model estimation is required to generate the D-GARCH, C-GARCH, and MARKOV forecasts. Influenced by the results of Engle et al. (1993), who compared the performance of volatility forecasts using GARCH models estimated over various sample sizes, we choose to estimate the parameters of each model using the most recent 1,000 observations.[24] The first 1,000 days in the data set are used to estimate the models and make volatility and correlation forecasts for December 7, 1976, the day we begin to apply our option trading methodology. Each model is estimated using maximum likelihood for each day in the sample.[25]

5.2 Summary of the five forecasting models and their forecasts

The means and standard deviations of the five forecasts of daily volatility and correlation over the 5,154 days in our sample are shown in Table 9. For those forecasts that take account of weekends and holidays, Table 9 incorporates those effects. In general, the NAIVE forecast has the least variability over time, EWMA has the most, and the three model-based forecasts have an intermediate amount of variability.

Table 10 shows the means of the estimated parameters for the D-GARCH, C-GARCH, and MARKOV models over the 5,154 days in our sample. Little in Table 10 will surprise readers who are

---

[22]A data series for the risk-free interest rate (3-month T-bill) was also taken from the Fed's database. Dividends were set to zero.

[23]Again, we follow Noh et al (1994) in excluding days around the October 1987 stock market crash.

[24]The two moving average forecasts (NAIVE and EWMA) were also computed using the most recent 1,000 observations.

[25]All estimation was done using GAUSS' MAXLIK module. We estimated 5,154 C-GARCH models, 5,154 D-GARCH models, and 5,154 MARKOV models. Using the previous day's parameter estimates as starting values for the maximum likelihood procedure expedited the estimation.

familiar with the literature on GARCH and Markov-switching models. One point of interest is the magnitude of the weekend effect. The delta parameters are all around 0.2, implying that variances and covariance from Friday's close to Monday's close are around 25 percent higher than over a weekday. Noh et al (1994) found a similar magnitude of the weekend effect for variances; our results show the same magnitude holds on average for a covariance.

5.3 Evaluating only correlation forecasts

We first evaluate the five forecasting models using our option package methodology, tailored to focus on correlation forecasts by minimizing the sensitivity to volatility. Table 11 shows the mean daily profits of the five forecasting models. MARKOV shows the highest profits per day, followed closely by D-GARCH. These two models model the time variation in correlation, and they also allow for the "weekend effect" to have different effects on variances and on covariances. C-GARCH does worse than D-GARCH, perhaps not unexpectedly, since it imposes an assumption of constant conditional correlation. C-GARCH also forces the "weekend effect" to be the same for variances and covariances, since conditional covariances are not forecast separately from conditional variances. The two "simple" forecasts, NAIVE and EWMA, have the lowest profits.

We assess statistical significance of these differences by computing asymptotic heteroscedasticity and autocorrelation-robust standard errors for the mean daily profits, shown in the second column of the table.[26] The third column of the table shows *t*-statistics for the null hypothesis that the mean of daily profits equals zero. The null hypothesis can be rejected for MARKOV, whose mean daily profit is statistically significantly greater than zero, and for NAIVE, whose mean daily profit is statistically significantly less than zero. The fourth column of the Table shows *t*-statistics for the null hypothesis that the mean of each agent's daily profits equals the mean of MARKOV's daily

[26]The only assumption necessary to permit us to use a standard robust covariance estimator is covariance stationarity of the underlying stock index returns. Because the daily profits are a bounded transformation of a moving 1000-day window of the stock index returns, they retain the covariance stationarity of the underlying data.

profits. The null can be rejected for NAIVE and EWMA, but not for the two GARCH forecasting agents. These results suggest that the three more sophisticated forecasting models (MARKOV, C-GARCH, D-GARCH) outperform simpler forecasting models (NAIVE and EWMA) but do not give us a basis to conclude that one of the three sophisticated forecasting models is superior to the other two.[27]

Figure 1 displays a graph of the daily profit series for the five forecasting agents. The figure shows that the difference between MARKOV and D-GARCH is not very large, compared to the gaps between the other three forecasting models. Different subsamples of the data could show a different ranking between MARKOV and D-GARCH. The figure confirms the extreme variability of the daily profits, and the importance of data outliers that lead to large spikes in the profit series. The figure also shows that most of the profits were made or lost in the first half of the sample period.

To investigate further the stability of the rankings over subsamples of our 20 years of data, Table 12 summarizes the relative position of the five forecasting agents when results are tabulated by calendar year. MARKOV still comes out on top, finishing first in 8 years and second in 5 years. A surprising finding in this table is the performance of NAIVE, finishing first in 6 years (although last in 6 years also).

5.4 Excluding days where we cannot isolate the effect of correlation

At the end of section 2, we discussed how our option package methodology could never completely remove the influence of different volatility forecasts on the results. If the package methodology is working, on each day the relationship between each agent's correlation forecast and her

---

[27]Our results using the call spread option should not be strongly affected by the fact that stock markets were on average rising during our sample period, because the results depend on the spread between NASDAQ and S&P 500 returns but not on their level. Although the spread was on average positive, it was much smaller than the level of returns.

It is possible that the high stock market returns during our sample period combined with our use of standard call options in the "option package" methodology could affect the results. To check whether this was important, we replaced the call options in the hedging portfolio with puts and found that the mean daily profits reported in Table 11 were virtually unchanged.

option package price should be monotonic. In this subsection we remove from the sample those days where the option package methodology cannot remove enough of the sensitivity to volatility. Specifically, on each day we add an extra step to the methodology following the determination of the contents of the option package: the Spearman rank correlation between the agents' correlation forecasts and their option package prices is calculated. If the rank correlation is greater than -.5, no trading for that day takes place and we skip forward to the following day. If the rank correlation is less than or equal to -.5, the remaining steps for that day are carried out.[28]

Excluding certain days creates the potential for selection bias. We do not believe that the selection mechanism favors one agent's forecasts over another's. The selection mechanism is triggered when volatility and correlation forecasts are too dissimilar for the option package methodology to remove the sensitivity of the call spread option's price to volatility. Selection bias would be a problem only if one forecasting model consistently produces better forecasts on days with heterogeneous variance-covariance forecasts compared to days with homogeneous variance-covariance forecasts.

Table 13 shows the mean daily profits under this alternative, along with robust standard errors and *t*-statistics. The ranking of forecasts is unchanged, and the range of profits is a little smaller than it was in Table 11. Figure 2 shows a graph of the five daily profit series. It too looks little changed from the previous results. Evidently the unwanted sensitivity of our methodology to volatility forecasts did not have a large effect on the ranking of the five forecasting models.

5.5 Forecasting volatility and correlation: four economic loss functions

Our methodology can also be used to evaluate forecasts of the entire variance-covariance matrix, not just correlation. Of course, the relative importance of volatility and correlation in

---

[28]The price of a call spread option is negatively related to correlation. A rank correlation of -1 means the relationship between correlation forecast and option package price is monotonic.

determining which forecast is "best" will depend on which rainbow option is traded by the forecasting agents. In this subsection, we present results for the four rainbow options defined in Table 2.

Table 14 shows the mean daily profit of the five forecasting agents when they trade and price the four different rainbow options alone, along with robust standard errors. The rankings of the five forecasting models are consistent across all four rainbow options: D-GARCH, C-GARCH, EWMA, MARKOV, NAIVE. There is one, important difference between these results and the results using the "option package" presented above. While the MARKOV model had the best performance when only correlation mattered, it does much more poorly when both volatility and correlation forecasts matter. There are two reasons behind this difference. First, the MARKOV model appears to do a much better job forecasting correlation than it does forecasting volatility. Second, while the rainbow options are sensitive to both volatility and correlation, the differences among agents' volatility forecasts are greater than the differences among their correlation forecasts, so the ability to forecast volatility well provides an opportunity for greater trading profits.

To confirm this point, we replicated the test of Engle et al (1993), who evaluated univariate volatility forecasting models by their ability to price and trade options. In place of a rainbow option, the forecasting agents are required to price and trade the combination of an at-the-money put option and an at-the-money call option (termed a straddle). The rankings of mean daily profit for trading a straddle are D-GARCH, EWMA, C-GARCH, MARKOV, NAIVE on the S&P data and C-GARCH, D-GARCH, EWMA, MARKOV, NAIVE on the NASDAQ data. These results confirm that MARKOV does a poor job forecasting volatility according to our option trading methodology.

5.6 Results for "statistical" loss functions

We have motivated our methodology by a desire to evaluate forecasts of joint returns based on their ability to forecast a tailored combination of correlation and volatility. An alternative way to evaluate volatility and correlation forecasts would be to use a statistical loss function, such as mean squared error, where the "error" is measured as the deviation from the forecast variance (or covariance) of the realized squared return (or cross-product of returns). Because the squared return (and the cross-product of returns) is a consistent estimator of the unobservable variance (or covariance), this method should be able to distinguish accurate from inaccurate variance-covariance matrix forecasts.

A loss function using mean squared forecasting error as the criterion for defining a good correlation forecast would be

$$L = \frac{1}{T}\sum_{t=1}^{T}(\rho_t - \rho_t^f)^2 \quad \textbf{(22)}$$

where T is the number of days in our data sample, ρ_t^f is the correlation forecast and ρ_t=sign($\varepsilon_{1t}\varepsilon_{2t}$)×1 reflects whether the correlation on day *t* was positive or negative. The top panel of Table 15 shows the mean squared error of our five forecasting models computed using (22). The ranking of forecasting models is NAIVE, EWMA, D-GARCH, MARKOV, C-GARCH. The differences between NAIVE and the three sophisticated forecasting models are statistically significant. The ranking is quite different from what we found in Table 11 using an option-pricing-based loss function.

A loss function using mean squared forecasting error as the criterion for defining a good variance-covariance matrix forecast would be

$$L = \frac{1}{T}\sum_{t=1}^{T}\left[\iota' \mathrm{vec}(H_t - H_t^f)^2\right] \quad \textbf{(23)}$$

where $H_t = \varepsilon_t\varepsilon_t'$, H_t^f is the forecast variance-covariance matrix, and ι is a 4x1 vector of ones. (If more generality were desired, the vector of ones could be replaced by an arbitrary weighting vector.) The bottom panel of Table 15 shows the mean squared error of our five forecasting models computed using (23). The ranking of forecasting models is D-GARCH, NAIVE, C-GARCH, MARKOV, EWMA. Compared with the results in Table 14, the ordering of forecasting models is again quite different, particularly the strong performance of NAIVE. The choice of loss function again appears to make a large difference to the results.[29]

6 Conclusions

We have described a methodology to evaluate volatility and correlation forecasts based on the forecasts' ability to accurately price options. The methodology can be tailored to emphasize accurate correlation forecasts or accurate volatility forecasts, or a weighted combination of the two. The methodology should be of interest to those who price and hedge options whose payoff depends on more than one financial time series. We chose to focus, in our empirical application, on the accuracy of correlation forecasts for equity returns. From the point of view of a seller of rainbow options, if volatility risk can be hedged but correlation risk cannot, an accurate correlation forecast is much more valuable than an accurate volatility forecast.

One advantage of our methodology over other ways to evaluate the accuracy of correlation forecasts is its use of an economic loss function, based on option pricing errors, rather than a statistical loss function. Engle et al (1993) found that different volatility forecast models were preferred under different statistical loss functions (for example, mean square error versus variance of forecast error). We find a similar result for the correlation forecasts we evaluate. One issue for future research is whether statistical loss functions are a poor choice for evaluating forecasts of variables that are unobservable (such as volatility and correlation).

In our empirical application, we compared five forecasting models for the variance-covariance matrix of two equity returns. The models ranged from simple averages of past squared returns to quite sophisticated GARCH and Markov switching models that take into account the difference in volatility

[29]Engle et al (1993) also found that mean squared forecasting error did not provide a good guide to which forecasting model did a better job of pricing options.

over trading and non-trading days. We found that the more sophisticated and more general models, the Markov switching model and the diagonal GARCH model, produced statistically significantly better forecasts of correlation. This result suggests that market participants should consider switching from forecasting correlation using weighted averages of past returns to using a GARCH model or a Markov switching model. When evaluating forecasting models on their ability to jointly forecast volatility and correlation, the Markov switching model performed poorly, while the GARCH models had the best performance.

Of course, these results depend on the particular forecasting models we chose. Other models to forecast correlation and volatility may exist that are superior to the ones we chose to implement. In particular, given the strong performance of the GARCH models in our empirical applications, it would be useful to evaluate more sophisticated GARCH models, such as a multivariate version of the exponential ARCH model of Nelson (1991) or a fully general bivariate GARCH model as formulated by Engle and Kroner (1995), to see if yet another level of additional complexity pays off in better correlation forecasts. For the small but growing number of financial time series for which "rainbow" options are actively traded, it would be interesting to evaluate forecasts based on implied correlation using our methodology. Also, it would be interesting to see if our results for equity markets carry over to other markets (fixed income, currencies, commodities).[30]

Several extensions of our methodology are possible avenues for future research. It would be interesting to see if the same results obtain when evaluating forecasts longer than one day. (The implied correlation literature typically evaluates forecasts of one-month or three-month duration.) For risk measurement, the entire variance-covariance matrix is important, but the relative importance of the various elements will depend on the particular portfolio whose risk is being measured. For risk

[30]Figlewski (1997) found that GARCH models did a better job forecasting volatility in equity markets than in other markets.

measurement applications, it would be interesting to use our methodology to evaluate variance-covariance matrix forecasts on their ability to price a portfolio of options rather than a single rainbow option.

References

Andrews, Donald W. K. and J. Christopher Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60:4, 953-966.

Baillie, Richard T. and Tim Bollerslev (1992), "Prediction in Dynamic Models with Time-Dependent Conditional Variances," *Journal of Econometrics* 52:1-2, 91-113.

Bera, Anil K. and Sangwhan Kim (1996), "Testing Constancy of Correlation with an Application to International Equity Returns," University of Illinois at Urbana-Champaign, CIBER Working Paper 96-107.

Bollerslev, Tim (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics* 31, 307-327.

Bollerslev, Tim (1990), "Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics* 72:3, 498-505.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge (1988), "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy* 96:1, 116-31.

Campa, Jose M. and P. H. Kevin Chang (1997), "The Forecasting Ability of Correlations Implied in Foreign Exchange Options," NBER Working Paper 5974 (March).

Diebold, Francis X. and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13:3 (July), 253-263.

Engle, Robert F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica* 50, 987-1008.

Engle, Robert F., Che-Hsiung Hong, Alex Kane, and Jaesun Noh (1993), "Arbitrage Valuation of Variance Forecasts with Simulated Options," *Advances in Futures and Options Research* 6, 393-415.

Engle, Robert F. and Kenneth F. Kroner (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory* 11:1, 122-50.

Engle, Robert F., Alex Kane, and Jaesun Noh (1997), "Index-Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts," *Review of Derivatives Research* 1, 139-157.

Figlewski, Stephen (1997), "Forecasting Volatility," *Financial Markets, Institutions and Instruments* 6:1, 1-88.

French, Kenneth R. and Richard Roll (1986), "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics* 17:1, 5-26.

Hamilton, James (1994), *Time Series Analysis*, Princeton University Press: Princeton, NJ.

Jorion, Philippe (1995), "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance* 50:2, 507-28.

Karolyi, G. Andrew and Rene Stulz (1996), "Why Do Markets Move Together? An Investigation of U.S.-Japan Stock Return Comovements," *Journal of Finance* 51:3 (July), 951-86.

Longin, François and Bruno Solnik (1995), "Is the Correlation in International Equity Returns Constant: 1960-1990?" *Journal of International Money and Finance* 14:1, 3-26.

Margrabe, William (1978), "The Value of an Option to Exchange One Asset for Another," *Journal of Finance* 33:1 (March), 177-86.

Mayhew, Stewart (1995), "Implied Volatility," *Financial Analysts Journal* (July-August), 8-20.

Nelson, Daniel B. (1991), "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica* 59: 347-70.

Noh, Jaesun, Robert F. Engle, and Alex Kane (1994), "Forecasting Volatility and Option Prices of the S&P 500 Index," *Journal of Derivatives*, 17-30.

RiskMetrics™ Technical Document (1996), Fourth edition, JP Morgan/Reuters.

Rubinstein, Mark (1995), "Return to Oz," in Robert Jarrow, ed., *Over the Rainbow: Developments in Exotic Options and Complex Swaps*, Risk Publications: London.

Sola, Martin and Allan Timmermann (1994), "Fitting the Moments: A Comparison of ARCH and Regime Switching Models for Daily Stock Returns," Birkbeck College Discussion Paper in Financial Economics 1/94 (February).

Strang, Gilbert (1980), *Linear Algebra and its Applications*, second edition, Harcourt Brace Jovanovich: San Diego.

Stulz, Rene (1982), "Options on the Minimum or the Maximum of Two Risky Assets," *Journal of Financial Economics* 10:2 (July), 161-85.

Walter, Christian and Jose Lopez (1997), "Is Implied Correlation Worth Calculating? Evidence from Foreign Exchange Options and Historical Data," Federal Reserve Bank of New York Research Paper #9730 (September).

West, Kenneth D. and Dongchul Cho (1994), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics* 69:2, 367-91.

Table 1. Forecasts used in the simulations

True parameters

Volatility = 14.1
Correlation = .6

Incorrect forecasts that are "close" to the truth

Volatility = {13.2, 13.6, 14.0, 14.3, 14.7, 15.0}
Correlation = {.5, .54, .58, .62, .66, .7}

Incorrect forecasts that are "far" from the truth

Volatility = {11.1, 12.2, 13.2, 15.0, 15.8, 16.6}
Correlation = {.3, .4, .5, .7, .8, .9}

Notes:
1. Volatility is expressed in percent per year.
2. The same set of incorrect volatility forecasts are used for both assets.
3. The volatility forecasts are evenly spaced when expressed as variances but look skewed when expressed as standard deviations.
4. The daily variance forecasts are annualized by multiplying by 250.

Table 2. Four rainbow options.

Name	Payoff function
Call spread	$\max(0,X_2-X_1-K)$
Outperformance option	$\max(X_1-K_1,X_2-K_2)$
Call worst	$\max(0,\min(X_1,X_2)-K)$
Basket call	$\max(0,w_1X_1+w_2X_2-K)$

Notes:
1. As described in the text, we focus on at-the-money options (K=0).
2. Analytic solutions for the prices of the call spread and the outperformance option for the case of K=0 were given by Margrabe (1978) and for the call worst by Stulz (1982). No analytic solution exists for the basket call (for any K) or for the call spread and the outperformance option when $K\neq0$. In those cases, we use the recombining binomial pyramid method of Rubinstein (1995) to price the option.

Table 3. Simulations using a call spread option and the basic methodology

Panel A: Incorrect forecasts are "close" to the truth

Agent:	1	2	3	4	5
	40	16	14	17	13
N=250	.047	.017	-.031	-.020	-.013
	.036	.053	.077	.073	.067
	58	20	4	12	6
N=1,000	.042	.023	-.030	-.023	-.013
	.020	.033	.050	.059	.048
	78	7	2	5	8
N=5,000	.043	.022	-.017	-.028	-.020
	.012	.021	.046	.051	.052

Panel B: Incorrect forecasts are "far" from the truth

Agent:	1	2	3	4	5
	59	11	2	13	15
N=250	.15	.077	-.089	-.081	-.052
	.060	.090	.16	.18	.17
	80	7	1	7	5
N=1,000	.15	.082	-.090	-.081	-.058
	.048	.085	.14	.17	.16
	85	2	0	6	7
N=5,000	.14	.078	-.066	-.090	-.067
	.038	.064	.14	.17	.18

Line 1 of each cell = number of trials out of 100 in which the agent had the highest profits
Line 2 of each cell = average across 100 trials of the mean profit per day.
Line 3 of each cell = average across 100 trials of the standard deviation of profit per day.

Agent 1 = correct correlation and volatilities
Agent 2 = correct correlation, incorrect volatilities
Agent 3 = correct volatilities, incorrect correlation
Agents 4 and 5 = incorrect correlation and volatilities

Table 4. Simulations using a call spread option and the option package methodology

Panel A: Incorrect forecasts are "close" to the truth

Agent:	1	2	3	4	5
N=250	30 .041 .047	23 .030 .053	16 -.033 .087	13 -.030 .090	18 -.010 .086
N=1,000	37 .036 .024	34 .036 .021	7 -.034 .059	10 -.023 .054	12 -.015 .050
N=5,000	48 .035 .013	38 .033 .014	4 -.018 .045	2 -.029 .044	8 -.021 .044

Panel B: Incorrect forecasts are "far" from the truth

Agent:	1	2	3	4	5
N=250	52 .13 .068	33 .12 .070	3 -.11 .16	6 -.087 .17	6 -.054 .16
N=1,000	55 .13 .039	43 .13 .037	1 -.11 .15	1 -.082 .15	0 -.065 .13
N=5,000	68 .13 .035	30 .12 .035	1 -.067 .14	0 -.10 .15	1 -.079 .14

Line 1 of each cell = number of trials out of 100 in which the agent had the highest profits
Line 2 of each cell = mean across 100 trials of the agent's mean profit per day.
Line 3 of each cell = standard deviation across 100 trials of the agent's mean profit per day.

Agent 1 = correct correlation and volatilities
Agent 2 = correct correlation, incorrect volatilities
Agent 3 = correct volatilities, incorrect correlation
Agents 4 and 5 = incorrect correlation and volatilities

Table 5. Simulations using an outperformance option and the option package methodology

Panel A: Incorrect forecasts are "close" to the truth

Agent:	1	2	3	4	5
	9	17	24	23	27
N=250	.0068	.0093	-.0026	-.029	.014
	.17	.17	.26	.33	.30
	23	25	14	20	18
N=1,000	.015	.0078	-.018	.0061	-.0068
	.10	.075	.16	.14	.12
	28	29	21	12	10
N=5,000	.019	.018	-.013	-.0019	-.020
	.029	.045	.076	.068	.067

Panel B: Incorrect forecasts are "far" from the truth

Agent:	1	2	3	4	5
	54	28	5	7	6
N=250	.068	.059	-.052	-.046	-.025
	.034	.039	.083	.090	.084
	62	30	2	2	4
N=1,000	.067	.063	-.049	-.038	-.034
	.025	.024	.077	.078	.070
	63	33	1	0	3
N=5,000	.065	.061	-.031	-.046	-.040
	.022	.022	.073	.079	.076

Line 1 of each cell = number of trials out of 100 in which the agent had the highest profits
Line 2 of each cell = mean across 100 trials of the agent's mean profit per day.
Line 3 of each cell = standard deviation across 100 trials of the agent's mean profit per day.

Agent 1 = correct correlation and volatilities
Agent 2 = correct correlation, incorrect volatilities
Agent 3 = correct volatilities, incorrect correlation
Agents 4 and 5 = incorrect correlation and volatilities

Table 6. Simulations using a call worst option and the option package methodology

Panel A: Incorrect forecasts are "close" to the truth

Agent:	1	2	3	4	5
	25	28	17	12	18
N=250	.021	.015	-.014	-.017	-.0037
	.023	.029	.049	.048	.046
	37	34	11	7	11
N=1,000	.018	.019	-.015	-.011	-.0079
	.016	.012	.032	.031	.027
	48	31	7	5	9
N=5,000	.018	.017	-.0091	-.013	-.0097
	.0078	.0079	.023	.022	.023

Panel B: Incorrect forecasts are "far" from the truth

Agent:	1	2	3	4	5
	55	27	2	10	6
N=250	.067	.059	-.051	-.045	-.026
	.034	.037	.082	.087	.082
	51	43	2	2	2
N=1,000	.065	.064	-.050	-.039	-.032
	.023	.020	.074	.075	.066
	69	29	1	0	1
N=5,000	.064	.061	-.032	-.046	-.038
	.018	.018	.070	.074	.071

Line 1 of each cell = number of trials out of 100 in which the agent had the highest profits
Line 2 of each cell = mean across 100 trials of the agent's mean profit per day.
Line 3 of each cell = standard deviation across 100 trials of the agent's mean profit per day.

Agent 1 = correct correlation and volatilities
Agent 2 = correct correlation, incorrect volatilities
Agent 3 = correct volatilities, incorrect correlation
Agents 4 and 5 = incorrect correlation and volatilities

Table 7. Simulations using a basket call option and the option package methodology

Panel A: Incorrect forecasts are "close" to the truth

Agent:	1	2	3	4	5
	20	25	17	19	19
N=250	.0097	.0099	-.0094	-.0089	-.0011
	.019	.018	.031	.034	.029
	33	11	26	14	16
N=1,000	.0079	.0097	-.0079	-.0062	-.0035
	.012	.0095	.020	.021	.018
	43	33	7	8	9
N=5,000	.0086	.0082	-.0047	-.0077	-.0044
	.0051	.0056	.013	.013	.012

Panel B: Incorrect forecasts are "far" from the truth

Agent:	1	2	3	4	5
	45	41	3	4	7
N=250	.030	.029	-.027	-.018	-.015
	.015	.015	.038	.036	.036
	42	57	1	0	0
N=1,000	.031	.035	-.028	-.020	-.015
	.0096	.0082	.032	.033	.032
	65	35	0	0	0
N=5,000	.030	.029	-.019	-.024	-.017
	.0077	.0075	.032	.031	.030

Line 1 of each cell = number of trials out of 100 in which the agent had the highest profits
Line 2 of each cell = mean across 100 trials of the agent's mean profit per day.
Line 3 of each cell = standard deviation across 100 trials of the agent's mean profit per day.

Agent 1 = correct correlation and volatilities
Agent 2 = correct correlation, incorrect volatilities
Agent 3 = correct volatilities, incorrect correlation
Agents 4 and 5 = incorrect correlation and volatilities

Table 8. Summary statistics on S&P 500 and NASDAQ composite continuously compounded returns

	S&P 500	NASDAQ
Daily returns		
Mean	.035 %	.042 %
SD	.878 %	.794 %
Daily returns expressed at an annual rate		
Mean	8.7 %	10.5 %
SD	13.9 %	12.5 %

Correlation of daily returns = 0.79

Table 9. Summary statistics of the five forecasts

	S&P volatility		Nasdaq volatility		Correlation	
Forecasting model	Mean	SD	Mean	SD	Mean	SD
NAIVE	14.9	3.7	13.1	2.9	.79	.036
EWMA	13.9	5.3	12.6	5.2	.79	.080
D-GARCH	14.0	4.0	11.8	4.1	.80	.062
C-GARCH	14.4	4.2	12.2	4.5	.83	.033
MARKOV	14.4	4.6	12.2	4.4	.81	.037

Notes:
1. Volatility forecasts are expressed at an annual rate.

Table 10. Mean parameters of the three forecasting models

	D-GARCH		C-GARCH		MARKOV	
AR(1) parameters						
S&P 500	a_0	.062	a_0	.065	a_0	.060
	a_1	.0050	a_1	.0015	a_1	.00081
Nasdaq	a_0	.067	a_0	.070	a_0	.073
	a_1	.25	a_1	.25	a_1	.24
S&P Variance						
	ω	.043	ω	.075	s_t=1	.45
	α	.053	α	.070	s_t=2	1.6
	β	.87	β	.81		
Nasdaq variance						
	ω	.049	ω	.056	s_t=1	.26
	α	.094	α	.13	s_t=2	1.5
	β	.81	β	.76		
Covariance/correlation						
	ω	.041	ρ	.83	s_t=1	.27
	α	.060			s_t=2	1.2
	β	.84				
Weekend effect						
	δ_1	.20	δ	.18	δ_1	.23
	δ_2	.21			δ_2	.24
Unconditional Markov state probabilities						
					Pr(s_t=1)	.77
					Pr(s_t=2)	.23

Table 11. Results using a call spread option and the option package methodology

Forecasting model	Mean daily profit	robust standard error	t-statistic (equal to zero)	t-statistic (equal to MARKOV)
NAIVE	-.034	.017	-2.0	3.1
EWMA	-.024	.015	-1.5	2.4
D-GARCH	.025	.017	1.5	0.4
C-GARCH	.008	.012	0.6	1.5
MARKOV	.035	.014	2.4	n/a

Note: the standard errors and t-statistics are robust to heteroscedasticity and autocorrelation and were computed using the technique of Andrews and Monahan (1992).

Table 12. Ranking of profits by year

Forecasting model	First	Second	Third	Fourth	Fifth
NAIVE	6	4	2	2	6
EWMA	0	3	7	5	5
D-GARCH	5	0	7	5	3
C-GARCH	1	8	3	4	4
MARKOV	8	5	1	4	2

Each cell in the table shows the number of calendar years in which the forecasting model in the row finished in the rank given by the column.

Table 13. Results using a call spread option and the option package methodology, excluding days when the Spearman rank correlation between correlation forecast and option package price is greater than -.5

Forecasting model	Mean daily profit	robust standard error	t-statistic (equal to zero)	t-statistic (equal to MARKOV)
NAIVE	-.020	.015	-1.4	2.4
EWMA	-.018	.015	-1.3	2.0
D-GARCH	.014	.015	0.9	0.6
C-GARCH	.008	.011	0.7	1.2
MARKOV	.029	.014	2.0	n/a

Notes:
1. 382 days were excluded because the Spearman rank correlation between correlation forecast and option package price was greater than -.5.
2. The standard errors are robust to heteroscedasticity and autocorrelation and were computed following the technique of Andrews and Monahan (1992).

Table 14. Results for the basic methodology (depend on both volatility and correlation forecasts)

Panel A: Call spread option

	Mean daily profit	robust standard error
NAIVE	-.049	.010
EWMA	.005	.009
D-GARCH	.034	.006
C-GARCH	.022	.007
MARKOV	-.012	.007

Panel B: Outperformance option

	Mean daily profit	robust standard error
NAIVE	-.15	.020
EWMA	.042	.015
D-GARCH	.079	.008
C-GARCH	.048	.009
MARKOV	-.023	.013

Panel C: Call worst option

	Mean daily profit	robust standard error
NAIVE	-.10	.015
EWMA	.011	.011
D-GARCH	.043	.008
C-GARCH	.042	.008
MARKOV	.006	.011

Panel D: Basket call option

	Mean daily profit	robust standard error
NAIVE	-.12	.017
EWMA	.023	.013
D-GARCH	.063	.007
C-GARCH	.041	.008
MARKOV	-.005	.011

Note: the standard errors are robust to heteroscedasticity and autocorrelation and were computed following the technique of Andrews and Monahan (1992).

Table 15. Mean squared error of "statistical" loss functions

Panel A: Correlation-specific loss function

	Mean squared error	p-value (MSE equal to NAIVE)
NAIVE	.711	n/a
EWMA	.713	.14
D-GARCH	.715	<.01
MARKOV	.718	<.01
C-GARCH	.724	<.01

Panel B: Equal weights on variances and covariances

	Mean squared error	p-value (MSE equal to D-GARCH)
D-GARCH	3.47	n/a
NAIVE	3.51	.32
C-GARCH	3.64	<.01
MARKOV	3.71	<.01
EWMA	3.73	.06

Note: p-values on the difference in mean squared error are based on the asymptotic test of Diebold and Mariano (1995). The asymptotic heteroscedasticity and autocorrelation robust standard errors of the difference in mean squared errors are computed following the technique of Andrews and Monahan (1992). The p-values shown are for a two-tailed test.

Figure 1. Profits using the call spread option and the option package methodology

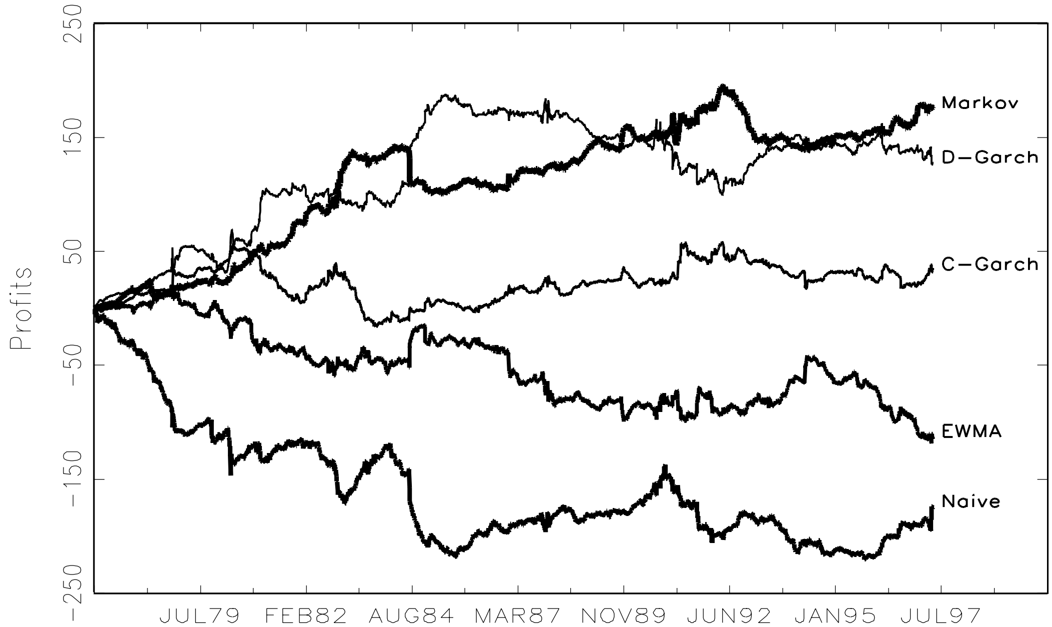

Figure 2. Profits using the call spread option and the option package methodology, excluding days when the Spearman rank correlation between correlation forecast and option package price is greater than -.5.