ABSTRACT

This paper proposes a multivariate nonparametric approach for hedging commodity risk in the financial market. The proposed approach is compared with the more conventional time series approach based on GARCH model. In our empirical analysis, the two approaches are applied to hedge against price risk inherent to crude oil contracts traded in the commodity exchanges. Diagnostic statistics are computed and out-of-sample performances are evaluated. In addition to the base models, our study examines the impact of incorporating the coefficient of absolute risk aversion and the transaction costs to the nonparametric and time series approaches. We find that the proposed nonparametric approach outperforms the time series approach as well as the un-hedged portfolio with the absence transaction costs. With transaction costs, nonparametric approach is better than the others if the hedger is risk averse. On the other hand, our empirical results indicate that explicit consideration of risk aversion does not alter the relative hedging performances of models and that the performance gaps among the models become more pronounced.

KEYWORDS: Financial market, commodity price risk, crude oil and energy sector, nonparametric and time series estimation, VEC-GARCH, Epanechnikov kernel regression, transaction cost, coefficient of absolute risk aversion

INTRODUCTION

Financial and commodity trading often involves a certain level of risk. Thus, it is not uncommon for institutional and individual investors to hedge against potential risk, such as undesirable / unfavorable price movements. Consequently, academic researchers and practitioners have developed numerous approaches to estimate the hedge ratios. Hedge ratio estimation of futures is important for investors and plays an important role when we estimate hedging effectiveness. Studies after studies have found that time-varying hedge ratios (THRs) lead to more risk reduction than traditional constant hedge ratios. This paper comparatively evaluates the use of nonparametric and times-series models to the observed time series of cash and futures in hedging. We apply the two approaches to obtain the THRs. Hence, the primary purpose of the paper is to assess the hedging effectiveness through out-of-sample testing. Also, our study attempts to answer the question of whether transaction costs alter the relative performances of
hedging schemes driven by nonparametric and time series approaches. Moreover, the empirical experiment investigates if the incorporation of coefficient of absolute risk aversion to the approaches affects our conclusions or not.

LITERATURE REVIEW

The applied economics literature has focused on the use of statistical models of observed time series of cash and futures prices in hedging. Early development of the type of optimal hedging is found in Johnson (1960), Peck (1975), and Kahl (1983), among others. This kind of hedgers considers not only risk but also returns. It makes progress to think of return of hedging-portfolio. In early stage, scholars hypothesize that hedgers only care risk, an assumption which lacks reasonable and supportive argument. Scholars, like Johnson, Peck, and Kahl, consider an agent with a non-tradable position in a cash commodity, who plans to buy or sell some number of commodity futures contracts that will maximize her utility. The notion traditionally involves choosing a level of hedging that would minimize the variance of changes in the hedger’s portfolio value by making static estimates of the variances of changes in the cash and futures prices and the covariance between those changes. This kind of estimation is still inadequate and out-of-date. Because it infers that hedge rations are constant, not time varying (i.e. dynamic). Since the price of financial asset varies from minute to minute, we must adopt more advanced technologies to estimate optimal hedge ratios. In this paper, we propose a time-varying hedging scheme guided by a nonparametric approach based on kernel regression. Its efficacy is compared to a more conventional time varying hedging method based on the parametric GARCH model. Both approaches are adapted to the traded crude oil contracts to estimate THRs.

Recently, the model of autoregressive conditional heteroskedastic (ARCH) [Engle (1982)] and generalized ARCH (GARCH) [Bollerslev (1986)] have been used to estimate time-varying hedge ratios (THRs). The time-varying joint distribution of cash and future price changes has been examined for hedging financial instruments [Cecchetti, Cumby and Figlewski (1988)]. Bivariate GARCH (BGARCH) models also have been used to estimate THRs in commodity futures [Bailie and Myers (1991); Myers (1991)], in foreign exchange futures [Kroner and Sultan (1991)], in interest rate futures [Gagnon and Lypny (1995)], and in stock index futures [Park and Switzer (1995)]. These more recent studies suggest that conventional hedging procedures can produce misleading results. It worth noting that the use of differenced data will lose information about the long-run relationship between two time series[Engle and Granger(1987)].To improve the appearance of things, we should consider error correction term into our model. Ghosh (1993) tests to determine whether the spot and futures price series are co-integrated and estimated the relevant error correction model. He found that the hedge ratio thus obtained was significantly better than the traditional regression approach. Kroner and Sultan (1993) use a bivariate ECM using GARCH error structure and find that their hedge ratio model provides more effective hedging. Chou, Denis and Lee (1996) apply ECM to Nikkei index and find that ECM gets more effective hedge ratio.

The nonparametric models we use in this paper are Kernel function, and its extension, kernel regression. Kernel function is also applied in many researches. Lien and Tse (2000, 2001) consider a futures hedge strategy that minimizes the lower partial moments (LPM). They use two statistical methods to estimate the optimal hedge ratios, and one of them is kernel function. Stanton and Whitelaw (1995) develop a new strategy for dynamically hedging mortgage-backed securities. They also use kernel function to estimate the time-varying hedge ratios.
Leung (2005) compare multivariate kernel function with Black-Scholes (adapted to account for skew) and the GARCH option pricing models.

MODELS AND METHODOLOGIES

Hedging Commodity Price Risk Using Time Series Approach

Typically, this type of hedging considers an agent with a non-tradable position in a cash commodity, who plans to buy or sell some of commodity futures contracts that will maximize her utility. We find a time-varying hedge ratio, $b_t$, which is the ratio of the size of the futures market position to the size of the cash market position. Then we apply the following equation to account for the change in the hedger’s portfolio value over the discrete interval from time $t-1$ to time $t$:

$$ P_t - P_{t-1} = (L_t - L_{t-1}) - b_{t-1}(F_t - F_{t-1}) $$

(1)

where $P_t$, $L_t$, and $F_t$ represent portfolio value, the local cash price of the commodity held by the hedger, and the futures price, respectively, in period $t$. There is one more thing we have to note when the hedger may not be able to deliver her commodity against the futures contract at par value locally, or she may be holding a different grade of the commodity than that specified in the futures contract. We therefore distinguish between a local cash price of an arbitrary commodity, and the price at the specified futures delivery location of the specified commodity. We refer to the former as a local cash price $L_t$ as above, and to the latter as the spot price $S_t$. In order to simplify the condition, we assume that the condition $L_t = S_t$ holds throughout the paper. In most situations, maximizing a mean-variance function is usually the hedger’s objective. This is equivalent to maximizing constant relative risk aversion utility when end-of-period terminal wealth is normally distributed. Furthermore, under such circumstances the mean-variance objective is the expected certainty equivalent income. The hedger’s objective can thus be represented by the following expression:

$$ \max_{b_{t-1}} \left( E \left( \Delta P_t \mid \Omega_{t-1} \right) - \frac{\lambda_U}{2} \text{Var} \left( \Delta P_t \mid \Omega_{t-1} \right) \right) $$

(2)

where $E()$ is the conditional expectation operator, $\Delta P_t$ is the change in portfolio value from $t-1$ to $t$, $\Omega_{t-1}$ is the information available as of $t-1$, $\lambda_U$ is the coefficient of absolute risk aversion, and $\text{var}()$ is the conditional variance operator. We can check if the CEI will be significantly different when $\lambda_U$ changes. In this paper, we let $\lambda_U$ take on the values of 0, 2, 4, and 10 for the sake of comparison. Risk-minimizing objective is a special case of Equation (2) when $\lambda_U = \infty$. Note that, given Equation (1), the conditional variance term in Equation (2) can be expanded to:

$$ \text{Var}(\Delta L_t \mid \Omega_{t-1}) + b^2_{t-1}\text{Var}(\Delta F_t \mid \Omega_{t-1}) - 2b_{t-1}\text{Cov}(\Delta L_t, \Delta F_t \mid \Omega_{t-1}) $$

(3)

where $\text{cov}()$ is the conditional variance operator. The objective-maximizing hedge ratio is then given by following equation:

$$ b_{t-1} = \frac{-\lambda_U \text{E}(\Delta F_t \mid \Omega_{t-1}) + \text{Cov}(\Delta L_t, \Delta F_t \mid \Omega_{t-1})}{\text{Var}(\Delta F_t \mid \Omega_{t-1})} $$

(4)
The second-order condition for this problem is the negative of the risk aversion coefficient multiplied by the conditional variance of changes in the futures price, and we are thus guaranteed a global maximum for a risk-averse hedger. We have the minimum-variance hedge ratio when the first term in the numerator is zero, and it means that \( \lambda_U = \infty \). When \( \infty > \lambda_U > 0 \), the optimal hedge ratio contains both the minimum-variance component, and a speculative component. This implies that an investor/trader should take into account adverse effects of risk as well as speculative motive. On the other hand, a hedger should predict the price of the futures to make avoid loss in or make profit from the portfolio. For example, an anticipated increase in the futures price will compel our hedger to increase the size of the futures position.

Given the above condition, we can now consider CEI with transaction costs. It is practical to combine CEI with transaction costs. It is because there are expenses incurred when investors buy futures for hedging. Further, financial assets’ prices changes as time goes by, and so do THRs. Hence, investors or hedgers can get better hedging effectiveness if they change their futures positions according to THRs. Nevertheless, it may not be wise to change futures positions every time THRs change. By doing so, it may incur excessive transaction costs. Investors should make transactions for rebalancing their portfolios when the benefits offset or exceed the costs. In other words, when the increased expected utility from rebalancing is great enough to offset the transaction costs expected to be incurred. This notion is summarized in Equation (5). Assume that \( y \) is the transaction cost of one futures contract. Observing from the futures market, the range of \( y \) is from 8~10 U.S dollars. Retail Investors have higher costs, 10 U.S dollars, while institutional investors only have 8 U.S dollars, and we decide to have \( y = 10 \) U.S dollars to be costs of each futures contract in the paper. A mean-variance expected utility-maximizing investor will rebalance at time \( t \) if and only if the following equation is satisfied (Kroner and Sultan (1993)):

\[
\text{CEI (brebalanced)} - 100^*y^*b_{\text{brebalanced}} > \text{CEI (bunrebalanced)} - 100^*y^*b_{\text{unrebalanced}}
\]  

where \( b_{\text{brebalanced}} \) is the newest rebalancing hedge ratio, and \( b_{\text{unrebalanced}} \) is the hedge ratio before rebalancing. Now we can decompose the time-varying hedge ratio, \( b \), into two parts. The conditional expected futures price change and conditional variance and covariance forecasts. The time-varying hedge ratio in Equation (4) requires the time-series modeler to provide the two kind of information.

Recent academic hedging research advocates obtaining the first piece of information using a vector error correction (VEC) model. This is an appropriate modeling technique in the event that each of the two price series is found to be non-stationary process, but a linear combination of the two is found to be stationary one (Engle and Granger, 1987). This linear combination is interpreted as representing a long-run equilibrium between the two levels series. The VEC model (VECM) is essentially a vector auto-regression model in which a deviation from the long-run equilibrium (the “error”) in one time period is subject to some degree of correction in the following time period. A basic representation of a VEC for two variables is as follow:

\[
\Delta y_t = \pi_0 + \sum_{i=1}^{r} \pi_i \Delta y_{t-i} + \alpha \beta y_{t-i} + \varepsilon_t
\]  

where \( y \) is the \( 2 \times 1 \) vector of observations at time \( t \), \( \pi_0 \) is a \( 2 \times 1 \) parameter vector, each \( \pi_i \) is a \( 2 \times 2 \) coefficient matrix, \( \beta \) is the co-integrating vector characterizing the long-run equilibrium, \( \alpha \) is a \( 2 \times 1 \) coefficient vector, and \( \varepsilon \) is a vector of innovations. The inner term \( \beta y_{t-i} \) is the deviation from the long-run equilibrium, and \( \alpha \) characterizes the rate at which each of the two
variables responds to this deviation. Equation (6) can then be used to generate forecasts of futures price changes – \( E(\Delta F_t | \Omega_{t-1}) \).

The other pieces of information that are required to calculate the THRs in equation (4) are the conditional variances and covariance-cov \( (\Delta L_t, \Delta F_t | \Omega_{t-1}) \) and \( \text{var}(\Delta F_t | \Omega_{t-1}) \). The two terms can be forecast using multivariate versions of the auto-regressive conditional heteroskedasticity (ARCH) model of Engle (1982) or the generalized ARCH (GARCH) model of Bollerslev (1986). A GARCH error structure implies that the conditional second moment of the innovation vector of a model follows an autoregressive, moving average process – it is a function of past innovation vectors and past second moments. Here we employ a multivariate GARCH \((1, 1)\) model with the diagonal vech parameterization of Bollerslev, Engle, and Wooldridge (1988). The conditional distribution of the error from Equation (6) is then given by:

\[
\begin{align*}
\epsilon_t | \Omega_{t-1} & \sim N(0, H_t) \\
\text{Vech}(H_t) &= \sigma^2 + A \text{vech}(\epsilon_{t-1} \epsilon_{t-1}^T) + \beta \text{vech}(H_{t-1})
\end{align*}
\]

were vech() is the column stacking operator that stacks the lower triangular portion of a symmetric matrix, \( W \) is a \( 3 \times 1 \) vector of constants, and \( A \) and \( B \) are a diagonal \( 3 \times 3 \) coefficient matrices. We show the diagonal vech parameterization of Bollerslev, Engle, and Wooldridge (1988) equation again.

\[
\begin{bmatrix}
H_{11t} \\
H_{12t} \\
H_{22t}
\end{bmatrix}
=\begin{bmatrix}
W_{01} \\
W_{02} \\
W_{03}
\end{bmatrix}
+\begin{bmatrix}
A_{11} & 0 & 0 \\
0 & A_{22} & 0 \\
0 & 0 & A_{33}
\end{bmatrix}
\times
\begin{bmatrix}
\epsilon_{11,t-1}^2 \\
\epsilon_{1,t-1} \epsilon_{2,t-1} \\
\epsilon_{22,t-1}^2
\end{bmatrix}
+\begin{bmatrix}
B_{11} & 0 & 0 \\
0 & B_{22} & 0 \\
0 & 0 & B_{33}
\end{bmatrix}
\times
\begin{bmatrix}
H_{11,t-1} \\
H_{12,t-1} \\
H_{22,t-1}
\end{bmatrix}
\]

or

\[
\begin{align*}
H_{11t} &= W_{01} + A_{11} \epsilon_{1,t-1}^2 + B_{11} h_{11,t-1} \\
H_{12t} &= W_{02} + A_{22} \epsilon_{1,t-1} \epsilon_{2,t-1} + B_{22} h_{12,t-1} \\
H_{22t} &= W_{03} + A_{33} \epsilon_{2,t-1}^2 + B_{33} h_{22,t-1}
\end{align*}
\]

where \( h_{11,t}, h_{22,t} \) are the conditional variance of the errors(\( \epsilon_{1t}, \epsilon_{2t} \)) from the mean equations, which in this application is the bivariate VAR model (with error correction term) and \( h_{12} \) represent the conditional covariance between spot and futures.

Equations (7) and (8) can be used to form forecasts of the variance of futures price changes and the covariance between futures and cash price changes. The VEC-GARCH model described by Equations (6) to (8) provides a way to estimate THRs and thus to assess the hedging effectiveness of both nonparametric and time series approaches in the out-of-sample period.

**Hedging Commodity Price Risk Using Multivariate Nonparametric Approach**

In the last section, we described the parametric VEC-GARCH model, to estimate THRs. Now, we introduce another approach, the nonparametric kernel regression, to estimate THRs. Here
are the differences between the two approaches. Parametric modeling enjoys a number of advantages. First, the approach is easy to apply and to understand, such as a linear model:

\[ y = \mu_1 + \mu_2 x_i + e; \ i = 1, 2 \ldots, n \] (9)

The properties of the estimators are known and there is a methodology available to compute confidence intervals, to carry out hypothesis tests, diagnostic checking etc. But the parametric approach has one major shortcoming, namely its lack of flexibility, especially when we have very limited information about the data that we have to analyze. The nonparametric approach avoids making stringent assumptions about the underlying relationship between the independent variables \((x_i)\) and the dependent variable \((y)\). Instead, it assumes that \(y = m(x) + e\); where \(m(\cdot)\) represents some function and the residual \((e)\) term is identical and independent distribution with \(E(e) = 0\) and \(Var(e) = \sigma^2\).

Multivariate nonparametric model is capable of performing a nonlinear projection and functional mapping in \(N\)-dimensional space. Classical parametric methods rely on a specific model of the signal of interest, and seek to compute the parameters of this model in the presence of noise. A generative model based upon the estimated parameters is then produced as the best estimate of the underlying model. In contrast, non-parametric methods rely on the data itself to dictate the structure of the model, in which case the implicit model is referred to as a regression function. So the nonparametric model does not require the pre-specification of functional forms prior to estimation. Hence, this model may be superior to traditional parametric models in generating an unknown conditional mean in the absence of knowledge regarding the functional forms for the conditional mean. The nonparametric model is also useful when the unknown functional form describing the process of interest is nonlinear. We use the nonparametric model to estimate the objective hedge ratio. And now we will introduce the nonparametric technique adopted in the paper. Interested readers should refer to Chen and Leung (2005), and Casdagli and Eubank (1992) for more detail exposition of the methodology.

The multivariate nonparametric approach proposed in the paper is the generalized kernel estimator of an unknown density function \(f(x)\), which is expressed by the equation:

\[
\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)
\] (10)

where \(x\) is a \(p\)-dimensional vector, \(n\) is the number of observations, \(K()\) represents the kernel function, and \(H\) is a bandwidth or smoothing parameter matrix. \(X\) is differenced from all the other \(X_i\) rows of the data set with each of these differences scaled by \(H\) and assigned a probability mass via \(K\). \(X\) is a given arbitrary row or vector. One can think of a kernel estimate in terms of a standardized distance between a point and each data point which is then converted into a probability based on this distance. Points close to the data get larger probability than points farther away. The multivariate nonparametric estimation contains two parts. One is kernel function \(K()\) and the other is the bandwidth \(H\). One popular class of kernel functions is the symmetric beta family which includes the normal density, the Epanechnikov (1969) kernel, and the bi-weight kernel as special cases. Our paper employs the product Epanechnikov kernel for the kernel function \(K()\). The Epanechnikov kernel is optimal based on the calculus solution of minimizing the integrated mean square error of the kernel estimator. The multivariate Epanechnikov kernel is given by the formula:
\[ K(z) = \prod_{j=1}^{n} k(z_j) \]  

for \( z = (z_1, z_2, \ldots, z_p) \) and 

\[ k(z_j) = \begin{cases} 
\frac{3}{4\sqrt{5}}(1 - \frac{1}{5} z_j^2) & \text{if } z_j^2 < 5.0 \\
0 & \text{otherwise}
\end{cases} \]  

(11) (12)

According to Scott (1992) and Hardle (1990, 1991), the choice of bandwidth is more crucial than the choice of kernel function for successful estimation of the proper response. We have mentioned that, given the value of \( x \), data closer to it get more probability than those far from it. Hence, we can roughly view that if data are not very meaningful in the estimation of the dependent variable if they are too far away from the vicinity of \( x \). This is due to the very limited information they have on the dependent variable. The function of bandwidth is to discriminate between the short distance and the long distance to the observed \( x \). If data do not locate in the range of \( x \)’s anticipated state space, then it is not effective too. Based on Scott’s conclusion, the asymptotically optimal bandwidth performs generally well for independent multivariate normal distribution. The expression of the asymptotically optimal bandwidth is:

\[ H = \text{diag}(\Omega^{-0.5}n^{-1/(p+4)}) \]  

(13)

where \( \Omega^{-0.5} \) is the variance-covariance matrix of covariate \( x \), \( n \) is the number of observations, and \( p \) is the number of variables.

Let \( f(y, x) \) denote the joint density of a set of random variables of interest \((Y, X)\), where \( Y \) is a scalar random variable, \( X \) is a \( p \)-dimensional vector of inputs (explanatory factors) and let \((y, x)\) be a realization of \((X, X)\). The density of \( Y \) conditional on \((X = x)\) will be denoted by \( g(x) = f(y, x) / f_i(x) \) where \( f_i(x) \) denotes the marginal density of \( X \). The conditional mean of \( Y \) with respect to a vector \( X \) of explanatory factors is defined as:

\[
E(Y|X=x) = \int y \frac{f(y, x)}{f_i(y, x)} dy = \int y \frac{f(y, x)}{f_i(x)} dy = \int y g(x) dy
\]  

(14)

Substituting the appropriate kernel estimators and simplifying the terms yield the following estimator:

\[
Y_i^* = m(X) = E(Y|X=x) = \frac{\sum_{i=1}^{n} K(H^{-1}(x-x_i))Y_i}{\sum_{i=1}^{n} K(H^{-1}(x-x_i))}
\]  

(15)

where \( H \) denotes the appropriate bandwidth parameter matrix for the covariate vector \( x \). In equation (4), we can see the hedge ratio is consisted of three variables, \( E(\Delta F_i|\Omega_{t-1}) \), \( \text{cov}(\Delta L_i, \Delta F_i|\Omega_{t-1}) \), and \( \text{var}(\Delta F_i|\Omega_{t-1}) \). So our explanatory variables \( X \) are \( E(\Delta F_i|\Omega_{t-1}), \text{cov}(\Delta L_i, \Delta F_i|\Omega_{t-1}), \) and \( \text{var}(\Delta F_i|\Omega_{t-1}) \) in the out-of-sample sub-period. In order to get the out-of-sample hedge ratio \((Y_i^*)\), we put the in-sample data \( E(\Delta F_i|\Omega_{t-1}), \text{cov}(\Delta L_i, \Delta F_i|\Omega_{t-1}), \)
and \( \text{var}(\Delta F_t|\Omega_{t-1}) \) as \( x_i \)'s in Equation(15), and then apply the estimated model to the out-of-sample \( X \). The purpose that we put the out-of-sample forecast \( \text{cov}(\Delta L_t \Delta F_t|\Omega_{t-1}) \), and \( \text{var}(\Delta F_t|\Omega_{t-1}) \), and \( \text{E}(\Delta F_t|\Omega_{t-1}) \) into \( X \) is to see whether or not the \( X \) has any information about \( Y_i^* \) (the out-sample hedge ratios). In the next step, we estimate the optimal in-sample hedge ratio \( b_{t-1}^* \), which makes CEI maximized. Finally, we put the in-sample \( b_{t-1}^* \) into \( Y_i \). After the above process, we get the objective out-of-sample hedge ratios (\( Y_i^* \)). By applying the Equation (2), we can determine the hedging effectiveness of the two approaches.

We can show the nonparametric kernel regression used for the estimation of THR in a rather straightforward equation:

\[
Y_i^*(E, C, V) = \sum_{i=1}^{n} k\left( \frac{E - E_i}{h_E} \right) * k\left( \frac{C - C_i}{h_C} \right) * k\left( \frac{V - V_i}{h_V} \right) * Y_i
\]

\[
= \sum_{i=1}^{n} k\left( \frac{E - E_i}{h_E} \right) * k\left( \frac{C - C_i}{h_C} \right) * k\left( \frac{V - V_i}{h_V} \right)
\]

where \( E \) is \( \text{E}(\Delta F_t|\Omega_{t-1}) \), \( C \) is \( \text{cov}(\Delta L_t \Delta F_t|\Omega_{t-1}) \), and \( V \) is \( \text{var}(\Delta F_t|\Omega_{t-1}) \). \( h_E, h_C, \) and \( h_V \) are the respective bandwidths for the three variables; \( K \) is the kernel function; \( Y_i \) is the in-sample hedge ratios; \( Y_i^* \) is the final estimated out-of-sample hedge ratios.

**DATA AND DIAGNOSTIC STATISTICS**

**Data Description**

Our dataset, provided by the CRB Database, is based on end-of-the-week observations of the New York Mercantile Exchange (NYMEX) crude oil futures contracts, and their associated spot prices. The futures and spot price data are observed over the period from January 1995 through June 2014. We split each data series into two sub-periods. The first time period, from January 1995 through December 2007, is used for in-sample parameter estimation. Out-of-sample hedging effectiveness is evaluated over the second time period, from January 2008 through June 2014. It should be noted that there is one NYMEX crude oil futures delivery in each month. Hence, the price data for individual futures contracts are patched to construct a rolling nearby futures series (\( \text{NEAR} \)), which are then used to estimate the parameters in the in-sample sub-period and to evaluate hedging effectiveness during the out-of-sample sub-period. Cash position is 100,000 barrels. One futures contract contains 1,000 barrels.

**Descriptive and Diagnostic Statistics**

In order to have a better understanding of the data, basic descriptive statistics for the time series (both cash and futures) are computed and various diagnostic tests on autocorrelation, normality, heteroskedasticity, and stationarity are performed. Results are obtained with respect to the in-sample and out-of-sample sub-periods.

Generally speaking, statistical results (not shown) suggest that the data series do not conform to normal distribution, are skewed, and have excessive kurtosis. It is a common feature that financial assets follow an asymmetric fat-tail distribution. The Ljung-Box Q(12) statistic tests for
autocorrelation is significant. Statistics from $Q^2(12)$ along with ARCH (6) indicate that heteroskedasticity condition exists. All statistical tests are significant at the 1% level.

Now we have to analyze the in-sample time series data. Augmented Dickey-Fuller (ADF) tests for unit roots are applied to the series over the in-sample estimation period. We utilize an ADF function with no time-trend but an intercept. The optimal lag length ($K$) was chosen according to Schwarz (1978) information criterion. The results in Table 1 indicate that the t-tests for cash and futures do not reject the null hypothesis, implying that cash and futures series are non-stationary. Given this, we test the first-difference of the series and find that the differenced series are both stationary (difference-stationary process). It follows that the first-order differenced cash and futures are used for subsequent estimations. Subsequently, the existence of co-integration relationship is confirmed by analyses based on the Engle-Granger (1987) test and ADF test on the residuals. From the last term in Table 1, we indicate that ECT series strongly rejects the null hypothesis of a unit root, and we can trust that S and NEAR have co-integration relation.

<table>
<thead>
<tr>
<th>Series</th>
<th>Lag length (k)</th>
<th>t-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash</td>
<td>0</td>
<td>-2.162</td>
</tr>
<tr>
<td>Cashdiff</td>
<td>0</td>
<td>-20.035</td>
</tr>
<tr>
<td>Futures</td>
<td>1</td>
<td>-3.120</td>
</tr>
<tr>
<td>Futuresdiff</td>
<td>2</td>
<td>-12.613</td>
</tr>
<tr>
<td>ECT</td>
<td>0</td>
<td>-18.985</td>
</tr>
</tbody>
</table>

*Tests for the presence of unit roots, using an intercept but no time trend. The critical value $-3.43$ (1%) is given in Fuller (1976). The optimal lag length ($K$) was chosen using the Schwarz (1978) information criterion.

**Parameter Estimation and Parameter Inference**

Before employing the multivariate GARCH (1, 1) model to estimate parameter, univariate GARCH (1, 1) is applied to see if the two first-difference series are appropriate or not. The specification of the GARCH model as well as the estimation results are shown in Table 2.

Table 2: Parameter estimates and residual diagnostics for the univariate GARCH(1,1) model:

\[
X_t = u + \varepsilon_t
\]

\[
\varepsilon_t | \Omega_{t-1} \sim N \left(0, h_t^2 \right)
\]

\[
h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}
\]
The numbers in parenthesis beside the parameter estimates are asymptotic standard errors. $m^3$ and $m^4$ are the sample skewness and sample kurtosis, respectively, of the standardized residuals. $Q(12)$ and $Q^2(12)$ denote Ljung-Box test statistics for 12th-order autocorrelation in the standardized and squared standardized residuals, respectively, with the numbers in parenthesis being the associated $p$-values.

Further, as reported in Table 2, $Q(12)$, which denotes the Ljung-Box (1978) test statistics for 12th-order autocorrelation in the standardized residuals, suggests no autocorrelation phenomenon while $Q^2(12)$, which denotes the Ljung-Box test statistics for squared standardized residuals, finds that ARCH effects do not exist. Hence, the univariate GARCH (1, 1) specification is confirmed and fits the data well. Under the assumption of normality, we use VECM-multivariate GARCH (1, 1) model to analyze our data further. Results from the analysis based on Schwarz (1978) information criterion, only the ECT term in the mean equation is required and neither constants nor autogressive terms are needed.

Given the findings, we generate parameter estimates with respect to the multivariate GARCH (1, 1) model. Results are obtained but not shown here. In summary, all estimated parameters are significant at the 5% significance level. The estimated coefficients, which serve as proxies for the speed of adjustment parameters, suggest that a considerable response of $s_t$ or $f_t$ to the previous period’s deviation from long-run equilibrium. Also, the results indicate that the error correction term is useful to increase the predictive ability of both the mean and the variance equations (Lee, 1994). The significance of the estimated coefficients in the model also points to the existence of both ARCH and GARCH effects.
RESULTS AND DISCUSSION

A primary purpose of our empirical analysis is to compare the hedging effectiveness of the proposed multivariate nonparametric approach and the more conventional parametric time series approach. To achieve this goal, we create two commodity portfolios made up of cash and futures positions. These portfolios are then guided by the two hedging approaches separately over the experimental investment/trading horizon. Technically, each portfolio is re-balanced dynamically by the time-varying THR estimated by Epanechnikov kernel regression or VEC-GARCH model. This simulation mimics the fact that, as new innovation is exposed in the financial market, cash and futures price reactions will cause the time-varying hedge ratio change to respond to the new information. In order to truly observe the effectiveness on commodity hedging, an un-hedged portfolio is also added to the analysis for benchmark comparison. For the current study, we adopt CEI as a common evaluation criterion for the hedging effectiveness of all portfolios as described earlier in the paper. In other words, the CEI measure will quantify the relative efficacy of the underlying statistical approaches during crude oil trading period (i.e., the out-of-sample testing sub-period).

To provide a more comprehensive examination of the approaches and possibly to yield additional insights on how externalities influence hedging performance, we incorporate two control factors into our empirical investigation. The first one is hedger’s transaction costs. The second one is hedger’s risk aversion level, which is represented by the coefficient of absolute risk aversion, \( \lambda_U \), in Equation (2). When \( \lambda_U \) is equal to 0, an investor is, essentially, fearless and completely risk neutral. When the parameter is equal to 2, an investor is risk averse and his utility decreases one unit for each unit of increase in variance (i.e., proportionally). However, when lambda is greater than 2, an investor is highly risk averse and his utility decreases more than one unit for each unit of increase in the variance (i.e., exponentially). Past literature such as Gagnon, Lypny, and McCurdy (1998) and Haigh and Holt (2000) simply set \( \lambda_U \) equal to 2 in their experimental analyses. Our paper explores the impact of risk aversion level on hedging through the use of a spectrum of values of \( \lambda_U = 0, 2, 4, \) and 10.

Hedging effectiveness (CEI) of the un-hedged portfolio along with the hedged portfolios guided by the two statistical approaches over the out-of-sample sub-period are displayed in Tables 3 and 4. Table 3 reports the scenario of no transaction costs whereas Table 4 shows the case with transaction costs. The tables also show the empirical results when different degrees of hedger’s risk aversion (\( \lambda_U = 0, 2, 4, 10 \)) are considered.

<table>
<thead>
<tr>
<th>( \lambda_U )</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average CEI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Un-hedged</td>
<td>-1.75E+03</td>
<td>-1.35E+09</td>
<td>-2.16E+09</td>
<td>-5.72E+09</td>
</tr>
<tr>
<td>Hedged by VEC-GARCH</td>
<td>-3.74E+02</td>
<td>-1.97E+08</td>
<td>-2.68E+08</td>
<td>-7.05E+08</td>
</tr>
<tr>
<td>Hedged by kernel regression</td>
<td>-3.62E+02</td>
<td>-1.80E+08</td>
<td>-2.47E+08</td>
<td>-6.36E+08</td>
</tr>
</tbody>
</table>

*The best performer in each category is highlighted.

With the absence of transaction costs, it can be observed in Table 3 that the hedging performance is the best when Epanechnikov kernel regression is used to guide the oil commodity portfolio position. Its average CEI is larger than the hedged portfolio driven by VEC-
GARCH time series approach which, in turn, is superior to the un-hedged portfolio. This rank is true across different hedger’s risk aversion level ($\lambda_U = 0, 2, 4, 10$). Also, the value of CEI becomes smaller when a hedger tends to be more and more risk averse. Besides, the performance gaps among kernel regression, VEC-GARCH and un-hedged portfolios become more pronounced as risk aversion increases.

Table 4 shows the corresponding results when transaction costs are incorporated into the hedging decision. The findings generally conform to those in Table 3 with a major exception – the un-hedged portfolio performs the best when $\lambda_U = 0$ (i.e., risk neutral), although it is quite close to the results of kernel regression. Essentially, the portfolio guided by multivariate kernel regression performs the best when transaction costs and risk aversion exist. Also, the performance gaps among the three portfolios become more pronounced as $\lambda_U$ becomes larger. However, a comparison between Tables 3 and 4 indicates that these gaps are much wider when transaction costs are considered in decision making. To further explore the reason behind these observations, we collect the numbers of re-balances occurred during the out-of-sample sub-period. They are tabulated inside the parentheses in Table 4. It can be seen that VEC-GARCH initiates much more re-balances than kernel regression, pointing to the possibility of over-trading or excessive overall transaction costs.

<table>
<thead>
<tr>
<th>$\lambda_U$</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average CEI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Un-hedged</td>
<td>$-4.254E+02$</td>
<td>$-1.607E+09$</td>
<td>$-4.613E+09$</td>
<td>$-3.836E+09$</td>
</tr>
<tr>
<td>Hedged by VEC-GARCH</td>
<td>$-4.79E+02$ (189)</td>
<td>$-1.592E+08$ (154)</td>
<td>$-3.54E+08$ (152)</td>
<td>$-6.798E+08$ (149)</td>
</tr>
<tr>
<td>Hedged by kernel regression</td>
<td>$-4.61E+02$ (128)</td>
<td>$-1.165E+08$ (70)</td>
<td>$-2.684E+08$ (69)</td>
<td>$-5.035E+08$ (69)</td>
</tr>
</tbody>
</table>

*The best performer in each category is highlighted.
*Y is the cost of each contract = 10 U.S dollars.
*Numbers in the parenthesis are rebalancing times.
*CEI (b_{rebalanced})-100*y* b_{rebalanced} > CEI (b_{unrebalanced})-100*y* b_{unrebalanced}

CONCLUSIONS

Determining a good approach to estimate accurate THRs is a key aim for every rational trader / hedger. Essentially, we pick up the approach which maximizes the CEI. In this paper, we analyze the relative hedging effectiveness of multivariate nonparametric and parametric time series approaches. The analysis is subject to with and without transaction costs and different coefficient of absolute risk aversion ($\lambda_U$). It is obvious that different $\lambda_U$ have little effect on our result. It is found that the nonparametric kernel regression is the best strategy for hedging with the absence of transaction costs. With transaction costs, it is also the best when the hedger is risk averse. Moreover, performance gaps among nonparametric kernel regression, parametric VEC-GARCH and un-hedged portfolios become more pronounced as risk aversion increases. Crude oil is a pearly resource and is pretty hard to be replaced given its pervasive role in the global energy sector. Many factors can influence its price. Further studies should point to development of models capturing academic theories and industrial practices.
REFERENCES


