Modelling and forecasting the volatility of petroleum futures prices

Sang Hoon Kang [a], Seong-Min Yoon [b, *]

[a] Department of Business Administration, Pusan National University, Busan 609-735, Korea

[b] Department of Economics, Pusan National University, Busan 609-735, Korea

Abstract

We investigate volatility models and their forecasting abilities for three types of petroleum futures contracts traded on the New York Mercantile Exchange (West Texas Intermediate crude oil, heating oil #2, and unleaded gasoline) and suggest some stylized facts about the volatility of these futures markets, particularly in regard to volatility persistence (or long-memory properties). In this context, we examine the persistence of market returns and volatility simultaneously using the following ARFIMA-GARCH-class models: ARFIMA-GARCH, ARFIMA-IGARCH, and ARFIMA-FIGARCH. Although the ARFIMA-FIGARCH model better captures long-memory properties of returns and volatility, the out-of-sample analysis indicates no unique model for all three types of petroleum futures contracts, suggesting that investors should be careful when measuring and forecasting the volatility (risk) of petroleum futures markets.

JEL Classification: C32; C52; G17; Q40

Keywords: DM test; forecasting ability; long memory; persistence; petroleum futures

[*] Corresponding author. Tel.: +82-51-510-2557; fax: +82-51-581-3143.
E-mail address: smyoon@pusan.ac.kr (S.-M. Yoon).

1. Introduction

Modelling and forecasting petroleum futures prices and their volatility are of great interest because accurately measuring the volatility of petroleum futures prices is an important component of the price linkage between spot and futures markets (Silvapulle and Moosa, 1999; Lin and Tamvakis, 2001; Hammoudeh, Li and Jeon, 2003; Hammoudeh and Li, 2004; Huang, Yang and Hwang, 2009), risk management, such as value-at-risk (Sadorsky, 2006; Aloui and Mabrouk, 2010), jump, or regime switching in energy futures markets (Fong and Kim, 2002; Lee, Hu and Chiou, 2010), and option pricing formulas for futures contracts (Wang, Wu and Yang, 2008). Thus, a better understanding of the dynamics of petroleum futures prices and their volatility should be useful to energy researchers, market participants, and policymakers.

Previous empirical studies have examined stochastic properties of petroleum futures prices by considering various econometric techniques and data frequencies. In particular, some have investigated whether times series of petroleum prices demonstrate long-memory properties of returns or volatility (Brunetti and Gilbert, 2000; Serletis and Andreadis, 2004; Elder and Serletis, 2008; Tabak and Cajueiro, 2007; Aloui and Mabrouk, 2010). Long memory is a particularly interesting feature in that its presence directly conflicts with the validity of the weak-form efficiency of the petroleum market. Thus, the presence of long memory provides evidence of nonlinear dependence and of a predictable component of returns and volatility.

Some empirical studies have addressed the modelling and forecasting of long-memory volatility in crude oil or petroleum markets using GARCH-type models (Sadorsky, 2006; Agnolucci, 2009; Kang, Kang and Yoon, 2009; Mohammadi and Su, 2010), but they have considered long memories in returns and volatility to be irrelevantly appearing phenomena. It is well-known that market shocks have considerable influence on returns and volatility at the same time, and thus they have dual long-memory properties. On the basis of this idea, some empirical studies have considered the relationship between returns and volatility for various economic and financial time series using a joint ARFIMA-FIGARCH model (Conrad and Karanasos, 2005a, 2005b; Kang and Yoon, 2007; Kasman, Kasman and Torun, 2009). The ARFIMA-FIGARCH model can facilitate the analysis of a relationship between returns and volatility for a process exhibiting dual long-memory properties.

The primary objective of this study was to model and forecast price volatility for three types of petroleum futures contracts traded on the New York Mercantile Exchange (NYMEX): West Texas Intermediate (WTI) crude oil, heating oil #2, and unleaded gasoline. This study extends the work of Kang, Kang and Yoon (2009) using ARFIMA-FIGARCH models that can capture long-memory properties of returns and price volatility for petroleum futures simultaneously. Additionally, this study demonstrates the superior predictability of ARFIMA-FIGARCH models using two forecast error statistics with multiple forecast horizons (e.g., 1-, 5-, and 20-day-ahead horizons).

The rest of this paper is organized as follows. Section 2 presents the statistical characteristics of the data. Section 3 discusses the ARFIMA-GARCH-class models and forecast error statistics. Section 4 presents the volatility model estimation and out-of-sample forecasting results, and Section 5 provides conclusions.

2. Data

We investigated the dynamics of futures prices of WTI crude oil, heating oil #2, and unleaded gasoline. In this report, "futures contracts" refer to those contracts with the earliest delivery date (Contract 1). Such futures contracts are traded on NYMEX, and data regarding these contracts are available from the U.S. Energy Information Administration (EIA).

Table 1
Descriptive statistics and unit root test results for petroleum futures returns

	WTI crude oil	Heating oil #2	Unleaded gasoline
Panel A: Descriptive statistics			
Mean	0.042	0.038	0.036
Std. dev.	2.302	2.415	2.578
Maximum	14.23	10.40	19.49
Minimum	-16.54	-20.97	-25.45
Skewness	-0.285	-0.673	-0.305
Kurtosis	6.335	8.681	9.349
Jarque-Bera	1431***	4260***	5085***
$Q(24)$	33.21	44.88***	26.41
$Q_s(24)$	182.32***	151.48***	76.97***
Panel B: Unit root tests			
ADF	-40.82***	-55.81***	-53.12***
PP	-53.35***	-56.26***	-53.18***
KPSS	0.070	0.075	0.039

Notes: The Jarque-Bera test corresponds to the test statistic for the null hypothesis of normality in the distribution of sample returns. The Ljung-Box statistics, $Q(n)$ and $Q_s(n)$, check for serial correlation of the return series and the squared returns up to the n^{th} order, respectively. MacKinnon's (1991) 1% critical value is -3.435 for the ADF and PP tests. The critical value for the KPSS test is 0.739 at the 1% significance level. *** indicates rejection of the null hypothesis at the 1% significance level.

The data used were of daily frequency for the period 3 January 1995 to 29 December 2006; data for the last one year were used to evaluate the accuracy of out-of-sample volatility forecasts.[1] The price series were converted into logarithmic percentage return series; that is, $y_t = 100 \times \ln(P_t / P_{t-1})$ for $t = 1, 2, \ldots, T$, where y_t indicates returns for each price at time t, P_t is the current price, and P_{t-1} is the price on the previous day. Following Sadorsky (2006), the actual daily volatility (variance) is measured by daily squared returns (r_t^2). Fig. 1 shows the dynamics of returns and price volatility for the three types of petroleum futures contracts.

Table 1 shows the descriptive statistics and the results of the unit root test for both sample returns. As shown in Panel A of Table 1, the mean of these return series is quite small, whereas the corresponding standard deviation of the returns is substantially higher. As indicated by the skewness, kurtosis, and Jarque-Bera results, the returns are not normally distributed. We also examined the null hypothesis of a white-noise process for sample returns using the Box-Pierce test for returns $Q(24)$ and squared returns $Q_s(24)$. The return series provides support for the null hypothesis of no serial correlation (except for heating oil futures), whereas the squared return series provides evidence of serial correlation at the 1% significance level.

Panel B of Table 1 presents the results of three types of unit root tests for each of the sample returns: augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski, Phillips, Schmidt, and Shin (KPSS). We considered the null hypothesis of a unit root for the ADF and PP tests and the null hypothesis of stationarity for the KPSS test. The null hypothesis of a unit root is rejected (large negative values),

[1] The long-memory property is often confused with structural breaks in a time series (Lamoreaux and Lastrapes, 1990; Diebold and Inoue, 2001). Structural breaks distort the long-memory property in returns and volatility. To avoid possible structural breaks, this paper excludes recent volatile price data.

whereas the null hypothesis of stationarity is not rejected at the 1% significance level. Thus, we concluded that the return series is a stationary process.

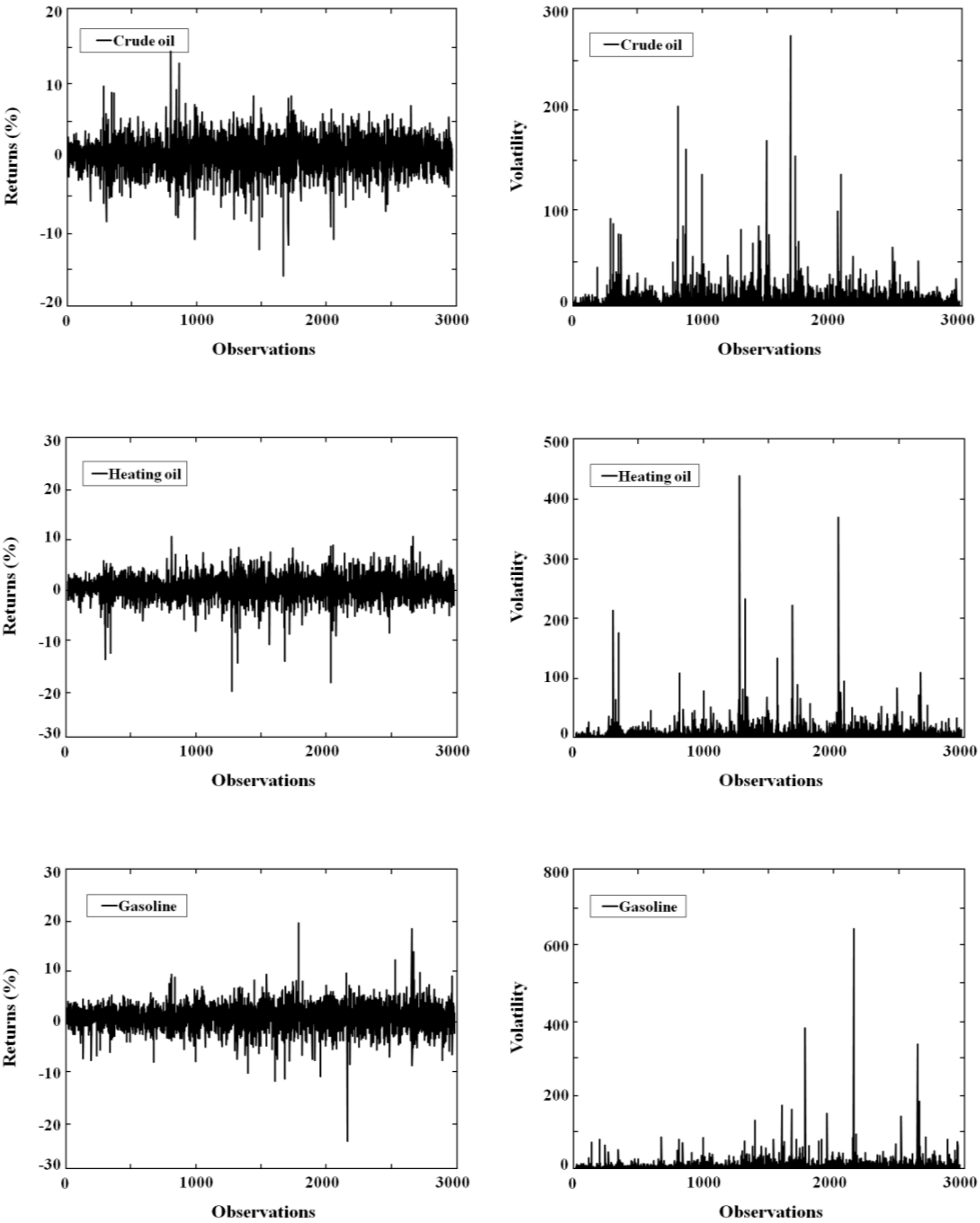

Fig. 1. Dynamics of daily returns and price volatility for three types of petroleum futures contracts.

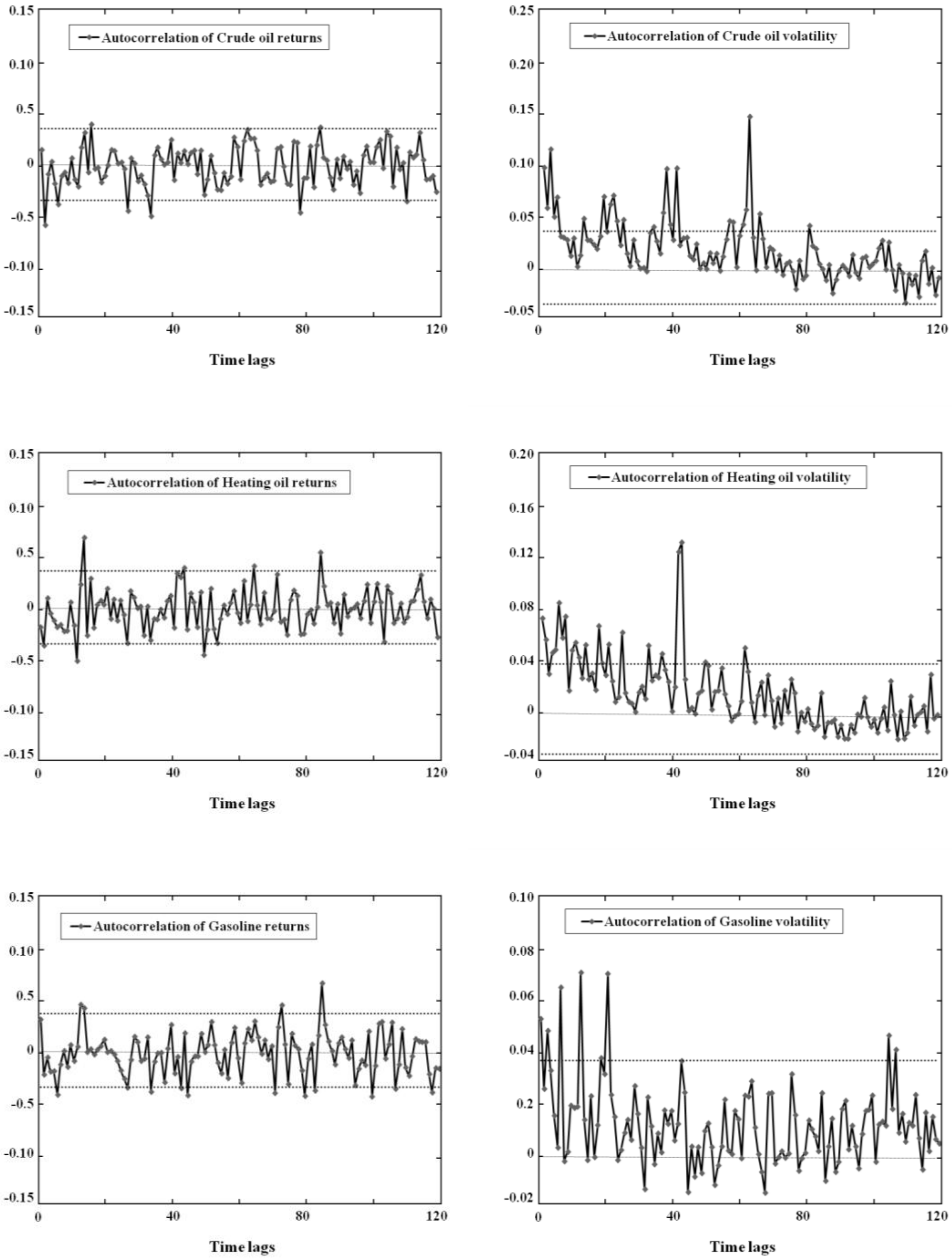

Fig. 2. Autocorrelation of daily returns and price volatility for three types of petroleum futures contracts.

Fig. 2 displays the autocorrelation function (ACF) of daily returns and volatility up to 120 time intervals with two-sided 5% critical values $\left(\pm 1.96\sqrt{1/T}\right)$. For the returns, most autocorrelations are small, and some significant autocorrelations die out quickly. There seems to be no systemic pattern in the return series of these petroleum futures contracts. However, the autocorrelations for the volatility series are significantly positive and persistence lasts for a substantial number of lags. This indicates that the volatility of petroleum futures contracts exhibits a long-memory process.

3. Model framework

3.1. ARFIMA-FIGARCH model

The ARFIMA model, a well-known parametric method for testing long-memory properties in financial time series, considers the fractionally integrated process $I(d)$ in the conditional mean. The ARFIMA (n,ξ,s) model can be expressed as a generalization of the ARIMA model, as follows:

$$\varepsilon_t = z_t \sigma_t, \quad z_t \sim N(0,1), \tag{1}$$

$$\Psi(L)(1-L)^{\xi}(y_t - \mu) = \Theta(L)\varepsilon_t, \tag{2}$$

where ε_t is independently distributed with variance σ_t^2, L denotes the lag operator, and $\Psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \cdots - \psi_n L^n$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_s L^s$ are, respectively, the autoregressive (AR) and moving-average (MA) polynomials for which all roots lie outside the unit circle. The parameters of the model are ξ, μ, ψ_i, and θ_i.

According to Hosking (1981), if $-0.5 < \xi < 0.5$, then the y_t process is stationary and invertible. For such processes, effects of shocks to ε_t on y_t decay slowly to zero. If $\xi = 0$, then the process is stationary (or short memory), and the effects of shocks to ε_t on y_t decay geometrically. For $\xi = 1$, the process follows a unit root process. If $0 < \xi < 0.5$, then the process exhibits positive dependence between distant observations, indicating long memory. If $-0.5 < \xi < 0$, then the process exhibits negative dependence between distant observations: that is, anti-persistence.

Similar research on volatility has extended the ARFIMA representation of ε_t^2, leading to the FIGARCH model of Baillie, Bollerslev and Mikkelsen (1996). The FIGARCH(p,d,q) can be expressed as follows:

$$\phi(L)(1-L)^d \varepsilon_t^2 = \omega + [1-\beta(L)]\nu_t, \tag{3}$$

where $\phi(L) \equiv \phi_1 L + \phi_2 L^2 + \cdots + \phi_q L^q$, $\beta(L) \equiv \beta_1 L + \beta_2 L^2 + \cdots + \beta_p L^p$, and $\nu_t \equiv \varepsilon_t^2 - \sigma_t^2$. The $\{\nu_t\}$ process can be interpreted as innovations for the conditional variance and is serially uncorrelated with a mean equal to zero. All the roots of $\phi(L)$ and $[1-\beta(L)]$ lie outside the unit root circle. The FIGARCH model provides greater flexibility for modelling the conditional variance in that it can accommodate the covariance stationary GARCH model for $d = 0$ and the nonstationary IGARCH model for $d = 1$. Thus, the attractiveness of the FIGARCH model is that for $0 < d < 1$, it is sufficiently flexible to allow for an intermediate range of persistence.

The parameters of the ARFIMA-FIGARCH model can be estimated using nonlinear optimization procedures to maximize the logarithm of the Gaussian likelihood function. Under the assumption that the random variable $z_t \sim N(0,1)$, the log-likelihood of the Gaussian or normal distribution (L_{Norm}) can be expressed as

$$L_{Norm} = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(2\pi) + \ln\left(\sigma_t^2\right) + z_t^2\right], \tag{4}$$

where T is the number of observations. The estimation procedure for ARFIMA-FIGARCH-class models requires a minimum number of observations. This minimum number is related to the truncation order of fractional differencing operators $(1-L)^{\xi}$ and $(1-L)^{d}$. Following the standard procedure used in previous research, we set the truncation order of infinite $(1-L)^{\xi}$ and $(1-L)^{d}$ to 1,000 lags as follows:

$$(1-L)^{\xi} = \sum_{k=0}^{1000}\frac{\Gamma(k-\xi)}{\Gamma(k+1)\Gamma(-\xi)}L^{k}. \tag{5}$$

3.2. Evaluation of forecasts

To measure forecasting accuracy, we calculated the mean square error (MSE) and mean absolute error (MAE) of volatility forecasts as follows:

$$MSE = \frac{1}{T}\sum_{i=1}^{T}(\sigma_{f,t}^2 - \sigma_{a,t}^2)^2, \tag{6}$$

$$MAE = \frac{1}{T}\sum_{i=1}^{T}\left|\sigma_{f,t}^2 - \sigma_{a,t}^2\right|, \tag{7}$$

where T denotes the number of forecast data points, $\sigma_{f,t}^2$ is the volatility forecast for day t, and $\sigma_{a,t}^2$ is the actual volatility on day t. A smaller forecast error statistic indicates the superior forecasting ability of a given model.

Although the above forecast error statistics are useful for comparing estimated models, they do not allow for statistical analyses of differences in forecast accuracy between two forecasting models. Thus, it is important to determine whether any reduction in forecast errors is statistically significant instead of comparing forecast error statistics between forecasting models. For this reason, Diebold and Mariano (1995) developed a test of forecast accuracy for two sets of forecasts. Having generated n, h-step-ahead forecasts from two different forecasting models, the forecaster has two sets of forecast errors $e_{1,t}$ and $e_{2,t}$, where $t = 1,2,\ldots,n$. With $g(e_{1,t})$ as a function of forecast errors, the hypothesis of equal forecast accuracy can be represented as $E\left[d_t\right] = 0$, where $d_t = g(e_{1,t}) - g(e_{2,t})$ and E is the expectation operator. The mean of differences between forecast errors $\bar{d} = n^{-1}\sum_{t=1}^{n} d_t$ has the approximate asymptotic variance of

$$V\left(\bar{d}\right) \approx n^{-1}\left[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right], \tag{8}$$

where γ_k is the k^{th} autocovariance of d_t, which can be estimated as

$$\hat{\gamma} = n^{-1}\sum_{t=k+1}^{n}\left(d_t - \bar{d}\right)\left(d_{t-k} - \bar{d}\right). \tag{9}$$

Diebold and Mariano's (1995) test statistic for the null hypothesis of equal forecast accuracy is

$$DM = \left[V\left(\hat{\bar{d}} \right) \right]^{-1/2} \bar{d}, \tag{10}$$

where DM has an asymptotic standard normal distribution under the null hypothesis. In this study, the DM test was conducted using the loss differential based on the MSE and MAE of the different forecasting models.

4. Empirical results

4.1. Testing long memory in petroleum futures contracts

To examine the long-memory property, we used several long-memory tests: Lo's (1991) modified R/S analysis and two semi-parametric estimators of the long-memory parameter: the log-periodogram regression (GPH) of Geweke and Porter-Hudak (1983) and the Gaussian semi-parametric (GSP) of Robinson and Henry (1999).[2] Panel A of Table 2 provides the results of Lo's R/S test for daily returns and volatility. For returns, the value of the modified R/S statistic supports the null hypothesis of short memory, while the volatility displays strong evidence of persistence. However, Panels B and C of Table 2 show that both the semi-parametric test (GPH and GSP tests) results reject the null hypothesis of short memory in returns and volatility of sample prices.[3] As a result, the evidence of long memory in returns is inconclusive by

[2] The choice of these alternative tests is justified by the fact that several authors have questioned the relevance of Lo's (1991) modified R/S. Practically, Lo's modified R/S analysis has a strong preference for accepting the null hypothesis of no long-range dependence, regardless of whether long memory is present in a time series (Hiemstra and Jones, 1997; Teverovsky, Taqqu and Willinger, 1999).

[3] The GPH test was implemented with different bandwidths: $m = T^{0.5}$, $m = T^{0.6}$, $m = T^{0.8}$. The GSP tes t statistic was also estimated with diverse bandwidths: $m = T/4$, $m = T/16$, $m = T/64$.

these different long-memory tests, while the volatilities of crude oil, heating oil, and gasoline seem to be well fitted by a fractionally integrated process. From this point, our research evolved with the ARFIMA-FIGARCH model to identify the long-memory property in returns and volatility in the three energy markets.

Table 2
Results of long-memory tests: Lo's R/S test, the GPH test, and the GSP test

	WTI crude oil	Heating Oil	Gasoline
Panel A: Lo's R/S test			
Returns	0.937	1.039	1.054
Volatility	3.452***	2.948***	2.988***
Panel B: GPH test			
Returns			
$m = T^{0.5}$	0.004	0.072	-0.046
$m = T^{0.6}$	-0.085	-0.059	-0.077
$m = T^{0.8}$	-0.045***	-0.053**	-0.068***
Volatility			
$m = T^{0.5}$	0.447***	0.433***	0.334***
$m = T^{0.6}$	0.316***	0.264***	0.264***
$m = T^{0.8}$	0.204***	0.153***	0.132***
Panel C: GSP test			
Returns			
$m = T/4$	-0.054***	-0.036**	-0.045***
$m = T/16$	-0.011	-0.036	-0.004
$m = T/64$	-0.071	-0.055	-0.124*
Volatility			
$m = T/4$	0.192***	0.150***	0.133***
$m = T/16$	0.289***	0.301***	0.245***
$m = T/64$	0.539***	0.435***	0.331***

Notes: The critical value of Lo's modified R/S analysis is 2.098 at the 1% significance level. m denotes the bandwidth for the GHP and the GSP tests. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

4.2. Estimation results using the ARFIMA-FIGARCH model

We estimated the ARFIMA-FIGARCH model described by Equations (2) and (3) to capture possible long-memory properties in both the mean and conditional variance. We also evaluated the performance of the ARFIMA-GARCH, ARFIMA-IGARCH, and ARFIMA-FIGARCH models in terms of their ability to capture long-memory properties of returns and volatility simultaneously. We estimated the ARFIMA-FIGARCH model using the quasi-maximum likelihood estimation method of Bollerslev and Wooldridge (1992).

Table 3 shows the estimation results obtained using these models.[4] In the mean equation, an ARFIMA $(0,\xi,1)$ model is the best representation of both WTI and heating oil returns for a long-memory process, whereas an ARFIMA $(1,\xi,0)$ is the best representation of the long-memory process for the gasoline returns. This suggests that gasoline prices have more pronounced short-run dynamics relative to those of WTI and heating oil prices, which are affected by market shocks. Generally, the results confirm the ability of the ARFIMA-FIGARCH model to capture the dynamics of returns for the three types of petroleum futures contracts. For example, the estimated values of the parameter ξ are negative and statistically different from zero, providing evidence of negative dependence (or anti-persistence) among the returns.[5] This result is consistent with the findings of Elder and Serletis (2008), who examined prices of crude oil futures contracts using GPH and wavelet OLS estimators.

---

[4] For the ARFIMA (*n*, *d*, *s*) specification in Equation (2), an MA (1) specification was retained for WTI and heating oil futures returns, whereas an AR (1) specification was chosen for unleaded gasoline futures returns.

[5] Anti-persistence is a form of long memory characterized by negative autocorrelation that decays very slowly. Peters (1994, p. 61) argued that anti-persistence time series reverses itself more often than a random one would. An anti-persistence process refers to a mean-reverting process.

The estimates of the long-memory parameter d are positive and significant at the 1% level, indicating rejection of $d = 0$ (GARCH model) and $d = 1$ (IGARCH model). This suggests that the volatility of petroleum futures returns has long-memory properties. Previous studies have reported similar findings (Brunetti and Gilbert, 2000; Sadorsky, 2006; Kang, Kang and Yoon, 2009; Aloui and Mabrouk, 2010).

In Table 4, we present the accuracy of model specifications using several diagnostic tests: three residual tests and three model selection criteria. To check the residual test, we applied the Box-Pierce test, $Q(24)$ for up 24^{th}-order serial correlation in the residuals, Engle's (1982) LM ARCH (10) test for the presence of ARCH effects in residuals up to lag 10, and the RBD (10) test for conditional heteroscedasticity in residuals up to lags 10.[6] Additionally, the Akaike information criterion (AIC), the Shibata criterion (SC), and the Hannan-Quinn criterion (HQ) were used to choose the best specification model among the given models in Table 3.

As presented in Table 4, the results of $Q(24)$ and the ARCH (10) show no serial correlation and no remaining ARCH effect. The insignificance of RBD (10) statistics indicates that the ARFIMA-FIGARCH model is suitable for depicting heteroscedasticity exhibited in the petroleum futures markets, indicating that there is no statistically significant evidence of misspecification in the ARFIMA-FIGARCH model. Additionally, the lowest values of three model selection criteria (AIC, SB, and HQ) indicate that the ARFIMA-FIGARCH model best captures the long-memory dynamics of both returns and price volatility simultaneously for petroleum futures contracts.

[6] Tse (2002) developed residual-based diagnostics (RBD) for conditional heteroscedasticity to test the n ull hypothesis of a correct model specification.

Table 3
Estimation results of volatility models

Series	WTI crude oil			Heating oil #2			Unleaded gasoline		
Model	ARFIMA -GARCH	ARFIMA -IGARCH	ARFIMA -FIGARCH	ARFIMA -GARCH	ARFIMA -IGARCH	ARFIMA -FIGARCH	ARFIMA -GARCH	ARFIMA -IGARCH	ARFIMA -FIGARCH
Mean equation									
μ	0.052 (0.025)**	0.050 (0.024)**	0.054 (0.026)**	0.045 (0.030)	0.047 (0.030)	0.042 (0.031)	0.040 (0.031)	0.028 (0.030)	0.043 (0.031)
ψ_1	-	-	-	-	-		0.101 (0.040)**	0.107 (0.040)**	0.097 (0.041)**
ξ	-0.088 (0.026)**	-0.091 (0.026)**	-0.078 (0.026)**	-0.057 (0.028)**	-0.058 (0.027)**	-0.055 (0.028)**	-0.073 (0.029)**	-0.077 (0.029)**	-0.072 (0.030)**
θ_1	0.103 (0.031)**	0.106 (0.030)**	0.093 (0.032)**	0.050 (0.034)	0.051 (0.033)	0.049 (0.034)	-	-	-
Variance equation									
ω	0.082 (0.054)	0.027 (0.017)	0.513 (0.200)**	0.110 (0.056)*	0.051 (0.025)**	0.275 (0.177)	0.197 (0.174)	0.047 (0.093)	0.645 (0.280)**
α_1	0.047 (0.017)**	0.051 (0.015)**	-	0.074 (0.021)**	0.079 (0.020)**	-	0.059 (0.041)	0.057 (0.074)	-
β_1	0.938 (0.023)**	1-0.051	0.415 (0.097)**	0.910 (0.024)**	1-0.079	0.550 (0.185)**	0.913 (0.060)**	1-0.057	0.538 (0.138)**
ϕ	-	-	0.191 (0.090)**	-	-	0.225 (0.132)*	-	-	0.362 (0.169)**
d	-	-	0.285 (0.065)**	-	-	0.409 (0.130)**	-	-	0.260 (0.081)**

Notes: Standard errors are in parentheses below the corresponding parameter estimates. ***, **, and * indicate rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively.

Table 4
Diagnostic tests of volatility models

Series	WTI crude oil			Heating oil #2			Unleaded gasoline		
Model	ARFIMA -GARCH	ARFIMA -IGARCH	ARFIMA -FIGARCH	ARFIMA -GARCH	ARFIMA -IGARCH	ARFIMA -FIGARCH	ARFIMA -GARCH	ARFIMA -IGARCH	ARFIMA -FIGARCH
$Q(24)$	17.46 [0.785]	16.84 [0.816]	17.90 [0.762]	23.38 [0.438]	22.03 [0.518]	22.07 [0.515]	15.61 [0.871]	16.72 [0.822]	16.17 [0.847]
$Q_s(24)$	26.43 [0.233]	27.30 [0.199]	25.71 [0.264]	20.17 [0.572]	23.90 [0.352]	17.59 [0.729]	22.10 [0.453]	20.84 [0.530]	22.09 [0.454]
ARCH (10)	1.179 [0.299]	1.243 [0.257]	0.912 [0.520]	0.670 [0.753]	0.824 [0.604]	0.599 [0.815]	1.245 [0.256]	1.188 [0.293]	1.186 [0.294]
RBD (10)	9.336 [0.500]	13.38 [0.202]	9.873 [0.451]	6.665 [0.756]	9.645 [0.472]	5.511 [0.854]	10.85 [0.369]	11.77 [0.300]	9.583 [0.477]
AIC	4.434757	4.437472	4.433171	4.499449	4.501740	4.499199	4.688999	4.694626	4.680864
SC	4.434749	4.437467	4.433160	4.499442	4.501734	4.499188	4.688991	4.694621	4.680853
HQ	4.439078	4.441073	4.438212	4.503770	4.505340	4.504240	4.693320	4.698227	4.685905
$\ln(L)$	-6151.81	-6157.01	-6148.79	-6216.46	-6219.97	-6215.16	-6432.54	-6440.29	-6420.67

Notes: $Q(24)$ and $Q_s(24)$ are Box-Pierce statistics for return series and squared return series, respectively, for up to 24th-order serial correlation. ARCH (10) is Engle's (1982) ARCH LM test to check the presence of ARCH effects in residuals up to lag 10. RBD (10) is the residual-based diagnostic for conditional heteroscedasticity, using 10 lags. $\ln(L)$ is the maximized Gaussian log-likelihood value. Numbers in brackets are p-values.

Table 5
Accuracy of out-of-sample forecasts for petroleum futures volatility

		Mean square error (*MSE*)						Mean absolute error (*MAE*)					
Series	Models	1-day horizon		5-day horizon		20-day horizon		1-day horizon		5-day horizon		20-day horizon	
		MSE	*DM*	*MSE*	*DM*	*MSE*	*DM*	*MAE*	*DM*	*MAE*	*DM*	*MAE*	*DM*
WTI crude oil	ARFIMA-FIGARCH	**19.74**	-	**19.80**	-	**20.37**	-	**3.77**	-	**3.77**	-	**3.82**	-
	ARFIMA-IGARCH	33.12	-7.19**	34.02	-3.93**	35.64	-3.13**	5.10	-10.77**	5.20	-5.31**	5.36	-2.77**
	ARFIMA-GARCH	21.78	-8.40**	21.99	4.62**	22.75	-2.69**	4.04	-10.19**	4.07	-4.47**	4.15	-3.35**
Heating oil #2	ARFIMA-FIGARCH	37.04	-9.69**	37.02	-8.99**	37.42	-9.61**	5.28	-14.41**	5.27	-13.97**	5.27	-12.56**
	ARFIMA-IGARCH	117.0	-13.8**	121.8	-6.42**	127.5	-3.38**	9.83	-17.62**	10.08	-8.00**	10.37	-4.16**
	ARFIMA-GARCH	**35.32**	-	**35.21**	-	**35.66**	-	**5.09**	-	**5.08**	-	**5.09**	-
Unleaded gasoline	ARFIMA-FIGARCH	115.7	-1.90*	118.7	-2.01**	124.6	-2.49**	7.26	-9.69**	7.28	-10.70**	7.47	-14.82**
	ARFIMA-IGARCH	192.9	-6.12**	200.4	-4.63**	211.5	-3.88**	11.89	-14.01**	12.17	-8.07**	12.66	-4.65**
	ARFIMA-GARCH	**112.6**	-	**116.1**	-	**122.0**	-	**6.62**	-	**6.66**	-	**6.84**	-

Notes: Values in bold type refer to the lowest value for both *MSE* and *MAE* statistics. The *DM* test statistic was used to evaluate the null hypothesis of no difference in forecast accuracy between the FIGARCH model and the GARCH or IGARCH model. ** indicates rejection of the null hypothesis for the *DM* test at the 5% significance level.

4.3. Out-of-sample forecast results

Although the ARFIMA-FIGARCH model captures the dynamics of the three petroleum futures time series well, an important question remains as to which ARFIMA-GARCH-class model best forecasts volatility. To address this, we evaluated 249 out-of-sample volatility forecasts between 3 January 2006 and 29 December 2006 and assessed the accuracy of these forecasts.

We obtained out-of-sample forecasts using parameter estimates for the volatility models in Table 5. Additionally, we tested the null hypothesis of no difference in forecast accuracy between the models using the DM test statistic in Equation (10). The out-of-sample forecast analysis considered 1, 5, and 20 forecast horizons, corresponding to 1-day, 1-week, and 1-month trading periods, respectively.

Table 5 presents the calculated values of the out-of-sample volatility forecast error statistics and the results of the DM test. In the case of WTI crude oil futures, the ARFIMA-FIGARCH model provides the lowest MSE and MAE values and shows a superior ability to forecast volatility for all three forecast horizons. It is noteworthy that in the cases of both heating oil #2 and unleaded gasoline futures contracts, the ARFIMA-GARCH model is more suitable than the other models (i.e., the ARFIMA-IGARCH and ARFIMA-FIGARCH models). Additionally, the values of the DM test statistic are negative and reject the null hypothesis of no difference at the 5% significance level, indicating better performance by the ARFIMA-GARCH model than the other models in these cases. Thus, the results of the out-of-sample analysis indicate that none of the models assessed provides the best fit for all of the three series considered.

In contrast to Kang, Kang, and Yoon (2009), who suggested that the fractionally integrated model provided the best fit for the volatility of crude oil spot prices, the results of the present study indicate that shocks to the volatility of heating oil and unleaded gasoline futures returns dissipate exponentially, pointing to the GARCH model. These findings have important implications for measuring value-at-risk estimations, determining optimal hedging ratios, and pricing derivatives in petroleum futures markets. For example, (1) an appropriate volatility model provides accuracy for capital reserve requirements in quantifying value-at-risk estimations (Fan et al., 2008; Aloui and Mabrouk, 2010), (2) accurate conditional variance from the volatility model is used for calculating hedging ratios and enhancing hedging effectiveness in the price change regression (Wilson, Aggarwal and Inclan, 1996; Zanotti, Gabbi and Geranio, 2010), and (3) an accurate long-memory volatility model is an important input in measuring option pricing in the Black-Scholes model (Bollerslev and Mikkelsen, 1996; Taylor, 2000).

5. Conclusions

In this study, we sought to identify a good model for forecasting volatility and examined some stylized facts about the volatility (particularly in regard to long memory or persistence) of three types of petroleum futures contracts. For this, we calculated the out-of-sample forecasts of the volatility and evaluated the performance of the ARFIMA-GARCH, ARFIMA-IGARCH, and ARFIMA-FIGARCH models in terms of their ability to capture long-memory properties of returns and volatility simultaneously.

The estimation results suggest that the ARFIMA-FIGARCH model can better capture long-memory features than can the ARFIMA-GARCH or ARFIMA-IGARCH models, indicating that returns and volatility for the three types of petroleum futures contracts have dual long-memory properties. The presence of long-memory properties casts doubt on the weak-form efficiency of petroleum futures markets.

However, the out-of-sample analyses suggest that none of the volatility models is adequate for all three petroleum futures series. This suggests that investors should be careful when measuring volatility (risk) in petroleum futures markets. The findings of this study should be useful in facilitating accurate value-at-risk management, developing futures pricing models, and determining optimal hedge ratios with respect to petroleum markets.

A number of avenues could be followed to extend this research. First, it would be interesting to consider high-frequency data in measuring the long-memory property in energy markets. Second, our long-memory result would be sensitive to the presence of structural breaks in energy markets. Thus, it would be worthwhile to include Markov switching-type volatility models in capturing regime shifts and comparing the forecasting ability with the long-memory volatility models. Third, it would be interesting to check the relevance of different return distributions to enhance the forecasting ability of volatility models.

Acknowledgments

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-371-B00008). The authors thank Chae Woo Nam, Jong Hyun Choi, and seminar participants at the 6th Conference of the Asia-Pacific Association of Derivatives (APAD 2010) for many helpful comments and stimulating discussions.

References

Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Economics 31 (2), 316–321.

Aloui, C., Mabrouk, S., 2010. Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. Energy Policy 38 (5), 2326–2339.

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74 (1), 3–30.

Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long memory in stock market volatility. Journal of Econometrics 73 (1), 151–184.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation of dynamic models with time varying covariances. Econometric Reviews 11 (2), 143–172.

Brunetti, C., Gilbert, C.L., 2000. Bivariate FIGARCH and fractional cointegration. Journal of Empirical Finance 7 (5), 509–530.

Conrad, C., Karanasos, M., 2005a. Dual long memory in inflation dynamics across countries of the Euro area and the link between inflation uncertainty and macroeconomic performance. Studies in Nonlinear Dynamics & Econometrics 9 (4), Article 5.

Conrad, C., Karanasos, M., 2005b. On the inflation-uncertainty hypothesis in the USA, Japan and the UK: a dual long memory approach. Japan and the World Economy 17 (3), 327–343.

Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics, 105 (1), 131–159.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3), 253–263.

Elder, J., Serletis, A., 2008. Long memory in energy futures prices. Review of Financial Economics 17 (2), 146–155.

Engle, R.F., 1982. Autoregresssive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.

Fan, Y., Zhang, Y.-J., Tsai, H.-T., Wei, Y.-M., 2008. Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach. Energy Economics 30 (6), 3156–3171.

Fong, W. M., Kim, H.S., 2002. A Markov switching model of the conditional volatility of crude oil futures prices. Energy Economics 24 (1), 71–95.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4 (4), 221–238.

Hammoudeh, S., Li, H., 2004. The impact of the Asian crisis on the behavior of US and international petroleum prices, Energy Economics 26 (1), 135–160.

Hammoudeh, S., Li, H., Jeon, B., 2003. Causality and volatility spillovers among petroleum prices of WTI, gasoline and heating oil in different locations, North American Journal of Economics and Finance 14 (1), 89–114.

Hiemstra, C., Jones, J.D., 1997. Another look at long memory in common stock returns. Journal of Empirical Finance 4 (4), 373–401.

Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68 (1), 165–176.

Huang, B.-N., Yang, C.W., Hwang, M.J., 2009. The dynamics of a nonlinear relationship between crude oil spot and futures prices: A multivariate threshold regression approach. Energy Economics 31 (1), 91–98.

Kang, S.H., Kang, S.-M., Yoon, S.-M., 2009. Forecasting volatility of crude oil markets. Energy Economics 31 (1), 119–125.

Kang, S.H., Yoon, S.-M., 2007. Long memory properties in return and volatility: evidence from the Korean stock market. Physica A 385 (2), 591–600.

Kasman, A., Kasman, S., Torun, E., 2009. Dual long memory property in returns and volatility: evidence from the CEE countries' stock markets. Emerging Markets Review 10 (2), 122–139.

Lamoreaux, C.G. Lastrapes, W.D., 1990. Persistence in variance, structural change and the GARCH model. Journal of Business & Economic Statistics 8(2), 225–234.

Lee, Y,-H., Hu, H.-N., Chiou, J.-S., 2010, Jump dynamics with structural breaks for crude oil prices. Energy Economics 32 (2), 343–350.

Lin, S.X., Tamvakis, M.N., 2001. Spillover effects in energy futures markets, Energy Economics 23 (1), 43–56.

Lo, A.W., 1991. Long-term memory in stock market prices, Econometrica 59 (5), 1279–1313.

MacKinnon, J.G., 1991. Critical values for cointegration tests. In: Engle, R.F., Granger, C.W.J. (Eds.), Long-Run Economic Relationships: Readings in Cointegration. Oxford University Press, New York, 266–276.

Mohammadi, H., Su, L., 2010. International evidence on crude oil price dynamics: applications of ARIMA-GARCH models. Energy Economics 32 (5), 1001–1008.

Peters, E. E., 1994. Fractal Market Analysis. John Wiley & Sons, NY.

Robinson, P.M., Henry, M., 1999. Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels. Econometric Theory 15 (3), 299–336.

Sadorsky, P., 2006. Modeling and forecasting petroleum futures volatility. Energy Economics 28 (4), 467–488.

Serletis, A., Andreadis, I., 2004. Random fractal structures in North American energy markets. Energy Economics 26 (3), 389–399.

Silvapulle, P., Moosa, I., 1999. The relationship between spot and futures prices: Evidence from the crude oil market. Journal of Futures Market 19 (2), 175–193.

Tabak, B.M., Cajueiro, D.O., 2007. Are the crude oil markets becoming weakly efficient over time? A test for time-varying long-range dependence in prices and volatility. Energy Economics 29 (1), 28–36.

Taylor, S.J., 2000. Consequences for option pricing of a long memory in volatility. EFA 2001 Barcelona Meetings, EFMA 2001 Lugano Meetings. Available at SSRN: http://ssrn.com/abstract=269840 or DOI: 10.2139/ssrn.26984.

Teverovsky, V., Taqqu, M.S., Willinger, W., 1999. A critical look at Lo's modified R/S statistic. Journal of Statistical Planning and Inference 80 (1–2), 211–227.

Tse, Y.K., 2002. Residual-based diagnostics for conditional heteroscedasticity models. Econometrics Journal 5 (2), 358–374.

Wang, T., Wu, J.T., Yang, J., 2008. Realized volatility and correlation in energy futures markets. Journal of Futures Markets 28 (10), 993–1011.

Wilson, B., Aggarwal, R., Inclan, C., 1996. Detecting volatility changes across the oil sector. Journal of Futures Markets 16 (3), 313–330.

Zanotti, G., Gabbi, G., Geranio, M., 2010. Hedging with futures: Efficacy of GARCH correlation models to European electricity markets. Journal of International Financial Markets, Institutions & Money 20 (2), 135–148.