Return Distributions and Volatility Forecasting in Metal Futures Markets: Evidence from Gold, Silver, and Copper

AHMED A. A. KHALIFA
HONG MIAO
SANJAY RAMCHANDER*

The characterization of return distributions and forecast of asset-price variability play a critical role in the study of financial markets. This study estimates four measures of integrated volatility—daily absolute returns, realized volatility, realized bipower volatility, and integrated volatility via Fourier transformation (IVFT)—for gold, silver, and copper by using high-frequency data for the period 1999 through 2008. The distributional properties are investigated by applying recently developed jump detection procedures and by constructing financial-time return series. The predictive ability of a GARCH (1, 1) forecasting model that uses various volatility measures is also examined. Three important findings are

The authors are grateful to Nancy Jianakoplos, Charles Revier, and Deepankar Basu for their helpful advice on this research. We also wish to thank the editor, Bob Webb, and an anonymous reviewer for their thoughtful and constructive comments. The authors are responsible for any remaining errors.

*Correspondence author, Department of Finance and Real Estate, Colorado State University, Fort Collins, Colorado 80523. Tel: (970) 491-6681, e-mail: Sanjay.Ramchander@colostate.edu

Received August 2009; Accepted January 2010

- *Ahmed A. A. Khalifa is in the Department of Economics, Bloomsburg University of Pennsylvania, Bloomsburg, Pennsylvania.*
- *Hong Miao and Sanjay Ramchander are in the Department of Finance and Real Estate, Colorado State University, Fort Collins, Colorado.*

The Journal of Futures Markets, Vol. 31, No. 1, 55–80 (2011)

Published online March 29, 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/fut.20459

reported. First, the magnitude of the IVFT volatility estimate is the greatest among the four volatility measures. Second, the return distributions of the three markets are not normal. However, when returns are standardized by IVFT and realized volatility, the corresponding return distributions bear closer resemblance to a normal distribution. Notably, the application of financial-time sampling technique is helpful in obtaining a normal distribution. Finally, the IVFT and realized volatility proxies produce the smallest forecasting errors, and increasing the time frequency of estimating integrated volatility does not necessarily improve forecast accuracy. © 2010 Wiley Periodicals, Inc. Jrl Fut Mark 31:55–80, 2011

INTRODUCTION

Volatility modeling is the subject of extensive investigation in the financial economics literature (see, for instance, Andersen, Bollerslev, & Diebold, 2005; Andersen, Diebold, Labys, & Bollerslev, 2001; Figlewski, 1997; Poon & Granger, 2003). The measurement and forecasting of volatility is essential for the characterization of market dynamics, valuation of assets, pricing of derivative instruments, and choices that affect portfolio allocation decisions. Furthermore, policy makers, including central banks often rely on volatility estimates to assess the vulnerability of financial markets and the economy.

The increased availability of high-frequency data has led to a renewed interest in understanding the return distribution properties, and developing new and improved volatility measurements that make use of information available in intraday prices. As a result various volatility measures have been examined in studies that provide fresh insights related to the distributional properties and dynamic dependencies in financial markets (for example, Andersen & Bollerslev, 1998a; Andersen, Bollerslev, Frederiksen, & Nielsen, 2009; Hansen & Lunde, 2005, 2006; Nielsen & Frederiksen, 2008). These studies show how intraday volatility measures can be used in the formulation of highly informative and directly testable distributional implications for discretely observed asset returns.

However, surprisingly very little of this research has been tested in the futures market. This article conducts a comprehensive examination of return distribution, volatility estimation, and volatility forecasting for three important metals—gold, silver, and copper. Interest in metal futures stem from the fact that these contracts are heavily traded, and are often viewed as a hedge against inflation and market uncertainty. A study of the metal futures prices is also important because of its role as a monetary medium and its application in a wide range of industries.

Some of the major research strands in metals futures market include exploring stochastic or unit root properties of prices (Krehbiel & Adkins, 1993), testing market efficiency (Aggarwal & Soenen, 1991; Lucey & Tully, 2006), examining conditional dependence in prices (Akgiray, Booth, Hatem, & Chowdary, 1991),

investigating co-movement among different metals (Ciner, 2001), and identifying dynamic relationships between futures and spot prices (Kocagil, 1997).

The study uses intra-day trading data for the period spanning 1999 through 2008, an expansive time period which allows us to evaluate the return distribution and model daily volatility in a manner that fully accounts for the process governing price variability. Four different measures of daily integrated volatility are estimated—absolute returns, bipower volatility, realized volatility, and integrated volatility using the Fourier transformation (IVFT). Among the four volatility measures, IVFT which was introduced by Malliavan and Mancino (2002) and the realized bipower volatility which was introduced by Barndoff-Nielsen and Shephard (2004) are fairly new techniques and their applications in the financial economics literature are still nascent. Consequently, the study places additional emphasis on the relative efficiency of these new measures in modeling and forecasting volatility for gold, silver, and copper.

The present research extends our understanding of the metals market in several ways. First, we examine the distributional properties of the three metal futures returns. Second, we estimate volatility using various integrated volatility measures and then provide evidence on the forecasting performance of these measures. Finally, we document how volatility forecasting performance changes as the intraday time frequency used to measure daily returns increase from 15-minute intervals to 1-minute intervals.

The remainder of this article is organized as follows. Second section reviews the methodology with a discussion of different volatility models, moment-based normality tests used to assess the distributional properties of daily returns, and loss functions used to compare competing volatility forecasts generated from a generalized autoregressive conditional heteroskedasticity (GARCH) (1, 1) model. Subsequent section discusses the data, data cleaning methods, and preliminary data characteristics. Penultimate section provides the empirical results and last section concludes the study.

METHODOLOGY

Volatility Measures

There are four major approaches to estimating integrated volatility in financial markets.

Absolute returns

The absolute return measure is sometimes viewed as an alternative to constant standard deviation used to calculate asset return volatility, and can be applied for data of any time frequency. This measure is somewhat popular among

empirical studies (see Cumby et al., 1993; Figlewski, 1997; West & Cho, 1995). It is estimated as follows:

$$\sigma_t = |R_t| = |p_t - p_{t-1}|, \tag{1}$$

where σ_t is the conditional standard deviation at time t, R_t is the return during the time period $t-1$ to t, and p_t and p_t-1 are the logarithm of the asset prices (S) at time t and $t-1$, respectively. If daily open and close prices are available, we can estimate the absolute daily return as

$$|R_t| = \left|\log\left(\frac{S_{t,\,close}}{S_{t,\,open}}\right)\right| = |p_{t,\,close} - p_{t,\,open}|. \tag{2}$$

By definition, the above absolute return measure ignores intraday information embedded in the data.

Realized volatility

The realized volatility measure, also known as the cumulative intraday squared return measure of volatility, was introduced by Anderson and Bollerslev (1998a) for high-frequency data. This measure is estimated as follows:

$$\sigma_t = \sqrt{RV_t} = \left(\sum_{n=1}^{N} r_{t,n}^2\right)^{1/2} = \left(\sum_{n=1}^{N} (p_{t,n} - p_{t,n-1})^2\right)^{1/2}, \tag{3}$$

where σ_t is the volatility measure, $RV_t = \sum_{n=1}^{N} r_{t,n}^2$ is the realized variance, or the cumulative intraday squared return, $p_{t,n}$ is the logarithm of the nth observation during the period between $t - 1$ and t, and N is the number of observations during that period of time.

The realized volatility estimator is justified based on the quadratic variation theorem which implies that when asset prices are observed without errors, this estimator provides consistent estimates of integrated volatility of the underlying price process (see Mancino & Sanfelici, 2008). Barndorff-Nielsen and Shepard (2001) show that realized volatility is less subject to measurement error and provides estimates that are less noisy.

Integrated volatility via bipower variation

As an alternative estimator of integrated variance, the realized bipower variation was suggested by Barndoff-Nielsen and Shephard (2004). The realized bipower variation is defined as

$$BV_t = \frac{\pi}{2}\sum_{n=1}^{N} |r_{t,n}||r_{t,n-1}|, \;\; t = 1, \ldots, T. \tag{4}$$

where, r_t, n is the nth intra-period return during the period between $t - 1$ and t. Therefore, the integrated volatility (realized bipower volatility) can be estimated as

$$\sigma_t = \sqrt{BV_t} \tag{5}$$

Barndoff-Nielsen and Shephard (2004) shows that the realized bipower variation converges to the same probability limit as the realized variance when prices follow a stochastic volatility process and the limit does not change with added rare jumps. In contrast, the limit for realized variance changes when there are jumps. Therefore, the differences between the limit of the realized bipower variation and the realized variance are useful in indentifying the impact of jumps on the quadratic variation, or integrated variance (also see Anderson et al., 2009).

Integrated volatility via Fourier transformation (IVFT)

Malliavin and Mancino (2002) recently propose the Fourier estimator of integrated volatility wherein the instantaneous volatility is reconstructed as a series expansion with coefficients gathered from Fourier coefficients of the price variation (also see Reno & Barucci, 2002).

This measure can be described as follows. Let $S(t)$, $(0[t[T)$ be a time series of an asset prices and $p(t) = \log(S(t))$ which is the series of the logarithm prices. Without loss of generality, the series $p(t)$ can be described by the following stochastic process:

$$dp(t) = \sigma_t dW(t), \tag{6}$$

where, σ_t is the instantaneous volatility at time t, a time dependent random function, and $W(t)$ is a standard Brownian motion.

Given a time series of N observations $(t_i, p(t_i))$, $i = 1, \ldots, N$, data is compacted into the interval $[0, 2\pi]$. This interval is the normalization for the time window. Inside this time window, there is a partition determined by the number of observations during that day. Thus, an estimator of the integrated volatility can be obtained as

$$\begin{aligned} \int_0^{2\pi} \sigma^2(s)\, ds &= 2\pi a_0(\sigma^2) \\ \sigma_t &= (2\pi a_0(\sigma^2))^{1/2}, \end{aligned} \tag{7}$$

where, $a_0(\sigma^2)$ is the first Fourier coefficient,

$$a_0(\sigma^2) = \lim_{n\to\infty} \frac{\pi}{n + 1 - n_0} \sum_{s=n_0}^{n} \frac{1}{2} [a_s^2(dp) + b_s^2(dp)], \tag{8}$$

and, $a_k(dp)$, and $b_k(dp)$ are the Fourier coefficients of dp:

$$\begin{aligned} a_k(dp) &= \frac{p(2\pi) - p(0)}{\pi} - \sum_{i-1}^{N-1} p(t_i) \frac{1}{\pi} [\cos(kt_i) - \cos(kt_{i+1})], \\ b_k(dp) &= \frac{p(2\pi) - p(0)}{\pi} - \sum_{i-1}^{N-1} p(t_i) \frac{1}{\pi} [\sin(kt_{i+1}) - \sin(kt_i)]. \end{aligned} \tag{9}$$

Equation (7) gives the expression of $a_0(\sigma^2)$ which is a limit of the summation of $a_k(dp)$ and $b_k(dp)$. Equations (8) computes the Fourier coefficients $a_k(dp)$ and $b_k(dp)$. The integrated volatility can now be estimated without integration.

The properties of different volatility measures have been examined in the literature in the context of foreign exchange markets and simulated data. Anderson and Bollerslev (1998a) use currency data to show that the realized volatility is more efficient than absolute returns. Reno and Barucci (2002) compare IVFT with realized volatility in a Monte Carlo study to generate latent instantaneous volatility, and they show that the IVFT method is superior to the realized volatility approach. Nielsen and Frederiksen (2008) compare different measures of integrated volatility with a Monte Carlo simulated dataset that uses parameters estimated by Andersen and Bollerslev (1998b). They find the IVFT method to be superior compared with the realized volatility and wavelet transformation. More strikingly, even after using bias correction methods designed specifically to handle market microstructure effects, the IVFT method is shown to have a superior forecasting performance while having only a slightly higher bias.

Normality Tests

Andersen, Bollerslev, and Dobrev (2007) argue that traditional normality tests are biased in testing whether or not the return distributions are i.i.d. Gaussian. The authors develop a new scheme of normality tests based on the first four moments within the context of continuous time jump diffusion framework, summarized as follows:

Define the sample and population moments by

$$\hat{M}_T^{(k)} = \frac{1}{T} \sum_{t=1}^{T} R_t^k, \quad \text{and} \quad M_t^{(k)} = E[\hat{M}_T^{(k)}]. \tag{10}$$

Then, for a standard normal distribution, we have $M_T^{(1)} = M_T^{(3)} = 0$, $M_T^{(2)} = 1$, and $M_T^{(4)} = 3$. The individual moments normality tests are

$$T[\hat{M}_T^{(1)}]^2 \cong \chi^2(1), \frac{T}{2}[\hat{M}_T^{(2)} - 1]^2 \cong \chi^2(1),$$
$$\frac{T}{15}[\hat{M}_T^{(3)}]^2 \cong \chi^2(1), \frac{T}{96}[\hat{M}_T^{(4)} - 3]^2 \cong \chi^2(1). \tag{11}$$

The joint normality test for the first four moments is

$$T(\hat{M}_T^{(1)}\hat{M}_T^{(2)} - 1\hat{M}_T^{(3)}\hat{M}_T^{(4)} - 3)\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{bmatrix}^{-1}\begin{pmatrix} \hat{M}_T^{(1)} \\ \hat{M}_T^{(2)} - 1 \\ \hat{M}_T^{(3)} \\ \hat{M}_T^{(4)} - 3 \end{pmatrix} \cong \chi^2(4) \tag{12}$$

Here, T is the number of observations (returns) in the sample. Andersen et al. (2007) prove that this set of normality test is more powerful than conventional normality tests. We apply the four moment's individual normality tests and the joint test to the standardized returns.

Assessing Forecast Performance

As 'true' volatility can never be observed, it is necessary to establish a proxy for actual volatility in order to determine the forecast performance of volatility estimates. In this study the predictive ability of the GARCH model is examined by using the daily absolute return, realized volatility, realized bipower volatility, and IVFT measures. The use of the GARCH(1,1) forecasting model has precedence in studies such as Baillie and Bollerslev (1992), Andersen et al. (2001), Reno and Barucci (2002), and Koopman, Jungbacker and Hol (2005).

Suppose the mean equation of returns is specified by an AR(p) model as

$$R_t = c + \sum_{i=1}^{p} a_i R_{t-i} + \varepsilon_t. \tag{13}$$

Here R_t is the return at day t, c and a_i are parameters to be estimated, and the error term, ε_t, is factorized as

$$\varepsilon_t = z_t h_t^{1/2}, \tag{14}$$

where z_t is an i.i.d. sequence with mean zero and variance one. Then the GARCH (p, q) model can be written as

$$h_t = \omega + \sum_{i=1}^{F} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_i h_{t-j}. \tag{15}$$

A GARCH(1, 1) model is estimated on the daily return series. Here, p and q are the numbers of lags in the error term and GARCH volatility, and ω, α_i and β_i are coefficients to be estimated. The following equation is used to forecast the one-day ahead volatility:

$$F_{1,t} = \sigma_1^2(\alpha_1 + \beta_1)(h_{t-1} - \sigma_1^2) \tag{16}$$

where $\sigma_1 = \frac{\omega}{1 - (\alpha_1 + \beta_1)}$ and $h_t - 1$ is the conditional variance from GARCH model of the previous day and $F_{1,t}$ is the one-day-ahead forecast.

The predictive ability of the GARCH(1, 1) model for different volatility measures are evaluated using the new family of loss functions suggested by Patton (2009). The family of loss functions has been empirically tested and shown to be superior to the heteroscedastic root mean squared error (HRMSE) and the logarithmic loss function (LL)[1] which are two commonly used loss functions used in the literature. Patton's (2009) family of loss functions has the following expression:

$$L(\hat{\sigma}^2, h; b) =$$

$$\begin{cases} \dfrac{1}{(b+1)(b+2)}(\hat{\sigma}^{2b+4} - h^{b+2}) - \dfrac{1}{b+1}h^{b+1}(\hat{\sigma}^2 - h), & \text{for } b\, v\, \{-1, -2\} \\ h - \hat{\sigma}^2 + \hat{\sigma}^2 \log\dfrac{\hat{\sigma}^2}{h}, & \text{for } b = -1, \\ \dfrac{\hat{\sigma}^2}{h} - \log\dfrac{\hat{\sigma}^2}{h} - 1, & \text{for } b = -2. \end{cases} \tag{17}$$

Here, $\hat{\sigma}^2$ is the forecasted variance, $F_{m,t}$, and h is volatility proxy, σ^2. To evaluate forecast performance, we first apply Equation (16) to obtain the one-day-ahead volatility. Subsequently, the four integrated volatility estimators are used as "real" volatility proxies, σ^2, to see how far forecasted values deviate from "true" values. The criteria for evaluating forecast performance are dependent on the relative values of the loss functions (see Andersen et al. 2001; Baillie & Bollerslev, 1992; Koopman et al., 2005; Reno & Barucci, 2002).

The criteria for evaluating forecast performance are dependent on the relative values of the loss functions. Similar to Patton (2009), we report cases where $b = -5, -2, -1, 0$, and 1. Smaller absolute values of the loss functions indicate better forecasting performance. Note that the simple MSE loss function is obtained when $b = 0$, and the quasi-likelihood (QLIKE) loss function is obtained when $b = -2$. Patton (2009) argues that the QLIKE loss function is the most powerful one within this family of loss functions.

[1] $HRMSE = E[(1 - \frac{F_{m,t}}{\sigma_t^2})^2]^{1/2}, \quad LL = E[\log(\frac{F_{m,t}}{\sigma_t^2})]$.

DATA CHARACTERISTICS AND DATA CLEANING

Our primary data set includes intraday, tick-by-tick, futures prices for gold, silver, and copper for four time intervals: 1-, 2-, 5-, and 15-minute intervals. The data is obtained from the *Futures Industry Institute*. All three futures contracts are traded on NYMEX (New York Mercantile Exchange) and priced in US dollars. The sample period is from January 1999 to December 2008.

Data Characteristics

The tick-by-tick raw futures data specify the time, to the nearest second, and the exact price of the futures transaction. The contract with the nearest expiration is used to construct a continuous futures price series. Specifically, we consider the daily tick volume for the front and first back-month contracts and rollover to the next contract when the daily tick volume of the back-month contract exceeds the daily tick volume of the current front month contract. This procedure results in linking together multiple contracts using only the most active portion of each contract. For each time interval, the last recorded price for the nearby futures contracts are employed to obtain the 1-, 2-, 5-, and 15-minute prices. Based on the number of trading days in the sample period, the total numbers of observations obtained for each metal are about 702,000, 351,000, 140,500, and 35,100 for the 1-, 2-, 5-, and 15-minute frequencies, respectively.

Table I reports summary statistics for the daily raw return series for gold, silver, and copper. Gold has a daily mean return of 0.04% and standard deviation 1.16%. Both silver and copper have approximately the same mean return (0.03%), but silver has a higher standard deviation (1.89%). The distributions for silver and copper are negatively skewed. Kurtosis, which is a measure of the peak of the distribution relative to a normal distribution, is 9.1 for gold, 11.8 for silver,

TABLE I
Summary Statistics of Daily Raw Returns for Gold, Silver and Copper for the Period 1999–2008

	Gold Futures	*Silver Futures*	*Copper Futures*
Mean	0.04%	0.03%	0.03%
Median	0.04%	0.11%	0.04%
Standard deviation	1.16%	1.89%	1.82%
Skewness	0.19	−1.00	−0.34
Kurtosis	9.11	11.76	8.24
Jarque–Bera	3921.12	8450.26	2914.69
Probability	0.00	0.00	0.00
No. of observations	2512	2510	2506

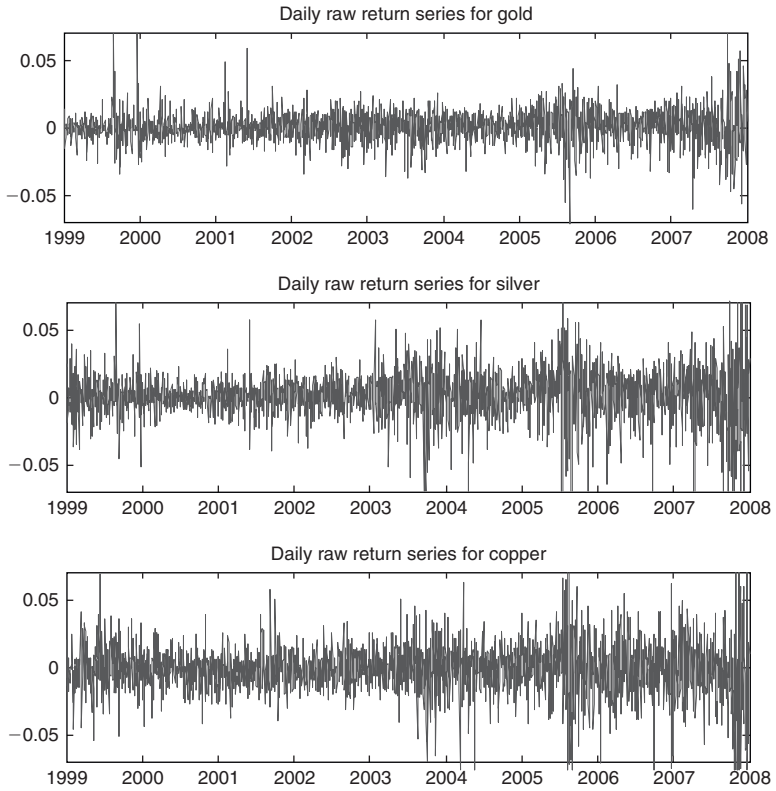

FIGURE 1
Daily raw returns for gold, silver, and copper for the period 1999–2008.

and 8.24 for copper. The Jarque–Bera tests reject the null hypothesis that the sample raw returns for gold, silver, and copper come from a normal distribution.

Figure 1 shows the daily raw return series for gold, silver, and copper for the period 1999–2008. All three return series exhibit a great deal of volatility, particularly after 2005, while showing a tendency for a constant mean. Across all three metals, the largest shocks to the return process took place towards the end of 2008 coinciding with the turmoil in financial markets. Also, in general, the volatility in silver and copper appear to be more pronounced than gold.

Figure 2 provides histograms of the raw return distribution for the three metals. All three return distributions show that they are concentrated around the mean, but exhibit fatter tails and sharper peaks at the mean of the returns in comparison to the standard normal distribution. The negative skewness coefficient (i.e., a large probability of outcomes above the expected value) for silver and copper are also visible from the graph. Once again, there is strong visual evidence to suggest that the returns for gold, silver, and copper futures do not appear to be normally distributed.

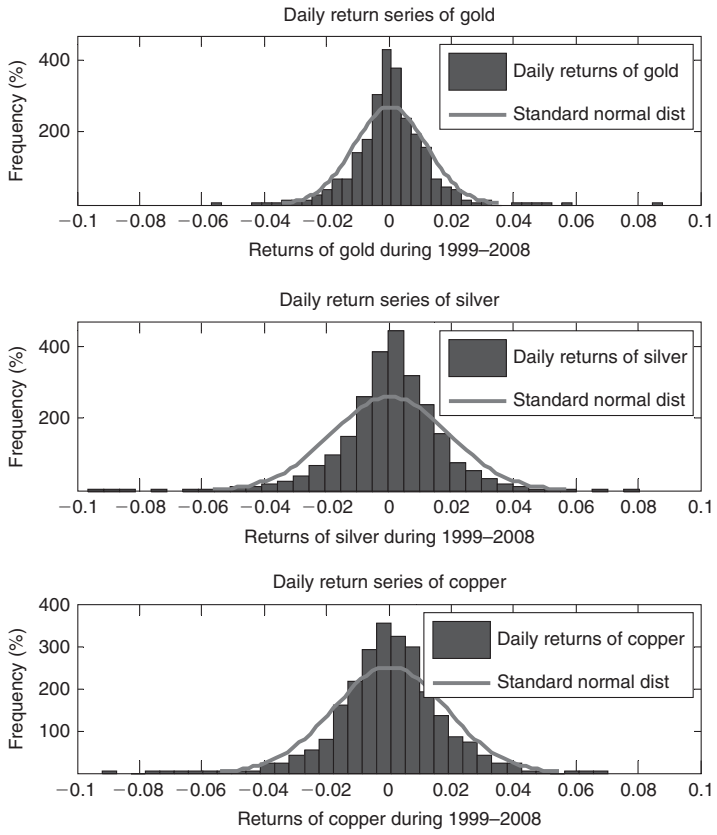

FIGURE 2
Histograms for gold, silver, and copper futures daily raw return series.

Figure 3 presents correlograms of the daily raw return series. Daily returns have very little correlation which means that returns are almost impossible to predict from their own past history (up to 20 lags), and this is evidence that the unconditional mean is roughly constant.

Data Cleaning

Prior to conducting empirical analysis, a two-step procedure is employed to clean the raw data. First, in order to minimize bias in the results we eliminate days that have infrequent trades. During each normal trading day, the high-frequency data at five-minute interval should have about 72 price observations for gold, 70 observations for silver, and 68 observations for copper. We remove those days from the sample where the number of five-minute price observations is less than 50% of the number of price observations found in a normal trading day. For instance, there are 33 days during which the gold contract had fewer than 36 five-minute price observations.

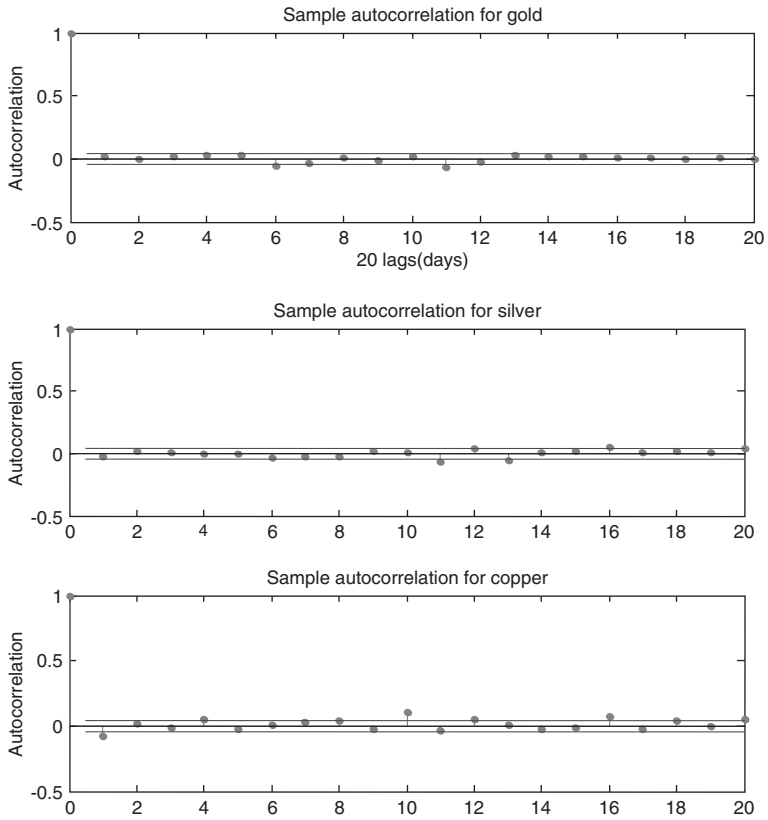

FIGURE 3
Return correlograms for daily raw returns of gold, silver, and copper.

Similarly, the corresponding number of days with insufficient data for silver and copper were 44 and 39, respectively. This resulted in a sample size of 2,480 days for gold, 2,467 days for silver, and 2,116 days for copper.[2]

Second, similar to Hansen and Lunde (2006) we eliminate days when the opening and closing transaction prices are the same. Zero daily returns are particularly a problem when computing the loss functions associated with daily absolute returns. After removing those days from the sample, we end up with data for 2,430, 2,375, and 2,061 days for gold, silver, and copper.

EMPIRICAL RESULTS

Comparing Volatility Measures

As a first step, we estimate integrated volatility using the four volatility measures (see Equations (1), (3), (5), and (7)) and volatility estimated using the

[2]We tried other filters that ranged from 30 to 60%, but the results were qualitatively similar. We thought that the 50% requirement was reasonable.

TABLE II
Annual Volatility Estimates of Gold, Silver and Copper for the Period 1999–2008

	Time Frequency			
Volatility Measure	*1 minute*	*2 minutes*	*5 minutes*	*15 minutes*
Panel A: Gold futures				
GARCH daily (%)	13.80	13.80	13.80	13.80
Daily absolute returns (%)	9.97	9.97	9.97	9.97
Bipower variation (%)	13.33	12.67	12.03	10.96
Realized volatility (%)	13.92	13.27	12.71	11.78
IVFT (%)	16.74	14.52	13.77	11.98
Panel B: Silver futures				
GARCH daily (%)	22.60	22.60	22.60	22.60
Daily absolute returns (%)	16.51	16.51	16.51	16.51
Bipower variation (%)	24.34	22.49	20.44	17.59
Realized volatility (%)	25.69	23.86	21.94	19.43
IVFT (%)	32.93	27.39	23.39	19.81
Panel C: Copper futures				
GARCH daily (%)	19.31	19.31	19.31	19.31
Daily absolute return (%)s	14.35	14.35	14.35	14.35
Bipower Variation	19.84	18.63	17.57	15.92
Realized volatility	20.59	19.44	18.42	17.01
IVFT	26.44	22.14	20.60	18.47

GARCH(1, 1) model. It is important to note that the GARCH model should not be viewed as a benchmark, but just as another competing model that is used for comparative purposes.

Table II summarizes the average annualized volatility, $\sqrt{252} \times \sigma_{Daily}$, of gold, silver, and copper for the *cleaned* data using different volatility measures and different sampling frequencies. In the case of gold (see Panel A), for the one-minute interval, the estimated annual volatility is 16.74% using IVFT, 13.92% using realized volatility measure, 13.33% using realized bipower volatility, and 9.97% using daily absolute returns. The results indicate that the realized volatility measure is closest to the GARCH volatility estimate of 13.80%. Furthermore, the magnitudes of IVFT, realized volatility, and realized bipower volatility decrease as the intraday time frequency decreases from 1-minute to 15-minute interval.[3]

The corresponding results for silver and copper are shown in Panels B and C, respectively. These results are qualitatively similar to those of gold. The absolute return measure gives the lowest volatility estimates. Furthermore, across all different frequencies, IVFT is always greater than realized volatility,

[3]As pointed out earlier, the absolute return measure does not depend on the price movements during the period. Therefore, the same value is obtained for all time frequencies.

and this in turn is greater than realized bipower volatility. This latter finding has been highlighted by both Andersen et al. (2009) and Barndoff-Nielsen and Shephard (2004). However, to the best of our knowledge, we are the first to document that the magnitude of the IVFT estimate is the greatest among the four volatility measures. This is perhaps because the IVFT measure is able to reflect more information about underlying price movements relative to other measures. Although it has been theoretically proven that both IVFT and bipower volatility measures are consistent estimators of integrated volatility, it is interesting to note that the empirical estimates of the magnitude of the two volatility measures are different from each other.

A couple of other interesting observations emerge. For the five-minute interval, the IVFT measure is the closest one to the GARCH (1,1) estimate. However, for the two-minute interval, the realized volatility estimates are very close to the GARCH (1, 1) results. We highlight the two- and five-minute frequencies, as these two frequencies are most commonly used in the literature to investigate intra-day volatility. An implication from these findings is that the IVFT measure or the realized volatility measure adds very little value relative to the use of a simple GARCH (1, 1) model on daily data.

Daily Return Distributions

In this section, the distributional properties of daily speculative returns are investigated by using the Anderson et al. (2007) and Anderson et al. (2009) framework. For the sake of brevity, only the results pertaining to the five-minute interval is presented.

Table III presents the first four moments (m_1 through m_4) for the different standardized return series associated with the three markets, along with the corresponding P-values for testing $m_1 = 0$, $m_2 = 1$, $m_3 = 0$, and $m_4 = 3$, respectively; except for the realized volatility standardized return series, for which the test for the fourth moment is based on the finite sample correction, $m_4 = 3\frac{M}{M+2}$. Here, M is the number of observations during a regular trading day. The results for gold, silver, and copper futures are reported in Panels A, B, and C in Table III. The QQ-plots for the corresponding distributions are reported in Figure 4.

Standardized return distributions

The first five rows in Table III report normality test results of returns standardized by unconditional volatility, GARCH(1, 1), and the three integrated volatility measures—realized volatility, realized bipower volatility, and IVFT—using five-minute interval. It is clearly evident from the table that the different standardized return series are not normally distributed.

TABLE III

Normality Test Results for Standardized Returns

	m_1	p_1	m_2	p_2	m_3	p_3	m_4	p_4	p_{joint}	$p_{joint\text{-}dm}$
Panel A: Gold										
$R_t/\sqrt{Var(R_t)}$	0.008245	0.681379	0.999665	0.990581	0.081351	0.295551	11.29416	0	0	0
$R_t/\sqrt{GARCH(1,1)}$	0.026025	0.19506	0.999409	0.983407	0.74405	1.12E−21	12.19727	0	0	0
$R_t/\sqrt{IVFT_t}$	0.033612	0.094157	0.702335	1.05E−25	0.097625	0.209374	1.376437	1.56E−16	5.12E−24	1.2E−23
$R_t/\sqrt{BV_t}$	0.043706	0.029514	0.925769	0.00895	0.172006	0.026988	2.301483	0.000385	0.000829	0.004115
$R_t/\sqrt{RV_t}$	0.041057	0.040894	0.806709	1E−11	0.139495	0.072868	1.68832	2.61E−11	8.5E−11	3.06E−10
$\tilde{R}_t/\sqrt{CV_t}$	0.056818	0.004662	0.744755	2.52E−19	0.150154	0.053518	1.524087	6.31E−14	1.87E−18	3.34E−17
$\hat{R}_t/\sqrt{CVS_t}$	0.057415	0.004247	0.729392	1.59E−21	0.148672	0.055921	1.495946	2.1E−14	7.03E−21	1.3E−19
$\hat{R}^*_k/\sqrt{E(CVS_t)}$	0.034544	0.085384	0.813092	4.65E−11	0.077759	0.317387	1.900882	2.32E−08	1.84E−09	6.56E−09
$\hat{R}^*_{5k,5}/\sqrt{5E(CVS_t)}$	0.077242	0.085384	0.857953	0.025289	0.275738	0.112832	2.421313	0.188384	0.046362	0.122006
Panel B: Silver										
$R_t/\sqrt{Var(R_t)}$	0.009597	0.633604	0.999687	0.991222	−0.76212	1.46E−22	10.17161	2.2E−289	0	0
$R_t/\sqrt{GARCH(1,1)}$	0.017240	0.3919326	1.000178	0.995001	−0.409277	1.542E−07	9.16865	1.41E−214	0	0
$R_t/\sqrt{IVFT_t}$	0.051616	0.010356	0.67485	3.34E−30	0.093081	0.232592	1.471384	9.26E−15	6.82E−33	9.26E−32
$R_t/\sqrt{BV_t}$	0.061098	0.002408	0.892035	0.00015	0.132526	0.089212	2.54773	0.021865	9.53E−06	0.000475
$R_t/\sqrt{RV_t}$	0.056171	0.005272	0.740786	8.71E−20	0.104399	0.180615	1.611917	1.97E−12	1.02E−19	3.05E−18
$\tilde{R}_t/\sqrt{CV_t}$	0.083243	3.56E−05	0.691512	2.36E−27	0.172176	0.02724	1.452469	4.33E−15	1.19E−30	8.89E−28
$\hat{R}_t/\sqrt{CVS_t}$	0.090475	7E–06	0.680571	3.3E−29	0.148079	0.05756	1.429912	1.73E−15	1.47E−34	7.43E−31
$\hat{R}^*_k/\sqrt{E(CVS_t)}$	0.041897	0.037438	0.780056	1.12E−14	0.070432	0.36639	1.793771	9.67E−10	1.2E−13	8.97E−13
$\hat{R}^*_{5k,5}/\sqrt{5E(CVS_t)}$	0.091305	0.042631	0.800641	0.001748	0.207365	0.234513	1.820289	0.007509	0.006463	0.030643

(Continued)

TABLE III (*Continued*)

	m_1	p_1	m_2	p_2	m_3	p_3	m_4	p_4	p_{joint}	$p_{joint\text{-}dm}$
Panel C: Copper										
$R_t/\sqrt{Var(R_t)}$	0.079138	0.000272	1.00579	0.85061	0.523112	5.2E−10	8.245035	6.9E−134	0	0
$R_t/\sqrt{GARCH(1,1)}$	0.079815	0.000242	0.997389	0.932324	0.459877	4.74E−08	5.845674	1.08E−40	1.8E−161	5.2E−151
$R_t/\sqrt{IVFT_t}$	0.080629	0.000208	0.693204	1.88E−23	0.195206	0.020422	1.558982	1.33E−11	9.46E−27	1.15E−24
$R_t/\sqrt{BV_t}$	0.101976	2.72E−06	0.940306	0.052177	0.352907	2.77E−05	2.628368	0.081027	2.43E−05	0.140342
$R_t/\sqrt{RV_t}$	0.093095	1.85E−05	0.834552	7.39E−08	0.263809	0.001729	2.011543	3.47E−06	1.27E−09	1.55E−06
$\tilde{R}_t/\sqrt{CV_t}$	0.09744	7.39E−06	0.766566	3.13E−14	0.243792	0.003785	1.781561	1.06E−08	8.38E−17	1.83E−13
$\hat{R}^*_k/\sqrt{(CVS_t)}$	0.097409	7.44E−06	0.74428	8.96E−17	0.218519	0.009448	1.673439	4.72E−10	1.15E−19	3.21E−16
$\hat{R}^*_k/\sqrt{E(CVS_t)}$	0.078491	0.000305	0.878134	7.37E−05	0.232824	0.005687	2.151055	6.73E−05	4.95E−06	0.000822
$\hat{R}^*_{5k,5}/\sqrt{5E(CVS_t)}$	0.173917	0.000348	0.86535	0.050204	0.514136	0.006329	2.236902	0.109194	0.002241	0.189111

This table reports the first four moments ($m_1 - m_4$) for the different standardized return series for the three markets using 5 minute interval, along with corresponding p-values for testing $m_1 = 0$, $m_2 = 1$, $m_3 = 0$, and $m_4 = 3$, respectively, except for the realized volatility standardized return series, for which the test for the fourth moment is based on the finite sample correction, $m_4 = 3\frac{M}{M+2}$. Here, M is the number of observations during a regular day. The column labeled p_{joint} gives the p-value for testing the four moment conditions jointly, while $p_{joint\text{-}dm}$ refers to the same test involving the (unconditionally) demeaned return series. The raw daily returns are denoted by R_t, while $\tilde{R}_t$ and $\hat{R}_t$ refer to the daily jump-adjusted returns based on the simple and sequential procedures, respectively. The daily realized volatility and the corresponding continuous component based on the simple and sequential jump-adjusted procedures are denoted by RV_t, CV_t, and CVS_t, respectively. $Var(R_t)$ refers to the unconditional variance of the returns; $GARCH(1, 1)$ is the conditional variance estimated by a GARCH(1, 1) model; $IVFT_t$, and BV_t represent the integrated variance using the Fourier transmission approach and the realized bipower volatility with 5-minute frequency intra-day data, respectively. $\hat{R}^*_k$ refers to the financial-time return series constructing from the sequential jump-adjusted intraday returns spanning $E(CVS)$ time-units. Finally, $\hat{R}^*_{5k,5} = \hat{R}^*_{5k} + \hat{R}^*_{5k-1} + \hat{R}^*_{5k-2} + \hat{R}^*_{5k-3} + \hat{R}^*_{5k-4}$ defines the financial-time return series spanning $5E(CVS_t)$ time units.

Panel A: Gold futures

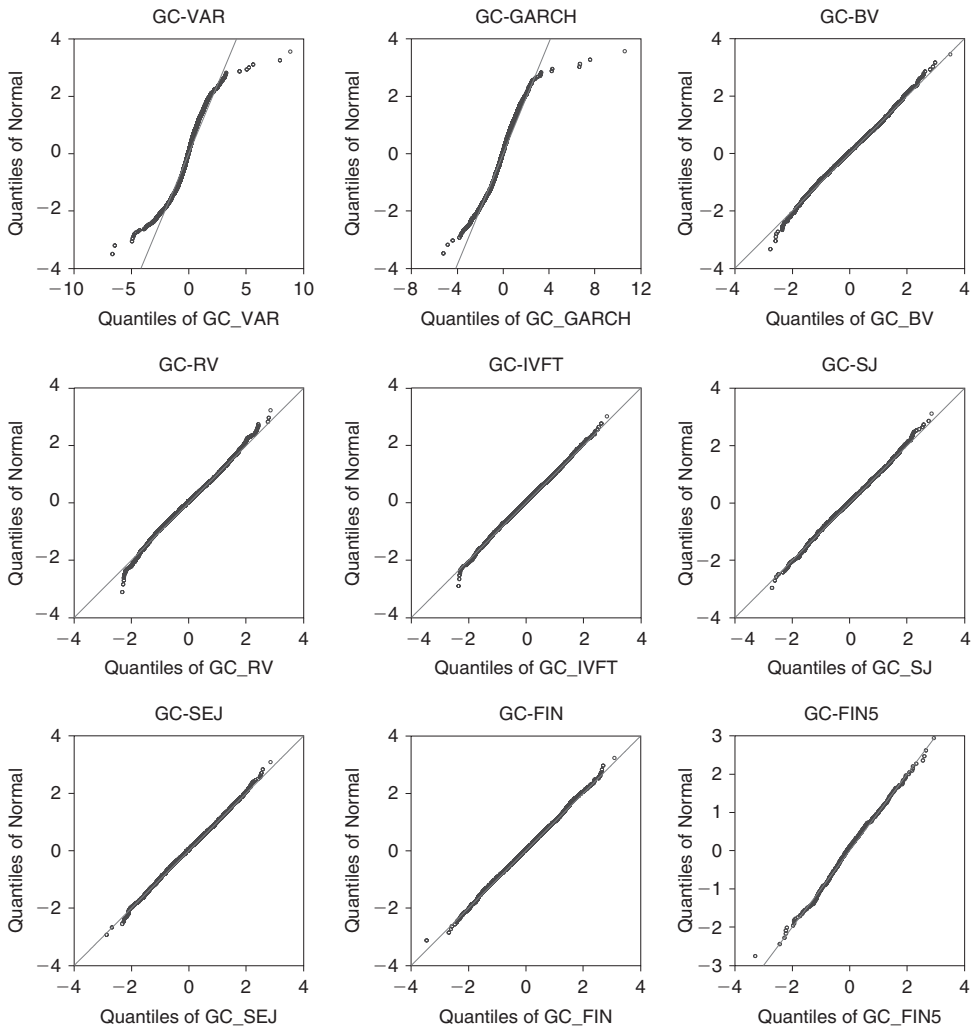

FIGURE 4. *(Continued)*

The unconditional returns and the GARCH(1, 1) standardized returns remain significantly leptokurtic (very high fourth moment) for all three markets. For instance, examining the unconditional returns in the gold market (see Panel A), the *P*-values of the first three-moment individual normality tests are 0.68, 0.99, and 0.29. However, the *P*-value for the fourth moment normality test is zero. In comparison, the normality of the GARCH (1, 1) standardized gold futures returns cannot be rejected only for the first two moments. Notably, the *p*-values of the first four-moment joint normality test is zero for both

Panel B: Silver futures

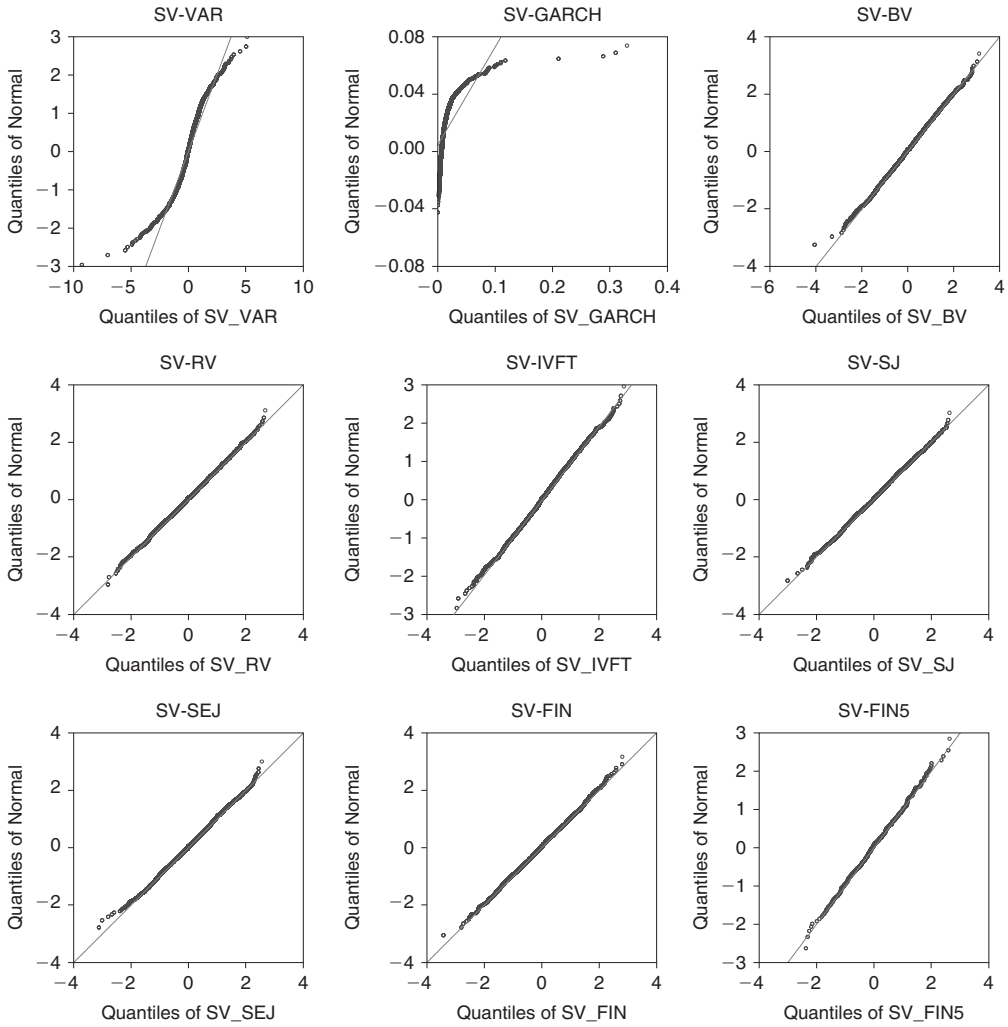

FIGURE 4. *(Continued)*

unconditional returns and GARCH (1, 1) standardized returns. One can arrive at a similar conclusion for silver and copper returns (see Panels B and C).

The normality of returns standardized by realized volatility, the realized bipower volatility and the IVFT is rejected by the joint test at the 1% significant level. The realized bipower volatility does a better job than the other two measures for all three markets. The *p*-values of the joint test for the bipower volatility standardized returns are significantly higher than the other two measures. The IVFT measure does a better job in normalizing the third moment for all three markets. The normality of IVFT standardized returns cannot be rejected by the first and third moment individual tests at the 1% significant level for all

Panel C: Copper futures

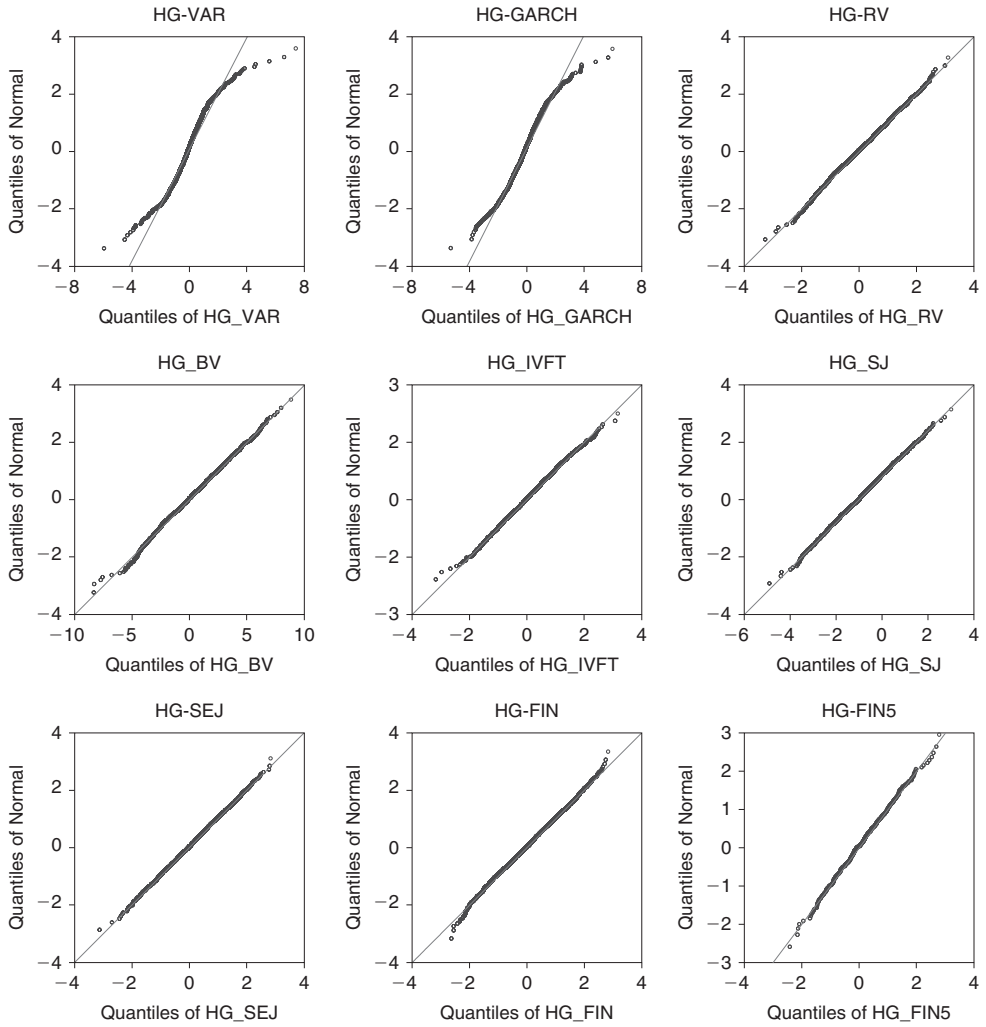

FIGURE 4
QQ-plots of standardized returns of gold, silver and copper. The figure reports the QQ-plots of standardized returns. GC, SV, and HG represent gold, silver and copper futures, respectively. **-VAR, **-GARCH, **-RV, **-BV, **-IVFT, **-SJ, **-SEJ, **-FIN, and **-FIN5 refer to the unconditional returns, returns standardized by GARCH(1, 1), realized volatility, realize bipower volatility, IVFT, the continuous component of realized volatility based on the simple and sequential jump-adjusted procedures, financial-time returns spanning $E(CVSV)$ time-units standardized by $E(CVSV)$, and financial-time return series spanning $5E(CVSV_t)$ time units standardized by $5E(CVSV_t)$.

three markets. Interestingly, the IVFT measure seems to over-adjust the fourth moment; in other words, IVFT-standardized daily returns show the lowest fourth moments across all three markets (1.376, 1.471, and 1.559 for gold, silver, and copper, respectively).

Standardized jump-adjusted returns

If there are no jumps and volatility leverage effects in the data, daily returns should be approximately normally distributed when standardized by integrated volatilities. However, as indicated in the previous section, all the standardized returns are not normal, which implies that there might be some jumps in the intraday returns.

In an attempt to improve the normality of the return series, we apply simple jump adjustment and sequential jump adjustment procedures proposed by Andersen et al. (2009) to the five-minute intraday returns for the three markets. The basic idea behind the jump detection procedure is that the realized bipower volatility measure is a consistent estimator of the integrated volatility even in the presence of jumps, but the realized volatility is a consistent estimator of the integrated volatility only in the absence of jumps. Therefore, the difference between those two measures identifies the jump component in realized volatility.

After identifying jumps, we remove the jump components from both daily returns and daily realized volatilities and arrive at daily jump-adjusted returns ($\tilde{R}_t$ and $\hat{R}_t$) and continuous components of the realized volatility (CV_t and CVS_t). Subsequently, we standardize the daily jump-adjusted returns by the corresponding continuous components of the daily realized volatilities. The 6th and 7th rows in Table III report the normality test results of the standardized daily jump-adjusted returns based on the simple and sequential jump identification procedures ($\tilde{R}_t/\sqrt{CV_t}$ and $\hat{R}_t/\sqrt{CVS_t}$).

The results indicate that even after adjusting for jumps, we are unable to obtain normality to the standardized returns. Andersen et al. (2009) also reach a similar conclusion and attribute this finding to the self-standardizing nature of jumps. The authors indicate that, "a large jump tends to inflate the absolute value of both the return and the realized volatility of standardized returns, so the impact is muted." Therefore, we extend the analysis to evaluate the properties of jump-adjusted standardized returns sampled in *financial-time* in the next section.

Standardized jump-adjusted financial-time returns

To sample in financial-time, or equal increments of integrated volatility, we include intraday returns until the cumulative squared returns exceeds the average continuous component of daily realized volatility in calendar time for the sequential jump-adjusted returns. That is, we only include the non-jump returns identified by the sequential jump detection procedure. For the purpose of comparing our results derived from financial-time with previous results, we calibrate the number of financial days to equal the number of calendar days in the sample. We consider one financial day if the cumulative squared returns

exceed the average continuous components of the daily realized volatility, or if the difference between the two values is less than a small value. We sample repeatedly by adjusting the magnitude of the small value until the number of financial days equals the number of calendar days.

The last two rows in Table III report the normality tests of the financial-time return series constructed from: (a) sequential jump-adjusted intraday returns spanning $E(CVS_t)$, which is the average continuous component of the sequential jump-adjusted realized volatility; and (b) $5E(CVS_t)$ time units. Evidence from all three markets indicates that the standardized daily financial-time returns are still not normal based on the joint test of the four moments. However, normality in the return distribution is restored for all three markets when considering the demeaned financial-time returns series spanning $5E(CVS_t)$ time units. In this case, normality cannot be rejected at the 1% level.

The QQ-plots in Figure 4 (Panels A, B, and C) show that in comparison to unconditional returns and GARCH (1, 1) standardized return distributions, the remaining standardized distributions are much closer to the reference Gaussian distribution. Notably, tails of the QQ-plots show significant improvement and display only slight deviations from the straight 45° line.

Comparing our results with Andersen et al.'s (2009) evidence on stock returns, we find that the normality for metal futures returns is somewhat more difficult to establish. Andersen et al. (2009) succeed in restoring normality for 24 out of the 30 stocks in the *Dow Jones Industrial Average* at the 1% significance level using financial time returns. However, in our study we reject normality of the standardized financial day returns at the 1% level of significance. To be sure we tried different sampling frequencies, but the results were qualitatively similar. We conjecture that this difficulty in establishing normality for metal futures returns might be attributable to some additional microstructure noise that may be prevalent in these markets.

In conclusion, the unconditional returns of futures prices are not normally distributed. When we standardize returns with IVFT and realized volatility measures, the corresponding return distributions bear closer resemblance to a normal distribution. By applying the newly developed financial-time sampling technique proposed by Andersen et al. (2009), we succeed in converting the return series into i.i.d. Gaussian for all three markets. The added difficulties of restoring normality for futures returns are a topic that may require further investigation.

Forecast Performance of GARCH(1, 1) Model Using Different Volatility Measures as Volatility Proxies

In the concluding step of the analysis, the predictive performance of the GARCH(1, 1) forecasting model is evaluated using the four different volatility

measures. As suggested by Patton (2009), forecasting accuracy is compared by using the QLIKE loss function, although we report the values for all five different loss functions.

For the purpose of comparing our results with evidence from prior literature, we first value from traditional loss functions such as HRMSE and LL.[4] The results indicate that both HRMSE and LL functions strongly document the GARCH model using the IVFT estimator which produces volatility forecasts that are more accurate when compared with other measures. The evidence provides support for Nielsen and Frederiksen (2008) and Reno and Barucci (2002) who posit that the IVFT measure is a better measure of integrated volatility than realized volatility. Also consistent with Anderson and Bollerslev (1998a) and others, the results using HRMSE and LL loss functions indicate that forecasting performance of the GARCH(1, 1) model improves as the time frequency increases from 15- to 1-minute intervals.

Table IV reports the values associated with the new family of loss functions using the 1-, 2-, 5-, and 15-minute time intervals for gold, silver, and copper. Several important insights are documented. First, forecasting performance of the GARCH (1, 1) model depends on both the volatility proxy and intraday time frequency that is used in the analysis. For the 5- and 15-minute frequencies, the GARCH (1, 1) model has the smallest forecast errors (i.e., least loss function values) when the IVFT measure. This is true for all three metals. However, if we increase the time frequency to two or one $$minute, the lowest forecasting error is obtained when the realized volatility proxy is used. The silver market provides the only exception to this finding; here the two-minute interval using IVFT produces the smallest forecasting error.

Second, unlike the results from HRMSE and LL, increasing the time frequency in general does not necessarily improve the predictive performance of the GARCH (1, 1) model. For instance, when IVFT is used, only the five-minute interval provides the lowest forecasting error. In the case of realized volatility, however, forecast performance improves with the intraday time frequency used to measure daily returns. It may be worth pointing out that these conclusions are broadly consistent with the results in Table II, which documents that the realized volatility estimate gets closer to the GARCH (1, 1) volatility estimate as the intraday frequency used to measure daily returns increases.

Third, and not surprisingly, the GARCH (1, 1) forecasting model yields the largest forecast error when using the daily absolute return measure. This is because by definition the daily absolute return measure does not integrate intraday information into the estimation procedure.

[4]These results are not reported but can be made available by the authors upon request.

TABLE IV

Forecast Performance of the GARCH (1, 1) Model

			Value of Loss Functions				
			b				
Time Frequency	*Market*	*Volatility Measure*	*−5*	*−2*	*−1*	*0*	*1*
1 Minute	Gold	*D–ABS*	3.39E+27	152328	0.314419	7.99E−05	1.84E−07
		IVFT	2.11E+15	473.3238	0.074112	2.56E−05	1.76E−08
		BV	3.71E+17	502.3164	0.037622	9.18E−06	5.89E−09
		RV	1.74E+17	429.6906	0.03828	1.06E−05	7.05E−09
	Silver	*D–ABS*	3.12E+24	72652.34	0.77424	0.000495	2.96E−06
		IVFT	1.01E+14	881.3926	0.368876	0.000268	4.24E−07
		BV	5.13E+16	428.1604	0.106149	7.81E−05	1.27E−07
		RV	1.37E+16	391.9233	0.119774	0.000113	2.68E−07
	Copper	*D–ABS*	1.16E+14	5725.963	17.11383	0.158213	0.003068
		IVFT	9.67E+13	491.7552	0.144392	7.8E−05	8.55E−08
		BV	3.95E+14	236.1186	0.04753	2.29E−05	2.41E−08
		RV	1.3E+14	202.5782	0.047681	2.56E−05	2.72E−08
2 Minutes	Gold	*D–ABS*	3.39E+27	152328	0.314419	7.99E−05	1.84E−07
		IVFT	1.02E+16	415.0369	0.047485	1.51E−05	1.06E−08
		BV	1.64E+18	742.5116	0.046472	1.21E−05	1.05E−08
		RV	5.28E+17	599.7234	0.044266	1.19E−05	8.88E−09
	Silver	*D–ABS*	3.12E+24	72652.34	0.77424	0.000495	2.96E−06
		IVFT	1.5E+14	539.111	0.16905	0.000107	1.59E−07
		BV	7.17E+16	484.5708	0.102449	6.72E−05	9.86E−08
		RV	2.16E+16	397.294	0.107412	8.53E−05	1.51E−07
	Copper	*D–ABS*	1.16E+14	5725.963	17.11383	0.158213	0.003068
		IVFT	1.18E+14	262.0463	0.065599	3.86E−05	5.56E−08
		BV	5.35E+14	280.5313	0.048792	2.47E−05	3.45E−08
		RV	2.56E+14	226.2115	0.047353	2.85E−05	4.48E−08
5 Minutes	Gold	*D–ABS*	3.39E+27	152328	0.314419	7.99E−05	1.84E−07
		IVFT	1.42E+16	508.5838	0.048723	1.87E−05	2.57E−08
		BV	3.11E+18	1261.425	0.060039	1.69E−05	2.16E−08
		RV	1.25E+18	976.6499	0.056502	1.86E−05	2.65E−08
	Silver	*D–ABS*	3.12E+24	72652.34	0.77424	0.000495	2.96E−06
		IVFT	5.2E+14	384.3816	0.113635	8.56E−05	1.39E−07
		BV	1.03E+17	815.1069	0.126349	6.36E−05	8.65E−08
		RV	8.47E+16	617.2162	0.125764	8.45E−05	1.35E−07
	Copper	*D–ABS*	1.16E+14	5725.963	17.11383	0.158213	0.003068
		IVFT	1.03E+14	183.4672	0.05021	4.14E−05	1.01E−07
		BV	2.75E+15	472.0762	0.063438	3.67E−05	8.61E−08
		RV	9.88E+14	346.381	0.058903	3.94E−05	9.49E−08
15 Minutes	Gold	*D–ABS*	3.39E+27	152328	0.314419	7.99E−05	1.84E−07
		IVFT	3.43E+18	1649.314	0.068689	1.48E−05	1.26E−08
		BV	2.84E+21	2651.288	0.077728	1.09E−05	6.88E−09
		RV	6.69E+18	1830.008	0.07072	1.42E−05	1.14E−08
	Silver	*D–ABS*	3.12E+24	72652.34	0.77424	0.000495	2.96E−06
		IVFT	3.8E+17	1270.958	0.169335	9.67E−05	1.98E−07
		BV	1.62E+27	8863.144	0.199	5.73E−05	5.25E−08
		RV	9.89E+17	1509.009	0.177228	9.78E−05	2.18E−07
	Copper	*D–ABS*	1.16E+14	5725.963	17.11383	0.158213	0.003068
		IVFT	2.85E+15	422.9513	0.067214	3.97E−05	8.72E−08
		BV	4.96E+17	1198.873	0.096562	2.74E−05	2.98E−08
		RV	1.27E+16	787.5803	0.084806	3.52E−05	5.93E−08

This table reports the values of the loss functions of the one day ahead forecasting of the GARCH (1,1) model using different volatility measures as volatility proxies. In the table, *D–ABS* refers to the daily absolute returns measure; *IVFT* and *BV* represent the integrated volatility using the Fourier transmission approach and the realized bipower volatility; *RV* refers to the realized volatility.

Finally, a comparison of different volatility forecasts indicates that copper has the lowest forecast error, followed by silver and gold. Although this finding cannot be explained within the context of the study, one might surmise that this result is being partly driven by the price discovery and volatility transmission process among strategically linked metals market (see Adrangi, Chatrath, & David, 2000).

CONCLUSIONS

A large amount of literature has been devoted to estimating and forecasting volatility for financial time series. In this field, the importance of high-frequency data has been stressed, in particular for evaluating the predictive abilities of volatility models. This study uses intraday futures prices for gold, silver, and copper for the period 1998–2009 to examine the return distribution properties and estimate four integrated measures of volatility—absolute returns, realized volatility, realized bipower volatility, and IVFT.

The study finds that the daily futures returns are generally not normal, even when standardized by various integrated volatility measures. However, restoration of normality is possible by applying the recently developed sequential jump detection scheme and financial-time sampling techniques. It is interesting to note that establishing a normal distribution for the metal futures market is found to be more difficult than that for the stock market, an issue that is left for future investigation.

The study further documents that, in general, IVFT and realized volatility are relatively efficient proxies of volatility in the metal futures market. The IVFT measure is always greater than realized volatility followed by realized bipower volatility for all different frequencies—an indication that the IVFT measure may reflect more information about underlying price movements than other volatility measures.

Finally, the forecast accuracy of the GARCH (1, 1) model is assessed by using a set of robust loss functions. The results indicate that for the two- and five-minute time frequencies, which are most commonly used in the literature, both realized volatility and IVFT measures provide low forecasting errors. Notably, an increase in the time frequency does not necessarily improve the forecasting performance of the GARCH (1, 1) model.

The research highlights the importance of integrated volatility measures and high-frequency sampling techniques in analyzing the distributional properties of futures returns. In particular, gauging the usefulness of volatility forecasts requires a refined articulation of volatility, as well as construction of integrated volatility measures that captures the concept in an empirically robust

fashion. Finally, an evaluation of competing forecasting models is partly dependent on the specific loss function that is used.

There are many possible ways to extend this study. For instance, it would be interesting to examine the response of price movements to new information or "news," and investigate the connections between the identified jump components with news arrival. An analysis of jumps across markets, or co-jumps, along with volatility transmission would also be a worthwhile endeavor. These topics are left for future research.

BIBLIOGRAPHY

Adrangi, B., Chatrath, A., & David, C. (2000). Price discovery in strategically-linked markets: The case of the gold-silver spread. Applied Financial Economics, 10, 227–234.

Aggarwal, R., & Soenen, L. A. (1991). The nature and efficiency of the gold market. The Journal of Portfolio Management, 14, 18–21.

Akgiray, V., Booth, G., Hatem, J., & Chowdary, M. (1991). Conditional dependence in precious metal prices. The Financial Review, 26, 367–386.

Andersen, T., & Bollerslev, T. (1998a). DM-Dollar intraday volatility: Activity pattern, macroeconomic announcemnets, and longer run dependences. Journal of Finance, 53, 219–265.

Andersen, T., & Bollerslev, T. (1998b). Answering the skeptics. International Economic Review, 39, 885–905.

Andersen, T., Bollerslev, T., & Diebold, F. (2005). Parametric and nonparametric volatility measurement. In Y. Aot-Sahalia, & L. Hansen (Ed.), Handbook of financial econometrics (pp. 67–138). Amsterdam: North-Holland.

Andersen, T., Bollerslev, T., & Dobrev, D. (2007). No-Arbitrage semi-martingle restrictions for continous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. Journal of Econometrics, 138, 125–180.

Andersen, T., Bollerslev, T., Frederiksen, P., & Nielsen, M. Ø. (2010). Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. Journal of Applied Econometrics; Forthcoming, 25, 233–261.

Andersen, T., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility. Journal of Empirical Finance, 6, 457–477.

Andersen, T., Diebold, F., Labys, P., & Bollerslev, T. (2001). Modeling and forecasting realized volatility (NBER Working Paper, No. 8160) Econometrica, 71, 579–625.

Baillie, R., & Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. Journal of Econometrics, 52, 91–113.

Barndoff-Nielsen, O. E., & Shepard, N. (2001). Non-Gaussian OU based models and some of their uses in financial economics. Journal of the Royal Statistical Society, Series B 63, 167–241.

Barndoff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics, 2, 1–48.

Ciner, C. (2001). On the long run relationship between gold and silver rrices: A note. Global Finance Journal, 12, 299–303.

Cumby, R., Figlewski, S., & Hasbrouk, J. (1993). Forecasting volatility and autocorrelation in the EGARCH models. Journal of Derivative, 1, 51–75.

Figlewski, S. (1997). Volatility forecasting. Journal of Financial Markets, Institutions and Instruments, 6, 1–88.

Hansen, P. R., & Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. Journal of Financial Econometrics, 3, 525–554.

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. Journal of Business and Economic Statistics, 24, 127–161.

Kocagil, A. (1997). Does futures speculation stabilize spot prices? Evidence from metal markets. Applied Financial Economics, 7, 113–125.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurement. Journal of Empirical Finance, 12, 445–475.

Krehbiel, T., & Adkins, L. C. (1993). Cointegration tests of the unbiased expectation hypothesis in metals markets. The Journal of Futures Markets, 13, 753–763.

Lucey, B., & Tully, E. (2006). Seasonality, risk and return in daily COMEX gold and silver data 1982–2002. Applied Financial Economics, 16, 48–59.

Malliavin, P., & Mancino, M. E. (2002). Fourier series method for measurement of multivariate volatility. Finance and Stochastics, 6, 49–61.

Mancino, M. E., & Sanfelici, S. (2009). Robustness of Fourier estimator of integrated volatility in the presence of microstrucuture noise. Computational Statisitics & Data Analysis, 51, 2966–2989.

Nielsen, M., & Frederiksen, P. (2008). Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. Journal of Empirical Finance, 15, 265–286.

Patton, A. (2009). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics; Forthcoming.

Poon, S., & Granger, C. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41, 478–539.

Reno, R., & Barucci, E. (2002). On measuring volatility and the GARCH forecasting performance. Journal of International Financial Markets, Institutions and Money, 12, 183–200.

West, K., & Cho, D. (1995). The predictive ability of several models of exchange rate volatility. Journal of Econometrics, 69, 367–391.