Realize the Realized Stock Index Volatility*

Ho-Chuan (River) Huang

Chien-Chung Nieh

Department of Banking and Finance, Tamkang University

This paper constructs estimates of daily stock index volatilities and correlation using high-frequency (one-minute) intraday stock indices. The key feature of these 'realized' volatilities and correlations is that they are not only model-free but also approximately measurement-error-free. In fact, they can be treated as observed rather than latent, so that direct modeling and forecasting of the realized volatilities can be performed using conventional time series approaches. Some interesting results appear in the analysis. Despite the fact that the unstandardized returns are skewed to the right and have fatter tails than normal, the distributions of the raw returns scaled by the realized standard deviations appear to be approximately Gaussian. The unconditional distributions of the realized variances and covariances are leptokurtic as well as highly right-skewed, but the realized correlation tends to be approximately normally distributed. There is no evidence in support of asymmetric volatility effects commonly found in previous findings. However, we find strong evidence to support the fact that there exists high contemporaneous correlation between realized volatilities and high comovement between realized correlation and volatilities.

Keywords: quadratic variation, realized correlation, realized volatility.

JEL classification codes: C22, G11, G12.

I. Introduction

Without doubt, volatility in financial markets has been one of the most analyzed issues in the past decades. This may not appear so surprising due to the fact that volatility is a key element for pricing financial instruments such as options, is a measure of tradeoff between return and risk for allocating assets, and is closely related to portfolio return fractiles which is central to risk management measure like VaR (Value at Risk). Therefore, understanding how to obtain

* Constructive suggestions from the referees are highly appreciated. Any remaining errors are our own responsibility.

reliable measures of asset volatility and how their dynamics evolve over time can be very important for both academic research and practical use.

However, volatility is inherently unobservable. As a result, measures of volatility have to rely on alternative approaches by either parametric or non-parametric estimation. For example, the generalized autoregressive conditional heteroskedasticity (GARCH) model proposed by Engle (1982) and Bollerslev (1986) is a popular parametric method to estimate the conditional volatility. Alternatively, the stochastic volatility (SV) model is another parametric approach to obtain the latent volatility (see Ghysels, Harvey and Renault, 1996). Although these models can characterize the well-documented time-varying and clustering features of the conditional volatility, the validity of the volatility measure relies heavily on the parametric model specifications as well as the specific distributional assumptions. As argued in Andersen, Bollerslev, Diebold and Labys (2001), hereafter ABDL, the existence of multiple competing models indicates the problem of model misspecification. Thus, it remains unclear which or whether any of these specifications provides an appropriate description of the latent volatility behavior. After all, at most, one of the models could be correct and, even worse, they might be all wrong. In addition, the GARCH and SV models are difficult to implement in a multivariate framework for most practical applications.

As an alternative, the data-driven or model-free unbiased estimates of the *ex post* realized volatility can be proxied by the squared returns over the relevant horizons (see Ding, Granger and Engle, 1993). Nevertheless, it is found, for example in Andersen and Bollerslev (1998), that these volatility estimators can be very noisy in standard practice. Thus, the method does not provide a reliable inference for the true underlying latent volatility. On the other hand, Alizadeh, Brandt and Diebold (2002) use the price range, defined as the difference between the highest and lowest log asset price, as a proxy for the latent volatility. In contrast, conventional empirical approach employs data of different sampling frequencies to construct volatility estimators. For instance, Officer (1973) uses monthly returns on an equity index to calculate the annual volatilities. In line with this work, French, Schwert and Stambaugh (1987) construct monthly volatilities from daily returns, whereas Schwert (1990a, b) exploits 15-minute returns in the measurement of daily stock market volatilities. More recently, due to the availability of higher-frequency data, Taylor and Wu (1997), Barndorff-Nielsen and Shephard (2002) and ABDL (2001, 2003) rely on five-minute exchange rate returns for construction of daily exchange rate volatilities, and Andersen, Bollerslev, Diebold and Ebens (2001), henceforth ABDE, obtain the stock return volatilities using a five-year sample of five-minute returns for thirty Dow Jones Industrial Average (DJIA) stocks.

Notably, ABDE (2001) and ABDL (2001, 2003) introduce a novel measure of volatility, termed realized volatility. Motivated by the theoretical work of Merton (1980), they use the sum of high-frequency (five- or 30-minute) returns to construct the measure of volatility at lower frequencies dates. For example, we can obtain daily realized volatility by summing the intraday squared returns. Although the derivation seems trivial, the theory behind this is deep, as discussed in ABDL (2001, 2003). The theory of quadratic variation shows that realized volatility estimates constructed as above are not only model-free, and, as the sampling frequency of the returns approaches infinity, those estimates are measurement-error-free as well. Furthermore, the realized volatility is also an unbiased *ex post* estimator of the daily return volatility. Based on a simulation of integrated volatility implied by the GARCH(1,1) diffusion limit, Andersen and Bollerslev (1998) find that the realized volatility provides a less noisy estimate of the latent volatility than does the squared daily returns. More importantly, the measure of realized volatility allows us to deal with high-dimensional returns situations which might not be feasible using either the multivariate version of GARCH or SV models.

As argued above, the theory for continuous-time arbitrage-free processes indicates that, by sampling the intraday returns at a sufficiently high rate, the realized volatility can be made arbitrarily close to the underlying integrated volatility. This striking feature can allow us, in practice, to treat the volatility as effectively observable rather than latent. Therefore, conventional time series approaches can be readily implemented to directly characterize the distributional properties as well as to model and forecast the dynamics of the realized volatility and correlation. For example, ABDE (2001) use high-frequency intraday prices on thirty stocks in the DJIA over a five-year period to obtain realized volatilities and correlations. They show that the raw returns are symmetric, with fatter tails relative to normal, but returns standardized by the realized standard deviations are near normal. They also find that the realized variances and covariances for all stocks are right-skewed and leptokurtic. However, they document that the realized logarithmic standard deviations and correlations are approximately normal, display strong temporal dependence, and can well be characterized by a fractionally-integrated long-memory process. ABDL (2001) also find quite similar features when high-frequency intraday exchange rate data are used. Built on these empirical regularities, ABDL (2003) proceed to specify and estimate a multivariate fractionally-integrated long-memory Gaussian tri-variate VAR for a set of daily logarithmic realized volatilities of exchange rate returns. When compared with the resulting volatility forecasts obtained from GARCH and other related models, they argue that their approach provides strikingly superior volatility forecasts and well-calibrated density forecasts. In contrast, Maheu and McCurdy (2002) use the realized volatility to explore the nonlinear time series features of latent volatility. In particular, they consider a doubly stochastic process with duration-dependent mixing, and find evidence of time-varying persistence and time-varying variance of volatility. They also argue that the finding of nonlinearity in realized volatility has important implications for measuring forecast performance as well as pricing derivative securities.

Motivated by the work of ABDE (2001) and ABDL (2001, 2003), we examine the stock index return volatilities in both Taiwan Stock Exchange (TSE) and Over the Counter (OTC) using high-frequency intraday return data. However, our analysis is distinct in many aspects. First, ABDE (2001) examine

30 common stock return volatilities, while ABDL (2001, 2003) investigate exchange rate volatility. In contrast, our focus is on the stock index return volatilities. Second, the frequency of data used in ABDE (2001) and ABDL (2001) is five-minute, while we use the intradaily stock index returns data at one-minute frequency over the whole year in 2000.[1] This seems to be the highest frequency available in the literature for constructing realized volatility and makes the theory of quadratic variation more appropriate to apply.

The rest of the paper is organized as follows. In Section II, we provide a brief review of the theoretical background, that is, quadratic variation theory, for constructing the realized volatility. The data used are also detailed and the practical issues involved in measuring daily volatility is discussed. Section III provides a thorough description of the distributional characteristics of the returns (raw vs standardized), volatilities, covariance and correlation. In Section IV, we investigate the dynamics of realized volatilities and correlation. Brief conclusions are finally given in Section V.

II. Measures of Realized Stock Index Return Volatility

Although no formal justification is provided, conventional work such as French, Schwert and Stambaugh (1987) and Schwert (1989) obtained the monthly realized volatilities using daily return observations while Schwert (1990a, b) relied on the 15-minute returns to obtain the daily realized volatilities. A rigorous treatment of the theoretical background can be found in ABDL (2001, 2003) and is briefly reviewed in the following subsection.[2]

II.1 Quadratic variation theory

Consider the following simple multivariate continuous-time stochastic volatility diffusion process,

$$dp_t = \mu_t dt + QdW_t \tag{1}$$

where p_t is the $k \times 1$ instantaneous logarithmic price, μ_t is a drift parameter, and dW_t is a $k \times 1$ standard Brownian motion. The $k \times k$ positive definite diffusion matrix Q_t follows a strictly stationary process and satisfies $Q_t Q_t' = \Omega_t$. For this diffusion, the integral of the instantaneous variances over the day, that is,

$$\tilde{\Omega}_t = \int_t^{t+1} \Omega_\omega d\omega \tag{2}$$

provides an *ex post* measure of the true latent volatility associated with day *t*.

1. We do not use the most recent data since the trading period has been extended from 3 hours to 4.5 hours per day starting from January 1, 2001. In order to avoid possible biases caused by the different trading hours, we restrict ourselves in using the data in 2000 only.
2. The following review relies heavily on ABDL (2001, 2003).

By cumulating the intraday squared returns, as shown in Merton (1980), we can approximate the integrated volatility in equation (2) to any arbitrary precision. In particular, we can obtain an estimate, denoted by $\hat{\Omega}_t$, of $\tilde{\Omega}_t$ as

$$\hat{\Omega}_t = \sum_{j=1}^{\delta} r_{t+j/\delta} \cdot r'_{t+j/\delta} \quad (3)$$

where $r_{t+j/\delta} \equiv p_{t+j/\delta} - p_{t+(j-1)/\delta}$ denotes the continuously-compounded returns, sampled δ times per day. Note that the subscript t indexes the day while j indexes the time within day t. In our applications, we are interested in the analysis of daily volatility. The stock indices of TSE and OTC are available for every minute in the 180-minute trading period (three hours) of a day, implying that $\delta = 180$. The measure, $\hat{\Omega}_t$, is termed realized volatility as in ABDE (2001) and ABDL (2001, 2003).

By the theory of quadratic variation, it can be shown that (3) provides a consistent estimate of latent volatility as

$$\text{plim}_{\delta\to\infty} \hat{\Omega}_t = \tilde{\Omega}_t \quad (4)$$

In other words, as the sampling frequency of the returns increases, $\delta \to \infty$, the *ex post* realized volatility measures so constructed will converge to the integrated latent volatilities.

II.2 Data and construction of realized stock index volatilities

Our empirical analysis is based on the data downloaded from the websites of TSE and OTC, respectively.[3] Since its inception in 1961, TSE has kept pace with market development and the most up-to-date technologies. In contrast, an independent non-profit legal entity, GreTai Securities Market (the OTC market) was established on November 1, 1994.

For those listed companies in the TSE (OTC), the following criteria must be met. First, the corporate has to last over five (two) years since its incorporation. Second, the amount of paid-in capital in its final accounts for the most recent two fiscal years shall be NT$300 (NT$50) million or more. Third, each of the operating profit and before-tax profit for the most recent two years (one year) represents 6% (4%) or greater of the amount of paid-in capital in its final accounts; or the average operating profit and before-tax net profit for the most recent two years represent 6% (2%) or greater of the amount of paid-in capital in its final accounts, and the profitability for the most recent year is greater than that for the immediately preceding year. Besides, it does not have accumulative loss in the most recent fiscal year. Fourth, the number of name-bearing shareholders

3. Most of the introductory description is mainly taken from the following websites. Interested readers can visit http://www.tse.com.tw/docs/eng_home home.htm as well as http://www.gretai.org.tw/e_index.htm for more detailed information.

shall be one thousand (three hundred) or more. Among them, the shareholders holding between one thousand and fifty thousand shares shall not be less than five hundred (three hundred), and the total number of shares they hold shall be 20% (10%) or greater of the total issued shares (or at least 10 (5) million). It can easily be seen that the criteria for listing in the OTC are less stringent than those for listing in TSE.

The stock indices for TSE and OTC are continuously recorded each minute in every trading day.[4] The constituent stocks for the indices are often heavily-traded. One of the measures for liquidity, the overall turnover rate, is about 178.5% for TSE and 771.5% for OTC in the year 2000. Our sample covers the periods starting from January 4, 2000 to December 31, 2000, resulting a total of 271 trading days. In a typical trading day, the market opens at 9:00 and closes at 12:00. This three-hour trading period provides a total of 180 continuously compounded one-minute returns for each day, corresponding to $\delta = 180$ in the notation above.

Based on the one-minute return series, constructed from the logarithmic stock index difference, we can obtain the realized daily volatilities for the index returns of TSE ($s^2_{tse,t}$) and OTC ($s^2_{ots,t}$) as

$$s^2_{tse,t} = \sum_{j=1}^{180} r^2_{tse,\,t+j/180} \tag{5}$$

$$s^2_{otc,t} = \sum_{j=1}^{180} r^2_{otc,\,t+j/180} \tag{6}$$

and the realized covariance between $r_{tse,t}$ and $r_{otc,t}$, denoted by cov_t, can be formed by the cross product series,

$$cov_t = \sum_{j=1}^{180} (r_{tse,\,t+j/180})(r_{otc,t+j/180}) \tag{7}$$

ABDL (2001, 2003) show that, for sufficiently large sampled times (δ), the above realized volatilities and covariance will provide an arbitrarily good approximation to the quadratic variation and covariation.

In addition, we follow ABDE (2001) and ABDL (2001) to examine several related measures of realized volatilities and covariance, including realized standard deviations $s_{tse,t} = \sqrt{s^2_{tse,t}}$ and $s_{otc,t} = \sqrt{s^2_{otc,t}}$, realized logarithmic standard deviations $ls_{tse,t} = \ln(s_{tse,t})$ and $ls_{otc,t} = \ln(s_{otc,t})$, and realized correlation $cor_t = cov_t/(s_{tse,t} \cdot s_{otc,t})$.

4. Since the stock indices are computed using the most recent transaction prices of constituent stocks for every minute, one might expect the indices to be highly autocorrelated as suggested by the referee.

Table 1 Summary Statistics of Daily Return distributions, unstandardized as well as standardized, for the TSE and OTC indices

	r_{tse}	r_{otc}	sr_{tse}	sr_{otc}
mean	−0.2081	−0.2775	−0.2394	−0.2722
std.	1.8852	2.2490	1.4792	1.8000
skew.	0.3576	0.3019	0.2609	0.2338
kurt.	4.5648	4.1139	2.7947	3.0262
JB	33.4246***	18.1282***	3.5507	2.4761
$Q(22)$	50.3350***	50.0631***	34.2600**	78.5455***
$Q^2(22)$	64.0659***	47.7661***	30.0565	37.7552**
ADF	−7.9560***	−6.8351***	−7.5431***	−6.8450***
PP	−19.1971***	−16.7272***	−17.9057***	−15.8567***

Notes: *** Significant at the 1% level. ** Significant at the 5% level. JB represents the Jarque-Bera normality test statistic. $Q(22)$ denotes the Ljung-Box test for returns up to 22nd order autocorrelation whereas $Q^2(22)$ is for squared returns. The Augmented Dickey-Fuller and Phillips-Perron tests for a unit root null hypothesis involving 4 augmentation lags are denoted by ADF and PP, respectively.

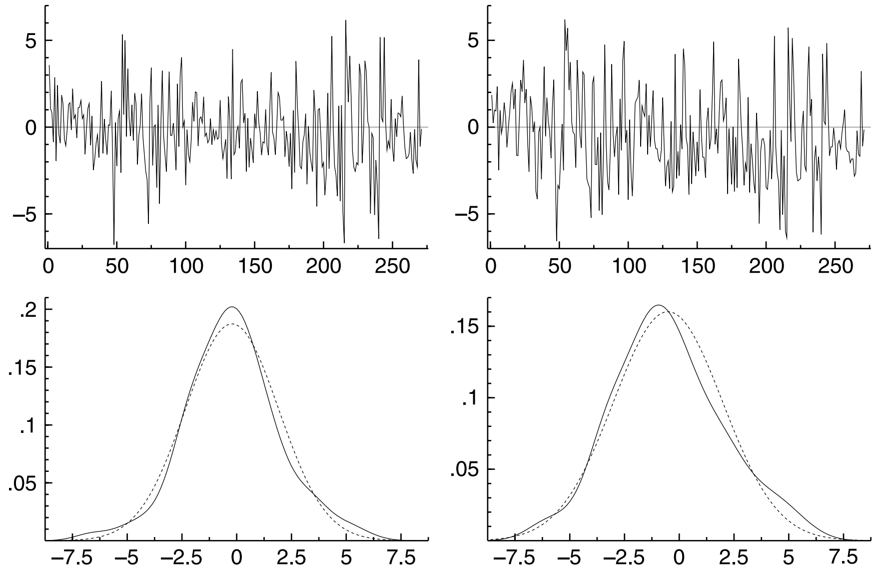
III. The Univariate Distributions

III.1 Returns

Our analysis begins with a summary of the distributions for the raw, or unstandardized, daily TSE (r_{tse}) and OTC (r_{otc}) stock index returns. A standard menu of moments, including mean, standard deviation, skewness and kurtosis, are reported in the first two columns of Table 1 to summarize the unconditional distributions of the daily returns series r_{tse} and r_{otc} respectively. In addition, Figure 1 provides the time series plots as well as the kernel densities of r_{tse} and r_{otc}.

The mean returns for both markets in the year 2000 are both negative, −0.2081 for r_{tse} and −0.2775 for r_{otc}, possibly reflecting a consequence of the bear stock markets in 2000. The stock index of TSE drops from the highest 10393.50 to the lowest 4555.91 while that of OTC falls from the highest 329.47 to the lowest 99.86, with corresponding falling percentages being 56.17% and 69.69% respectively. The standard deviation of r_{otc} is 2.2490 which is larger than that of r_{tse} (1.8852), meaning that the OTC is more volatile. As regards sample skewness, both return series have positive estimates as 0.3576 and 0.3019, implying that the distributions of the returns are not symmetric and are actually slightly right-skewed. This asymmetry is further confirmed by the kernel density estimates shown in the bottom panel of Figure 1. The estimates of the sample kurtoses are well above the normal value of 3, indicating that the distributions of the returns are leptokurtic. These findings are consistent with those found in ABDE (2001). The Jarque-Bera statistics for normality test of r_{tse} and r_{otc} are 33.4246 and 18.1282 with p-values being 0.0000 and 0.0001, respectively. Without doubt, the null of normality can be easily rejected for both series.

Figure 1 The Time Series Plots (top) and Kernel Densities (bottom) of r_{tse} (left) and r_{otc} (right)



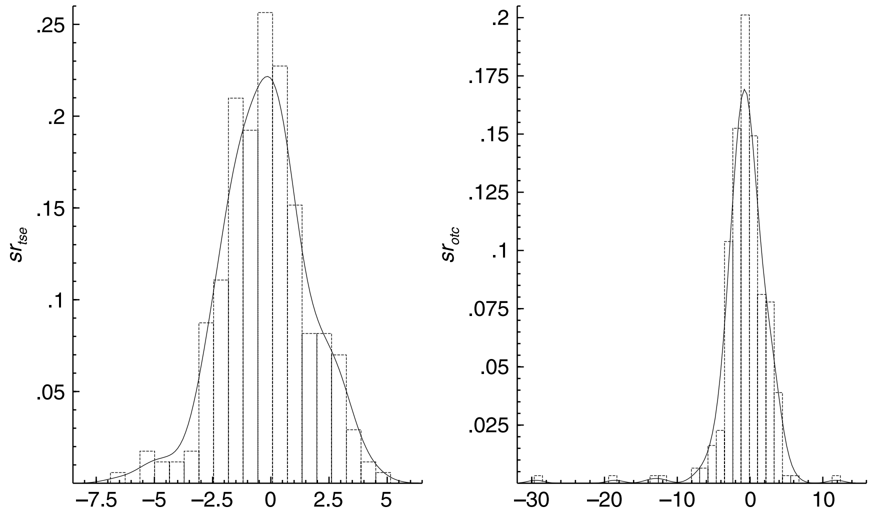
Note: The dotted lines in the bottom panels refer to normal densities.

The standard Ljung-Box portmanteau test for the joint significance of the first 22 auto-correlations (about one month of trading days) of r_{tse}, r_{otc} and the corresponding squared terms r^2_{tse} and r^2_{otc} are also provided in Table 1. Judged by the p-values of the corresponding $Q(22)$ statistics, the hypothesis of zero autocorrelations is clearly rejected for both series, suggesting the returns are highly persistent. On the other hand, the $Q^2(22)$ statistics for both the squared returns series overwhelmingly reject the null of no serial correlation, indicating that there is strong volatility clustering effect in asset returns. However, when we test for the unit root for both returns series, the empirical results reject the null hypothesis of unit root and strongly favor that the returns series are stationary, both by ADF and PP tests.

We now proceed to examine the standardized returns, denoted by sr_{tse} and sr_{otc}, which are obtained by dividing the raw returns with their corresponding realized standard deviation s_{tse} and s_{otc}. Notationally, they are expressed as $sr_{tse} = r_{tse}/s_{tse}$ and $sr_{otc} = r_{otc}/s_{otc}$.

The last two columns of Table 1 also present the moments of the standardized returns. The mean returns for both standardized series remain negative while the standard deviations become smaller. Although the sample skewness coefficients are still positive, the values have been decreased. More strikingly, the sample kurtosis coefficient of sr_{tse} now has a value 2.7947 and that of sr_{otc} becomes

Figure 2 Kernel Densities and Histograms



3.0262. Both values are reduced significantly and are very close to the normal value of 3. When we perform the normality test on the standardized returns, we find that the Jarque-Bera statistics provide values of 3.5507 and 2.4761 for the series sr_{tse} and sr_{otc} respectively. The corresponding p-values are 0.1694 and 0.2900, indicating that the null hypothesis of normality can not be rejected at the commonly-used significance level such as 5%. Similarly, the results can also be seen from the kernel densities and histograms of sr_{tse} and sr_{otc} shown in Figure 2. In contrast to the unstandardized returns, the distributions of the standardized returns are approximately normal. Our findings are consistent with ABDE (2001) and ABDL (2001, 2003), who show that both the stock returns and exchange rate returns standardized by their respective realized standard deviations are approximately Gaussian.

Similarly, we also report the Ljung-Box serial correlation test statistics for the standardized returns sr_{tse} and sr_{otc}. The values are 34.2600 and 78.5455, and are significant at 5% and 1% levels respectively. As a result, we conclude that the strong persistence in standardized returns still remains. As regards the test for autocorrelation in squared standardized returns, we find that the $Q^2(22)$ test statistic for sr_{tse} has been highly decreased and appears insignificant, despite the value for sr_{otc} remaining significant at 5% level. The results suggest that the volatility clustering effects are reduced, or even disappear, once the raw returns are standardized by the realized standard deviations. Not surprisingly, the ADF and PP tests, again, strongly reject the unit root null hypotheses and confirm that both the standardized returns are stationary.

III.2 Volatilities

As discussed earlier, the realized volatilities (s^2_{tse} and s^2_{otc}) can be computed by cumulating intradaily one-minute squared returns as shown in equations (5) and (6). The realized standard deviations, that is, s_{tse} and s_{otc}, can be calculated by taking square root of s^2_{tse} and s^2_{otc} and the realized logarithmic standard deviations, i.e., ls_{tse} and ls_{otc}, are obtained by taking the natural logarithm of s_{tse} and s_{otc} respectively.

From the first two columns of Table 2, we find that the mean realized volatilities of s^2_{tse} and s^2_{otc} for the TSE and OTC index returns are approximately equal, with values being 1.4143 and 1.4890 respectively. Clearly, the standard deviation of the realized volatility is relatively smaller for s^2_{otc} with a value of 0.7485. The sample skewness coefficients are both positive, implying that the distributions for s^2_{tse} and s^2_{otc} are skewed to the right. Specifically, the skewness value of s^2_{tse} is 2.2085 which is larger than that of s^2_{otc}. This suggests that the distribution of s^2_{tse} is more right-skewed than that of s^2_{otc}, and this result is confirmed from the kernel density estimates in the top panels of Figure 3. Turning to the coefficients of the sample kurtoses, we find that the values are much larger than the normal value 3, indicating that the distributions are highly leptokurtic. This is especially true for the realized volatility of TSE returns, that is, s^2_{tse}. The top panels (left panel for s^2_{tse} and right panel for s^2_{otc}) of Figure 3, again, confirm these impressions. As argued in ABDE (2001) and ABDL (2001, 2003), the strong persistence in intraday returns renders the normal distribution a poor approximation, even though the realized volatilities are constructed by cumulating the 180 squared one-minute returns.

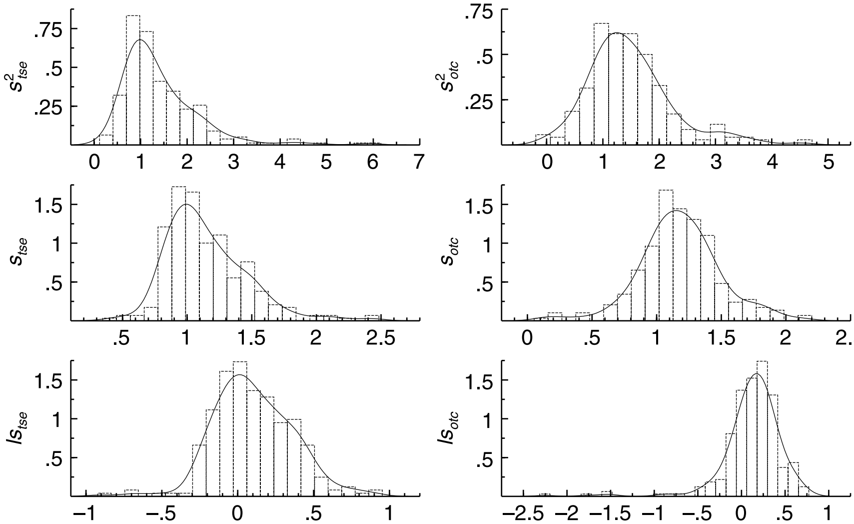
While the distributions of the realized volatilities s^2_{tse} and s^2_{otc} are obviously skewed to the right, transforming to the realized standard deviations, that is, s_{tse} and s_{otc}, moves them toward symmetry. The evidence can be seen from the third and fourth columns of Table 2 and the kernel density estimates as well as histograms in the middle panels of Figure 3. The means of s_{tse} and s_{otc} become smaller, as do their standard deviations. Both the sample skewness coefficients have reduced remarkably. In particular, the skewness of the s_{otc} has a value of −0.0760 which is quite close to symmetric value of 0. Combined with the impression from the middle right panel of Figure 3, we can conclude that the distribution of s_{otc} is symmetric. In contrast, the s_{tse} has a sample skewness value of 1.1258, suggesting that the distribution of s_{tse} remains slightly right-skewed, as can be seen from the left middle panel of Figure 3. However, again, the large kurtosis coefficients indicate that the distributions are still leptokurtic. When turning to the logarithmic standard deviations, we find that the distribution of ls_{tse} appears to be more symmetric, judged by the near-zero skewness coefficient 0.0710 and the kernel density in the bottom left panel of Figure 3. Although the kernel density estimate of ls_{otc} in the bottom right panel of Figure 3 looks like a normal one, it is actually skewed to the left and has a fat tail due to a few outliers. In sum, the evidence leads us to conclude that the distributions of ls_{tse}

Table 2 Summary Statistics of Daily Volatilities, Covariance and Correlation for the TSE and OTC index returns

	s^2_{tse}	s^2_{otc}	s_{tse}	s_{otc}	ls_{tse}	ls_{otc}	*cov*	*cor*
mean	1.4143	1.4890	1.1484	1.1800	0.1046	0.1195	0.5142	0.3289
std.	0.8342	0.7485	0.3096	0.3112	0.2589	0.3439	0.4100	0.1502
skew.	2.2085	1.1250	1.1258	−0.0760	0.0710	−2.8073	1.3289	−0.1275
kurt.	10.2772	5.1160	5.1896	4.4066	4.0687	17.3566	4.6992	2.9404
Q(22)	573.0324***	182.5943***	645.7015***	142.3981***	544.7126***	78.7582***	187.2189***	122.1705***
ADF	−4.3128***	−4.2910***	−3.9490***	−4.5027***	−3.8727***	−5.0330***	−5.0492***	−5.2346***
PP	−8.3775***	−11.9453***	−8.6096***	−11.5581***	−9.5217***	−12.1715***	−11.6895***	−12.6011***

Notes: *** Significant at the 1% level. ** Significant at the 5% level. *Q*(22) denotes the Ljung-Box test up to 22nd order autocorrelation. The Augmented Dickey-Fuller and Phillips-Perron tests for a unit root null hypothesis involving 4 augmentation lags are denoted by ADF and PP, respectively.

Figure 3 Kernel Densities and Histograms



and ls_{otc} are not normal. This is in contrast to ABDE (2001) and ABDL (2001, 2003), who find that the realized logarithmic standard deviations for all stock returns and exchange rate returns appear to be approximately Gaussian.

Turning again to the serial correlation tests, from Table 2, the Ljung-Box statistics indicate strong autocorrelation in the realized volatilities s^2_{tse} and s^2_{otc} at any reasonable significance level. Clearly, the tests also show that realized standard deviations as well as realized logarithmic standard deviations are strongly persistent. This result is consistent with the significant Ljung-Box statistics for the raw squared returns in Table 1. However, as noted in ABDL (2003), the $Q^2(22)$ statistics in Table 1 have smaller magnitudes than those in Table 2, reflecting the fact that the daily squared returns are very noisy volatility proxies relative to the daily realized volatilities. The implication is that the realized volatility should be viewed as the object of intrinsic interest and this is exactly the main topic investigated in later sections. For completeness, we also perform unit root tests. The results from both ADF and PP tests do not show any evidence of nonstationarity. The null hypotheses of unit root can be rejected at 1% significance level for all series, including realized volatilities, standard deviations and realized logarithmic standard deviations.

III.3 Covariance and correlation

Many key financial and economic applications rely on the particular way that volatilities move across assets and markets. In the literature, these issues have been addressed using multivariate ARCH or SV models (Bollerslev, Engle and

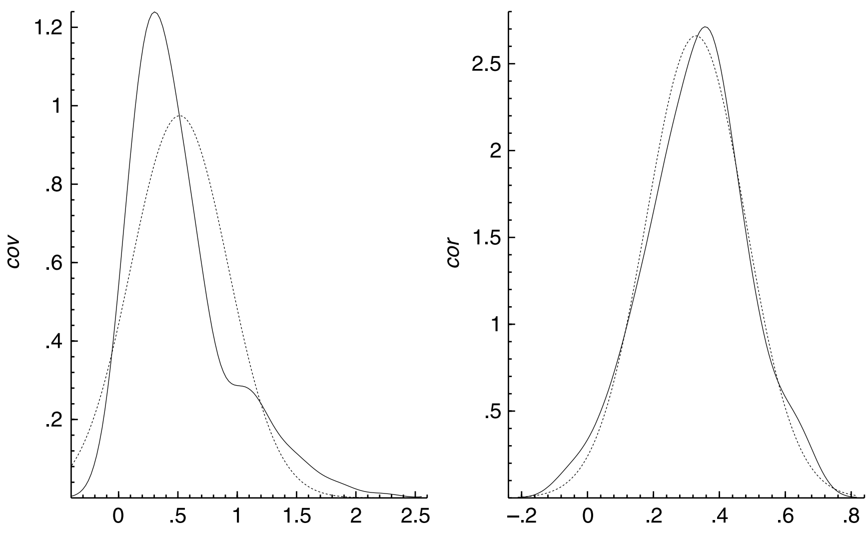
Nelson, 1994; Ghysels, Harvey and Renault, 1996; and Kroner and Ng, 1998). Due to the curse of dimensionality, the target analyzed in those specific parametric models is limited to only a few assets and, thus, severely restricts their practical applications. In contrast, the realized covariance and correlation constructed by the high-frequency intraday returns provide us with a straightforward way to analyze the high-dimensional returns dynamics.

Similarly, we also discuss the distributional properties of realized covariance and correlation. The basic statistics for both series are reported in the last two columns of Table 2. First, we find that both realized covariance (*cov*) and correlation (*cor*) between r_{tse} and r_{otc} are positive, often strongly so. The mean values of the covariance and correlation are 0.5142 and 0.3289, respectively. This suggests that the indices of TSE and OTC move, in general, in the same direction.

The realized covariance has sample skewness and kurtosis coefficients of 1.3289 and 4.6992, suggesting that the distribution of the realized covariance is skewed to the right and has fat tails relative to the normal. The Jarque-Bera test of normality provides a statistic 112.3603 with p-value 0.0000, suggesting that the distribution of *cov* is not Gaussian. The kernel density estimates provided in the left panel of Figure 4 confirm this finding. In addition, the Ljung-Box test indicates that the realized covariance series is highly persistent while no evidence of nonstationarity is found by either ADF or PP test.

In contrast, the realized correlation has a near-zero sample skewness coefficient of −0.1275 and a sample kurtosis value of 2.9404 which is very close to

Figure 4 The Kernel Densities



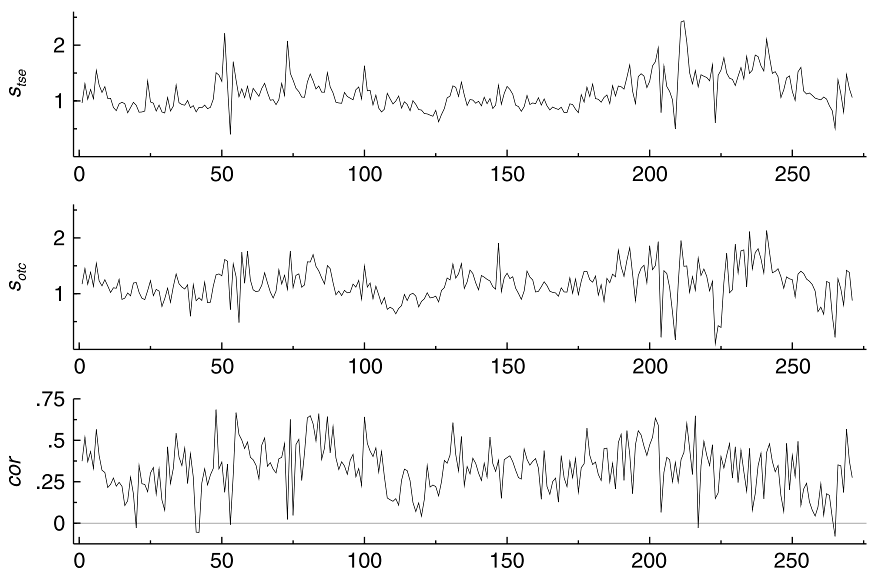
Note: The dotted lines refer to normal densities.

the normal value of 3. The Jarque-Bera normality test provides a statistic of 0.7746 with corresponding p-value 0.6789, indicating that the null hypothesis of normality can not be rejected at any conventional significance level. To illustrate this result, the kernel density of the realized correlation in the right panel of Figure 4 was graphed. Clearly, the normal reference distribution affords a close approximation. Furthermore, the long-run dependence found in realized volatilities and covariance is also present in realized correlation. This is justified by the highly significant value of the Ljung-Box portmanteau statistic of 122.1705 reported in the last column of Table 2. However, the null hypothesis of a unit root for the realized correlation is overwhelmingly rejected when judged by the conventional –3.4566 (–3.4562) 1% critical value using the ADF (PP) test.

IV. The Dynamics of Realized Standard Deviations and Correlation

In this section, the different aspects of volatilities, correlation and their relationship will be examined. In particular, following ABDL (2003), we focus on the realized standard deviations for analysis of volatilities. The main reason for doing this is because the realized standard deviations are measured on the same scale as the returns, and thus, provides a naturally interpretable measure of volatilities. The time series plots of the realized standard deviations $s_{tse,t}$, $s_{otc,t}$ and the realized correlation *cor* are provided in Figure 5. Clearly, all series display large variations.

Figure 5 The Time Series Plots of the Realized Standard Deviations



IV.1 Asymmetric volatility effect

It is well documented in the literature that there is an asymmetric association between volatility and return, namely, negative returns have a larger impact on future volatility than positive returns of the same absolute magnitude. As argued in Bekaert and Wu (2000), the asymmetry may be caused by either the leverage effect (e.g., Christie, 1982), or volatility feedback effect (e.g., Campbell and Hentscel, 1992). In particular, Bekaert and Wu (2000) use data constructed from Nikkei 225 stocks to find support for a volatility feedback explanation but not for leverage effect. ABDE (2001) also confirm the asymmetric relationship that positive returns are associated with smaller volatility innovations than negative returns of the same absolute magnitude.

It is naturally interesting to see if the asymmetry also exists using the realized standard deviations constructed by the intraday stock index returns. First, we report the preliminary results based on the following standard regressions,

$$\underset{}{s_{tse,t}} = \underset{[0.0605]}{0.4884} + \underset{[0.0506]}{0.5690}\, s_{tse,t-1} + \underset{[0.0266]}{0.0107}\, s_{tse,t-1} I(r_{tse,t-1} < 0) \tag{8}$$

$$s_{otc,t} = \underset{[0.0698]}{0.7464} + \underset{[0.0585]}{0.3475}\, s_{ots,t-1} + \underset{[0.0294]}{0.0339}\, s_{otc,t-1} I(r_{ots,t-1} < 0) \tag{9}$$

where figures in brackets are standard errors and $I(\cdot)$ refers to the indicator function. It is found that the coefficients of the lag realized standard deviations, that is, $s_{j,t-1}$, $j = tse$, otc, are both positive and significant at 1 percent level. This strong dynamic dependence seems to suggest the existence of volatility clustering effect. Although the estimates of the terms for capturing asymmetry, that is, $s_{j,t-1} I(r_{j,t-1} < 0)$, are positive, they are not significantly different from 0 at any conventional level. Thus, it provides no evidence to support the fact that there is asymmetry in the impact of past returns on future volatility. This finding is in sharp contrast with previous work where the asymmetric effect is commonly documented.

Note that the above approach uses the lagged returns as the threshold variable with threshold value 0 to examine the possibility of different volatility response to return shocks. In the following, we actually estimate the unknown threshold value rather than restrict the value to be 0. In particular, we first estimate the following threshold autoregressive (TAR) regression for $s_{tse,t}$ using the lagged returns $r_{tse,t-1}$ as the threshold variable. The results are reported as follows,

$$s_{tse,t} = \begin{cases} \underset{[0.1436]}{1.0897} + \underset{[0.1161]}{0.2777}\, s_{tse,t-1} & \text{if } r_{tse,t-1} < -1.9774 \\ \underset{[0.0810]}{0.3046} + \underset{[0.0773]}{0.6834}\, s_{tse,t-1} & \text{if } -1.9774 \le r_{tse,t-1} < 1.9224 \\ \underset{[0.1852]}{0.7793} + \underset{[0.1579]}{0.3422}\, s_{tse,t-1} & \text{if } 1.9224 \le r_{tse,t-1} \end{cases} \tag{10}$$

We find that there are two thresholds with values being −1.9774 and 1.9224. The LM-test for no threshold (linearity) can be rejected at 1% and 5% significance level, respectively. As a result, there are three distinct regimes with a linear first-order autoregressive (AR(1)) process within each regime. Clearly, all AR(1) coefficients are significantly positive. Among which, the AR estimate in the first regime, that is, $r_{tse,t-1} < -1.9774$, is the smallest while the AR coefficient is the largest in the middle regime, that is, $-1.9774 \leq r_{tse,t-1} < 1.9224$. Again, there seems no evidence of asymmetry but we find that the volatility dynamics can be characterized by a nonlinear threshold process.

Turning to the modeling of $s_{otc,t}$, we also use the lagged returns $r_{otc,t-1}$ as the threshold variable to fit a TAR regression as follows,

$$s_{otc,t} = \begin{cases} 1.1413 + 0.2315\ s_{otc,t-1} & \text{if } r_{otc,t-1} < -3.1324 \\ [0.1319]\ \ [0.1044] & \\ 0.7185 + 0.3580\ s_{otc,t-1} & \text{if } -3.1324 \leq r_{otc,t-1} \\ [0.0945]\ \ [0.0781] & \end{cases} \tag{11}$$

In contrast to the above case, we only find a threshold value −3.1324. The null of linearity can be rejected at 1% significance level, suggesting that the volatility movement can be well characterized by a two-regime TAR process. The AR coefficients in both regimes are significantly positive. Specifically, the smaller AR estimate 0.2315 in the first regime, that is, $r_{otc,t-1} < -3.1324$, offer little support of asymmetric volatility effect.

In sum, either judged by the empirical results via the conventional regression approach or the nonlinear threshold modeling, we find no evidence in support of the well-documented asymmetric effect in volatilities $s_{tse,t}$ or $s_{otc,t}$. We also find that the volatility dynamics can be characterized by the multiple-regime nonlinear TAR regressions. In addition, no formal justification, at least statistically, is provided by restricting the threshold values of the lagged returns to be 0 as is done in the literature.

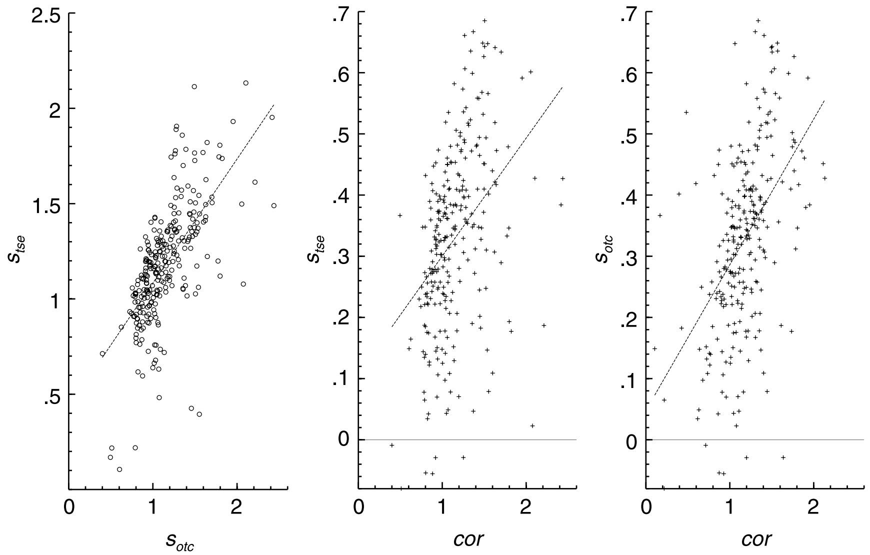
IV.2 The relation between realized standard deviations

Here we investigate various aspects of the relation between the realized standard deviations s_{tse} and s_{otc}. First, in the left panel of Figure 6, we show the bivariate scatter plots of these two series. Clearly, the graph and the regression line indicate a strong positive relation between these two index return volatilities. This is further confirmed by the large value of correlation, 0.6474, between s_{otc} and s_{tse}. In particular, we regress s_{otc} on a constant and s_{tse} to obtain

$$\underset{}{s_{otc,t}} = \underset{[0.0761]}{0.4327} + \underset{[0.0687]}{0.6508}\ s_{tse,t} \tag{12}$$

The slope coefficient estimate is 0.6508 with t test statistic 9.4731. The p-value is 0.0000 and the R^2 is 0.4191, confirming that the relation between s_{otc} and s_{tse} is

Figure 6 The Bivariate Scatterplots



significantly positive. Thus, we find that not only do the two stock index return series move together in the same direction, as indicated by the positive means of *cov* and *cor* in Table 2, but so also do their volatilities.

On the other hand, the causation in conditional variance across various financial asset price movements has attracted increasing attention from both investors and academics. The reason is that understanding how the volatilities transmit across markets may allow us to assess how the market reacts, assimilates and evaluates the arrival of new information. In addition, the causation pattern in variance may provide information in understanding the temporal dynamics of returns series. The examples include Cheung and Ng (1996), Hu, Chen, Fok and Huang (1997), Laopodis (1998), Ng (2000), among others. These studies often require the estimation of the conditional variance by, for instance, a parametric model such as GARCH. In contrast, the realized standard deviations obtained using intraday returns allow us to directly deal with the problem by the conventional Granger-causality test.

Specifically, we consider the following bivariate regression:

$$s_{tse,t} = \alpha_{10} + \sum_{j=1}^{p} \alpha_{1j} s_{tse,t-j} + \sum_{j=1}^{p} \beta_{1j} s_{otc,t-j} + \varepsilon_{1t} \tag{13}$$

$$s_{otc,t} = \alpha_{20} + \sum_{j=1}^{p} \alpha_{2j} s_{otc,t-j} + \sum_{j=1}^{p} \beta_{2j} s_{tse,t-1} + \varepsilon_{2t} \tag{14}$$

First, we employ both Akaike information criterion (AIC) and Schwarz Bayesian information criterion to select the appropriate lag length p, both information criteria suggesting the identical lag length 2, that is, $p = 2$. We further apply the Granger causality test to check for the causal relation between s_{tse} and s_{otc}. The F-statistic for testing the null hypothesis that s_{otc} does not Granger-cause s_{tse} is 1.553, which is not signficantly different from zero at any conventional level, indicating that we cannot reject the null hypothesis. In contrast, the F-statistic shows a value of 9.145, which is different from zero at 1% significance level, suggesting the null hypothesis that s_{tse} does not Granger-cause s_{otc} can be rejected. Thus, the test results show that Granger-causality runs one-way from s_{tse} to s_{otc} but not the other way around.

IV.3 Volatility-in-correlation effect

It is found, for example by ABDL (2001), that realized correlation is highly correlated with the realized volatilities, which they call the ‘volatility effect in correlation’. Along the same line, we now move our attention to the relation between realized volatilities and realized correlation. To get a preliminary idea, the bivariate scatterplots of the [cor, s_{tse}] and [cor, s_{otc}] in the middle and right panels of Figure 6 are shown. Not surprisingly, the plots and the regression line clearly indicate a positive association between realized correlation and volatilities. Specifically, cor on a constant and the realized volatility (either s_{tse} or s_{ots}) are also regressed by the following simple linear regressions:

$$cor_t = \underset{[0.0474]}{0.1076} + \underset{[0.0421]}{0.1927}\, s_{tse,t} \tag{15}$$

$$cor_t = \underset{[0.0450]}{0.0479} + \underset{[0.0374]}{0.2382}\, s_{otc,t} \tag{16}$$

In equation (15), the estimated slope coefficient of cor on s_{tse} is 0.1927 with standard error 0.0421. The estimate is significantly different from 0 at any conventional level, showing that the positive association between s_{tse} and cor is significant. Similar results obtained by regressing cor on a constant and s_{otc} are also shown in equation (16). Again, the slope estimate is significantly positive and with a larger value of 0.2382. The preliminary analysis shows that the realized correlation is highly correlated with realized volatilities s_{tse} and s_{otc}. In fact, it is found that the correlation between cor and s_{tse} is equal to 0.3973 and that between cor and s_{otc} is 0.4935. The results tend to support that there exists a so-called volatility-in-correlation effect.

This so-called ‘volatility-in-correlation effect’ is well documented in the literature. King and Wadhwani (1990) find that correlation across major international stock market tends to increase during periods of market crises. Alternatively, Longin and Solnik (1995) use a bivariate GARCH model and find that the correlations between the major stock markets rise in periods of high volatility.

Ramchand and Susmel (1998), using a regime-switching ARCH model, document that the correlations between the USA and other world markets are on average 2 to 3.5 times higher when the USA market is in a high-variance regime as compared to a low-variance state. This phenomenon is also confirmed by using direct model-free measures of realized correlations and volatilities as in ABDL (2001) and ABDE (2001). In particular, ABDL (2001) plot two kernel density estimates of realized correlation, depending upon whether the volatility measures (logarithm of the realized standard deviations) are less than or larger than their median value of −0.46. They find evidence in support that the realized exchange rate correlations tends to rise on high-volatility days. Similar results are also found in ABDE (2001) when realized stock return volatility and correlation are examined.

In contrast to ABDE (2001) and ABDL (2001), who use the median realized volatility measure as the threshold value for separating small versus large volatility days, we formally estimate the threshold volatility value via the threshold regression of Hansen (2000). In particular, we first examine the following threshold regression

$$cor_t = \begin{cases} \mu_1 & \text{if } s_{j,t} \le \gamma \\ \mu_2 & \text{if } s_{j,t} > \gamma \end{cases} \tag{17}$$

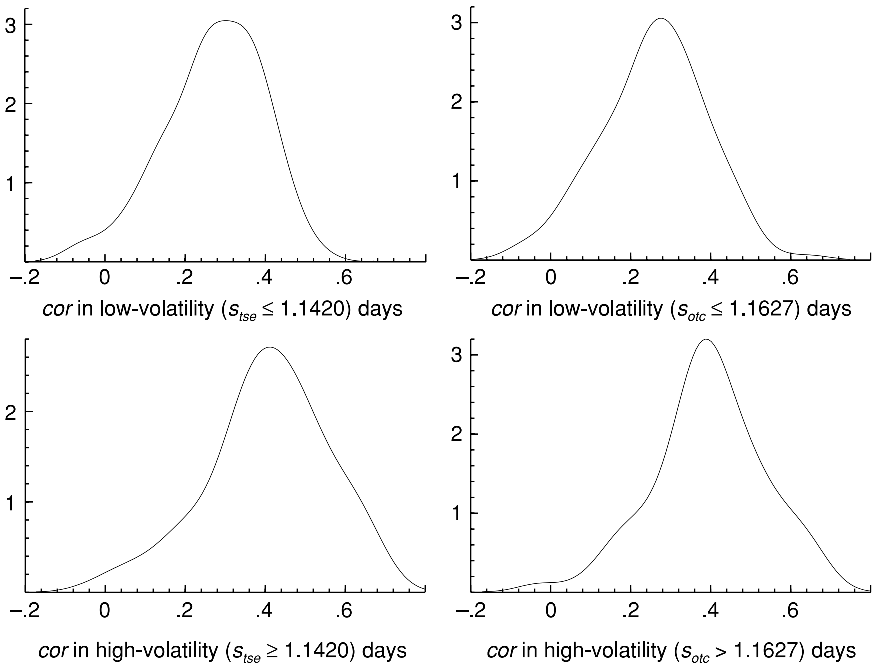
to see if the means (μ_1 and μ_2) of *cor* remain the same in both small ($s_{j,t} \le \gamma$) and large ($s_{j,t} > \gamma$) volatility regimes. The subscript j denotes either '*tse*' or '*otc*'. From Table 3, we find that the LM-tests of no threshold hypothesis using s_{tse} and s_{otc} as threshold variables yield F statistics 17.9750 and 12.5177, respectively. The bootstrap p-values in both cases are 0.0010 and 0.0080, suggesting that there is a threshold effect in *cor* with respect to the realized volatilities s_{tse} and s_{otc}. Furthermore, the mean correlation is 0.2721 (0.2594) in the low-volatility state, that is, $s_{tse} \le 1.1420$ ($s_{otc} \le 1.1627$). In contrast, in the high-volatility state,

Table 3 Volatility-in-Correlation Effect

	s_{tse}			s_{otc}		
	estimate	*2.5%*	*97.5%*	*estimate*	*2.5%*	*97.5%*
μ_1	0.2721	0.2529	0.2943	0.2594	0.2372	0.2989
μ_2	0.4060	0.3761	0.4343	0.3949	0.3721	0.4470
γ	1.1420	1.0879	1.2112	1.1627	1.1248	1.2747
F		17.9750			12.5177	
p		0.0010			0.0080	

Notes: The table summarizes the results of the threshold regressions in equation (17) using s_{tse} and s_{otc} as the threshold variable, respectively. The values of μ_1 and μ_2 denote the mean correlation in the low- and high-volatility regimes whereas the value of γ denotes the estimated threshold volatility value for classifying the correlation into different regimes. The F statistic tests whether there exists a threshold effect with p-value obtained from the bootstrap approach.

Figure 7 Kernel Densities



that is, $s_{tse} > 1.1420$ ($s_{otc} > 1.1627$), the mean correlation becomes larger, as 0.4060 (0.3949). Note that the threshold values for classifying *cor* into different regimes are quite close using either s_{tse} or s_{otc} as the threshold variable. This volatility-in-correlation effect is again confirmed by looking at the kernel density estimates illustrated in Figure 7. It is clear that the distribution of realized correlation shifts rightward when realized volatility increases.

V. Conclusions

This paper investigates the distributional characteristics of the realized volatilities and correlation dynamics using high-frequency intraday stock index returns observations from Taiwan Stock Exchange (TSE) and Over The Counter (OTC). Those measures so constructed are model-free, approximation-error-free and can be, in fact, treated as observed rather than latent. This striking feature can greatly facilitate modeling and forecasting using conventional time series approaches based directly on the observable variables.

Our findings are strikingly similar to those found in the existing work such as ABDE (2001) and ABDL (2001). In particular, the following interesting results come out. First, although the raw returns are right-skewed and leptokurtic, the distributions of the raw returns standardized by the realized standard deviations are approximately normal. Second, the unconditional distributions of the realized variances and covariance are leptokurtic and, in general, highly right-skewed, while the realized correlation appears to be approximately Gaussian. Third, no asymmetric volatility effects are found, that is, negative returns and positive returns have similar impact on future volatilities. In addition, the volatilities of TSE and OTC stock index returns series move closely in the same direction and we find that the volatility in TSE Granger-causes that in OTC but not the other way around. Fourth, we also find evidence in support of comovement between volatilities and correlation. This volatility-in-correlation effect may reduce the benefits to portfolio diversification when the market is most volatile as argued in ABDE (2001).

Last, but definitely not least, one important point made by one referee is that the intradaily one-minute returns are likely to be autocorrelated. The possible autocorrelation might have effect on the moment of the series, in a sense that the series is not drawn out independently from some particular distribution. As a result, it would be more meaningful and appropriate to analyze the residuals after taking into account the autocorrelation. As we concur, this important observation is very constructive and deserves further attention and examination in future related studies.[5]

References

Alizadeh, S., M. W. Brandt and F. X. Diebold, 2002, Range-based estimation of stochastic volatility models. *Journal of Finance*, **57**, pp. 1047–1091.

Andersen, T. G. and T. Bollerslev, 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39**, pp. 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold and H. Ebens, 2001, The distribution of stock return volatility. *Journal of Financial Economics*, **61**, pp. 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys, 2001, The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, pp. 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys, 2003, Modeling and fore-casting realized volatility. *Econometrica*, **71**, pp. 579–625.

Barndorff-Nielsen, O. E. and N. Shephard, 2002, Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B*, **64**, pp. 253–280.

Bekaert, G. and G. Wu, 2000, Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, **13**, pp. 1–42.

Bollerslev, T. 1986, Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, pp. 307–327.

Bollerslev, T., R. F. Engle and D. B. Nelson, 1994, ARCH Models. In: *Handbook of Econometrics*, Vol. 4 (R. F. Engle and D. McFadden) pp. 2959–3038. North-Holland, Amsterdam.

Campbell, J. Y. and L. Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, **31**, pp. 281–318.

Cheung, Y. W. and L. K. Ng, 1996, A causality-in-variance test and its application to financial market prices. *Journal of Econometrics*, **72**, pp. 33–48.

5. We thank a referee for pointing out this important observation to us.

Christie, A. A. 1982, The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, **10**, pp. 407–432.

Ding, Z., C. W. J. Granger and R. F. Engle, 1993, A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, pp. 83–106.

Engle, R. F. 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, **50**, pp. 987–1008.

French, K. R., G. W. Schwert and R. F. Stambaugh, 1987, Expected stock returns and volatility. *Journal of Financial Economics*, **19**, pp. 3–29.

Ghysels, E., A. Harvey and E. Renault, 1996, Stochastic Volatility. In: *Handbook of Statistics* 14, *Statistical Methods in Finance* (ed. G. S. Maddala) pp. 119–191. North-Holland, Amsterdam.

Hansen, B. E. 2000, Sample splitting and threshold estimation. *Econometrica*, **68**, pp. 575–603.

Hu, J. W. S., M. Y. Chen, R. C. W. Fok and B. N. Huang, 1997, Causality in volatility and volatility spillover effects between US, Japan and four equity markets in the South China Growth Triangular. *Journal of International Financial Markets, Institutions and Money*, **7**, pp. 351–367.

King, M. and S. Wadhwani, 1990, Transmission of volatility between stock markets. *Review of Financial Studies*, **3**, pp. 5–33.

Kroner, K. F. and V. K. Ng, 1998, Modeling asymmetric comovements of asset returns. *Review of Financial Studies*, **11**, pp. 817–844.

Laopodis, N. T. 1998, Asymmetric volatility spillovers in Deutsche mark exchange rates. *Journal of Multinational Financial Management*, **8**, pp. 413–430.

Longin, F. and B. Solnik, 1995, Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance*, **14**, pp. 3–23.

Maheu, J. M. and T. H. McCurdy, 2002, Nonlinear features of realized FX volatility. *Review of Economics and Statistics*, **84**, pp. 668–81.

Merton, R. C. 1980, On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, **8**, pp. 323–361.

Ng, A. 2000, Volatility spillover effects from Japan and the US to the Pacific Basin. *Journal of International Money and Finance*, **19**, pp. 207–233.

Officer, R. R. 1973, The variability of the market factor of the NYSE. *Journal of Business*, **46**, pp. 434–453.

Ramchand, L. and R. Susmel, 1998, Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, **5**, pp. 397–416.

Schwert, G. W. 1990a, Stock market volatility. *Financial Analysts Journal*, **46**, pp. 23–34.

Schwert, G. W. 1990b, Stock volatility and the crash of '87. *Review of Financial Studies*, **3**, pp. 77–102.

Schwert, G. W. 1989, Why does stock market volatility change over time? *Journal of Finance*, **44**, pp. 1115–1153.

Taylor, S. J. and X. Z. Xu, 1997, The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, **4**, pp. 317–340.