ECONOMETRIC ANALYSIS OF FINANCIAL MARKETS USING HIGH-FREQUENCY DATA

By

Kun Yang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

December, 2006

Nashville, Tennessee

Approved:

Professor Mototsugu Shintani

Professor Mario J. Crucini

Professor Yanqin Fan

Professor Clifford Ball

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This work was generously supported by the Household International Research Award, the Summer Research Award, the Walter Noel Dissertation Fellowship and the Richard Hoover Dissertation Enhancement Award at Vanderbilt University. Without these financial supports, this work would not have been possible.

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. I would especially like to thank Dr. Mototsugu Shintani, the chairman of my committee. As my teacher and mentor, he has taught me more than I could ever give him credit for here.

Nobody has been more important to me in the pursuit of my doctoral degree than the members of my family. I would like to thank my parents, whose love and guidance are with me in whatever I pursue. They are the ultimate role models.

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

TIME HORIZON SENSITIVITY OF THE FORWARD PREMIUM PUZZLE: IN-SAMPLE FIT AND OUT-OF-SAMPLE FORECAST

Introduction

One of the most examined hypotheses in international finance is whether the foreign exchange forward rate is an unbiased estimator of the future spot exchange rate, that is,

$$E_t[s_{t+k}] = f_{t,k}, \tag{1}$$

where s_{t+k} is the (log) spot exchange rate at time $t+k$, $f_{t,k}$ is the (log) forward exchange rate at time t with maturity k.[1] This unbiasedness hypothesis may be understood as a joint hypothesis of risk neutrality, efficient market, and rational expectations. Simply speaking, if investors are risk-neutral, ignoring the inflation factor and transaction costs, they should expect future spot exchange rates to be equal to the forward rates of corresponding maturities (that is, $E_t^m[s_{t+k}] = f_{t,k}$, where E_t^m refers to the investors' subjective expectation conditional on the information set at time t). Otherwise, they will take equal and opposite positions in the forward and the future spot transactions, expecting to make profits. These trading activities will continue until the equality ($E_t^m[s_{t+k}] = f_{t,k}$) holds. And if the investors' expectations are rational, that is, $E_t^m[s_{t+k}] = E_t[s_{t+k}]$, where E_t is the true population expectation conditional on information available at time t, then the unbiasedness hypothesis should hold.

Though theoretically justifiable, the unbiasedness hypothesis has often been rejected in the empirical literature. The common finding is that the forward rates predict the future exchange rate changes with the wrong sign. For example, the following regression:

$$\Delta s_{t,k} = s_{t+k} - s_t = \alpha + \beta(f_{t,k} - s_t) + \varepsilon_{t+k} \tag{2}$$

[1] Throughout this paper, we use small letters for variables in log, and capital letters for variables in level; also, for both spot and forward exchange rates, we express foreign currency in terms of domestic currency.

appears frequently in the literature and the joint hypothesis of $\alpha = 0$, $\beta = 1$ is tested. It is often found that the estimated slope coefficient is negative, and is significantly different from 1. As noted by Froot (1990), the average $\hat{\beta}$ over 75 published papers is -0.88. A huge number of papers in the literature have been written to explore the possible explanations, which include: risk premiums (e.g. Engel, 1996), irrational expectations (e.g. Frankel and Rose, 1994), learning and peso problems (e.g. Lewis, 1994; Evans, 1995), measurement errors (e.g. Cornell, 1989; Bekaert and Hodrick, 1993), long memory of forward premiums (e.g. Baillie and Bollerslev, 1994; Maynard and Phillips, 2001), and limits to speculation (e.g. Lyons, 2001; Sarno, Valente and Leon, 2004). So far, however, none of these explanations successfully accounts for the magnitude of the discrepancy between the forward rate and its associated future spot rate. As a result, this discrepancy is an unsolved anomaly called the "forward premium puzzle".[2]

In this paper, we aim at providing another perspective on the forward premium puzzle: its sensitivity to the horizon k. To determine the motivation behind it, let us first sketch a consumption-based asset pricing model, linking the future spot rates to the forward rates. Following Lucas (1978), Hansen and Singleton (1982), and Obstfeld and Rogoff (1996), consider an infinitely-lived representative agent who faces the optimization problem as follows:

$$Max\ E_0 \textstyle\sum_{s=t}^{\infty} \beta^t U(C_t),$$
$$s.t.\ C_t + \textstyle\sum_{k=1}^{K} P_t^k Q_t^k \leq \sum_{k=1}^{K} R_{t-k}^k Q_{t-k}^k + W_t,$$

where C_t and W_t are consumption and (real) wage in time period t. The collection of K assets have different investment horizons, P_t^k and Q_t^k are the price and quantity purchased of asset k at date t, and R_{t-k}^k is the date t (real) payoff from holding a unit of asset k purchased at date $t-k$. Solving the first order condition gives us the Euler condition:

$$1 = \beta^k E_t[\frac{R_t^k U'(C_{t+k})}{P_t^k U'(C_t)}],\ k = 1, 2, \cdots, K.$$

Let $r_{t,k} = \dfrac{R_t^k}{P_t^k}$, that is, the (real) gross return of holding one unit of asset k in time period t. We then get:

$$1 = E_t[r_{t,k} \cdot \beta^k \cdot \frac{U'(C_{t+k})}{U'(C_t)}].$$

[2] Sometimes it is called the forward discount puzzle or the forward discount/premium anomaly. In international finance literature, it is often referred to as a violation of the uncovered interest parity (UIP).

In words, the (real) gross return of an asset ($r_{t,k}$), when discounted by an appropriate factor ($\beta^k \cdot \frac{U'(C_{t+k})}{U'(C_t)}$), should be equal to 1. The result can be easily generalized to the case where there are N kinds of assets of the same horizon k. Under that situation, we will have:

$$1 = E_t[r_{t,k}^n \cdot \beta^k \cdot \frac{U'(C_{t+k})}{U'(C_t)}],\ n = 1, 2, ..., N;\text{ or,}$$

$$0 = E_t[(r_{t,k}^n - r_{t,k}^m) \cdot \frac{U'(C_{t+k})}{U'(C_t)}],\ \ n, m = 1, 2, ..., N,\ \ n \neq m. \tag{3}$$

Now let us see what it implies in our case. Consider a nominal home-currency bond with maturity k (denoted as asset 1), and a portfolio constructed as follows (denoted as asset 2): buy foreign currency and invest in a nominal foreign-currency bond with maturity k, and then sell for domestic currency at the maturity date. The (real) gross returns of these two assets will be:

$$r_{t,k}^1 = (1 + i_{t,k}^1) \cdot \frac{P_t}{P_{t+k}}\ ,\ r_{t,k}^2 = \frac{S_{t+k}}{S_t}(1 + i_{t,k}^2) \cdot \frac{P_t}{P_{t+k}}\ ,$$

where $i_{t,k}^n$ is the domestic ($n = 1$) or foreign ($n = 2$) nominal interest rate (in level) of period k, S_t is the spot exchange rate (in level), P_t is the domestic price level. Imposing the covered interest parity condition $S_t(1 + i_{t,k}^1) = F_{t,k}(1 + i_{t,k}^2)$,[3] where $F_{t,k}$ is the forward rate (in level) with maturity k, we have:

$$r_{t,k}^1 - r_{t,k}^2 = \frac{P_t}{P_{t+k}} \cdot (1 + i_{t,k}^1)(\frac{F_{t,k} - S_{t+k}}{F_{t,k}}). \tag{4}$$

Combining (3) and (4), and factoring out the term $P_t(1 + i_{t,k}^1)\frac{1}{F_{t,k}}$ (since they are in the information set at t), we get:

$$0 = E_t[(\frac{F_{t,k} - S_{t+k}}{P_{t+k}})\frac{U'(C_{t+k})}{U'(C_t)}].$$

Now assuming the CRRA preference ($U(C) = \frac{C^{1-\rho}}{1-\rho}$), and also assuming $\{F_{t,k}, S_{t+k}, P_{t+k}, C_{t+k}\}$ are jointly

[3] Covered interest parity is generally found to hold in real data (see Sarno and Taylor, 2002, chapter 2 for a survey of the evidence).

log-normally distributed, we arrive at the following equation:

$$E_t[s_{t+k}] = f_{t,k} - 0.5Var_t(s_{t+k}) + Cov_t(s_{t+k,}p_{t+k}) + \rho Cov_t(s_{t+k}, c_{t+k}).^4 \quad (5)$$

Though the model is built on some assumptions (CRRA preferences and log normal distributions), it demonstrates the sense that the unbiasedness hypothesis (1) may be contaminated by other factors in (5). Moreover, the extent of contamination may depend on the horizon k. On one hand, when the horizon k is short (say, one day or one week), macro factors (for example, p and c here) are unlikely to change much. Hence, we may expect that the conditional covariances in equation (5) are of very small magnitude. On the other hand, since the (log) exchange rates s_{t+k} are commonly found to be non-stationary (for example, see Hodrick, 1987, Section 3.5), the conditional variance $Var_t(s_{t+k})$ is likely to increase with k.[5] Combining these two arguments, we might expect that the unbiasedness hypothesis (1) is less contaminated at short horizons.[6] For medium and long horizons, because the interaction of the last three terms in (5) is unknown, we leave it to the empirical part to see how the unbiasedness hypothesis is affected.

Recent empirical evidence also suggests the time horizon-sensitivity of the unbiasedness hypothesis. For example, Chaboud and Wright (2005) test the uncovered interest parity (UIP) (which is equivalent to the unbiasedness hypothesis if the covered interest parity holds) using high frequency data, and find that UIP holds at the 1-day horizon. On the other end, Chinn and Meredith (2004) find supporting evidence of UIP over very long horizons (5 and 10 years).

The above empirical findings, however, may partly be due to a different methodology, or a different dataset, instead of purely due to a change of time horizon. For example, in Chaboud and Wright (2005), the evidence is found only when small-interval exchange returns are regressed on multi-day interest differentials, which is a different method from what has been used in the literature. In Chinn and Meredith (2004, page 415), their long-horizon interest rates data, “are inherently somewhat less pure from the point of view of

[4] Following the literature, $-0.5Var_t(s_{t+k}) + Cov_t(s_{t+k},\ p_{t+k})$ are referred to as the "Jensen's Inequality Terms" and $\rho Cov_t(s_{t+k}, c_{t+k})$ is referred to as the "risk premium."

[5] For a simple illustration, assume that s_t follows a random walk, $s_t = s_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d\ N(0,1)$. Then $s_{t+k} = s_t + \varepsilon_{t+1} + \varepsilon_{t+2} + ... + \varepsilon_{t+k}$, and $Var_t(s_{t+k}) = Var_t(s_{t+k} - s_t) = k$.

[6] We note that the Jensen's Inequality Terms are empirically found to be small (see, for example, Engel, 1996, page 133). However, we are not sure if this empirical finding is robust to different horizons. Even though the finding is true over all horizons, the risk premium term will still make the unbiasedness hypothesis sensitive to the time horizon.

the UIP hypothesis."

To get a precise sense of how the time horizon matters, we employ the commonly used in-sample method on a large dataset over a much wider variety of horizons than those examined in the existing literature. Specifically, we obtain the forward rates of nine horizons (ranging from 1 day to 1 year) from Bloomberg, one of the largest financial data providers over the world. We then run the regression (equation 2) to test the unbiasedness hypothesis for each horizon. By applying the same methodology to the data from the same source, we aim to control for factors other than the time horizon.

Meanwhile, to complement our understanding of the horizon-sensitivity of the puzzle, we also evaluate the out-of-sample forecastability of spot exchange rates using the forward premiums (without imposing theoretical restrictions; hereafter, we denote it as the forward premium model) across various horizons. As indicated by Chinn and Meredith (2004, note 8), even though the unbiasedness hypothesis may not hold in sample, we might still expect that the forward premiums have predictive power in forecasting the future spot rates. And similarly to the above argument, depending on the magnitude of macro factor contamination, the predictive power of the forward premiums may vary with the forecast horizon. Therefore, we examine the forecast performance of the forward premium model relative to that of eight linear/nonlinear exchange rate models across different horizons. Specifically, we first carry out a comparison between the forward premium model and the random walk model, which is the most popular benchmark model in forecasting exchange rates. We then use the reality check procedure proposed by White (2000) to implement multiple comparisons between the forward premium model and the competing forecast models.

The remainder of the paper is organized as follows: Section 2 explains the data; Section 3 analyzes the time series properties of the data; Section 4 and 5 examine the in-sample fit and the out-of-sample forecast, respectively; Section 6 check the robustness of the out-of-sample performance. Some concluding remarks are presented in Section 7.

Data

We obtain daily data of seven major currencies from Bloomberg: the Australian Dollar (AD), the British Pound (BP), the Canadian Dollar (CD), the Deutsche Mark (DM), the Euro, the Japanese Yen

(JY), and the Swiss Franc (SF), expressed in units of US Dollar per unit of currency. For each currency, we have 10 series: spot, spot/next, 1-week, 2-week, 3-week, 1-month, 2-month, 3-month, 6-month, 1-year (forward) rates. All are the closing middle prices. Weekend data are excluded because of the negligible trading during weekends. For missing weekday data, we substitute the previous day's prices.

All the data (except for the DM and the Euro) range from January 1, 1994, to April 13, 2004, with a total of 2,682 observations for each currency. For the DM and the Euro, the data range from January 3, 1994, to December 31, 1998, and from January 1, 1999, to April 13, 2004, with a total of 1,304 and 1,378 observations, respectively.

To closely align the future spot rates with their corresponding forward rates, we need to determine the settlement dates of forward contracts. The settlement convention in the foreign exchange market is as follows.[7] The spot rate settles in two business days after the trade. The spot/next forward rate settles on the next business day after the spot settlement date. For forward rates with a horizon above 1 week (including 1 week), the settlement is similar: if the horizon is 1-week (or 1-month, etc.), then the settlement occurs 1 week (or 1 month, etc.) after the spot settlement date. Exceptions regarding weekends and holidays are described in detail in Walmsley (2000).

In this paper, to comply with the above market convention, we construct the return series $\Delta s_{t,k}$ ($= s_{t+k} - s_t$), and their associated forward premium series ($f_{t,k} - s_t$), in the following way:

1) For spot/next forward rates, $k = 1$, that is, the spot/next rate is treated as 1-day forward rate;[8]

2) For 1-week forward rates, $k = 7$. This is based on the fact that the 1-week forward contract commences in 2 business days (spot settlement date), and there are 5 observations per week (weekend data are excluded);

3) For 2-week forward rates, $k = 7 + 5 = 12$;

4) For 3-week forward rates, $k = 7 + 2 * 5 = 17$;

5) For 1-month forward rates, $k = 24$. This is based on the fact that we have 2,682 daily data for each series, while they span 123.5 months. So on average, there will be 22 daily data per month. Since the contract commences in two business days, k should be $2 + 22 = 24$;

6) For 2-month forward rates, $k = 2 + 22 * 2 = 46$;

[7] Walmsley (2000) and Bloomberg terminal are good references for the settlement convention in the foreign exchange market.

[8] For more discussion of using spot/next rates as the 1-day forward rates, we refer to Chaboud and Wright (2003).

7) For 3-month forward rates, $k = 2 + 22 * 3 = 68$;

8) For 6-month forward rates, $k = 2 + 22 * 6 = 134$;

9) For 1-year forward rates, $k = 263$. This is based on the fact that we have around 261 daily data per year, and the commencing date is 2 business days from the contract date.

The final series of foreign exchange returns and their associated forward premiums range from January 3, 1994, to April 9, 2003, with a total of 2,418 observations per series. For the DM and the Euro, the final series ranges from January 3,1994, to December 31, 1997, and from January 1, 1999, to April 9, 2003, with a total of 1,043 and 1,112 observations per series, respectively.

Unit root and long memory properties

To obtain reliable regression results, we need to make sure that both the regressors and the regressands are stationary. There has been a universal agreement on the non-stationarity of both the (log) spot rates and the (log) forward rates. Less consensus, however, has been achieved on the properties of their first differenced series, that is, the foreign exchange returns and the forward premiums.[9] Therefore, we test the unit root hypothesis in the foreign exchange returns and forward premiums. Specifically, we use the Phillips-Perron tests, with an intercept and a time trend included in the standard Dickey-Fuller regression, and calculate the residual spectrum at frequency zero using the Bartlett kernel with bandwidth $k + 1$.[10]

Recent evidence also shows that the foreign exchange returns and the forward premium series have different memory properties, which makes conventional statistical theory inapplicable to the regression model (2) (e.g. Baillie and Bollerslev, 1994; Maynard and Phillips, 2001). Therefore, in addition to the unit root tests, we also estimate the long memory parameters of both the return and the forward premium series. In particular, we estimate the fractionally integrated model:

$$(1 - L)^d (x_t - \mu) = u_t,$$

[9] In a non-overlapping case, it is often documented that the exchange rate return series are serially uncorrelated and stationary. However, when overlapping return series, such as a 1-week or a 1-month return series at the daily frequency, are involved, we remain unsure until appropriate tests are implemented.

[10] The results using alternative methods are similar and available upon request.

where $\{x_t\}$ is the return or the forward premium series, $\{u_t\}$ is a short memory time series. The parameter d determines the memory behavior of the returns or the forward premiums series, as described in Maynard and Phillips (2001). We follow the modified log periodogram procedure by Kim and Phillips (1999), and report the point estimates, as well as the 95% confidence intervals for d.

The results are given in Table 1. In terms of unit root tests, we find that both the return and the forward premium series do not contain unit roots at short horizons (within 3-week). From 1-month up to the 6-month horizon (the cut-off horizon varies among different currencies), we find that the return and the forward premium series have different time series properties. On average, over these medium horizons, we can reject the unit root hypothesis in the return series, but not in the forward premium series. The only exception involves the DM, where the result is opposite. When the horizon is 1 year, however, both series contain unit roots.

With respect to the long memory behavior, there are also three cases. Over a very short horizon (1 day), both the return and forward premium series are stationary ($d < 1/2$). And for most of the currencies (except for the CD), these two series are likely to have short memory rather than long memory, that is, d is closer to 0 than to 1/2. Over intermediate horizons (from 1 week up to 1 month, the cut-off horizon varies among different currencies), the return series are likely to be long memory, non-stationary ($d > 1/2$), while the forward premium series are likely to be long memory, stationary ($d < 1/2$). When the horizon is beyond 1 month, both series are long memory, non-stationary ($d > 1/2$). And the longer the horizon, the more likely it is that both series contain unit roots ($d = 1$).

In summary, the time series properties of the return and forward premium series vary over different horizons. Generally, over short horizons, both series are likely to be short memory, stationary. Over intermediate horizons, the two series tend to have divergent time series properties, in terms of both unit root and long memory results. Over long horizons, both series become non-stationary, and seem to contain unit roots.

In-sample fit

Having examined the time series properties of the return and the forward premium series, we now turn to the in-sample analysis. Specifically, for each currency, we run regression (2) over the nine horizons, and report the point estimates of α and β, their standard errors, and the adjusted R^2.[11] At the 1-day horizon, there are no overlapping observations, and hence the error terms are not autocorrelated. Therefore, we compute White's standard errors. For other horizons, overlapping observations are involved, which induces moving average terms of order $k-1$ in the errors. Therefore, the Newey-West heteroskedasticity and autocorrelation consistent (HAC) standard errors are reported. For the convenience of a visual check, we also construct Figure 1, where, for each currency, the center line connects the point estimates over all horizons, and the lower and upper lines connect the lower and upper bounds of the 95% confidence intervals.

The results are reported in Table 2 and several patterns can be observed. First, consistent with the literature, the point estimate of β is often negative, and significantly different from 0 for horizons longer than 1 month (including 1 month). For horizons shorter than 1 month, which are less frequently examined in the literature, the results are mixed. At the 1-day horizon, $\hat{\beta}$ is insignificantly different from 0 for all currencies. Over horizons from 1 week to 3 weeks, $\hat{\beta}$ is significantly negative for AD, EURO, SF at all the three horizons, and for JY at the 1-week and 2-week horizons; while $\hat{\beta}$ is insignificantly different from 0 for BP, CD, DM at all the three horizons, and for JY at the 3-week horizon. In all cases, the unbiasedness hypothesis ($\beta = 1$) is rejected.

Second, the adjusted R^2 increases with the horizon for all currencies. At the 1-day horizon, the adjusted R^2 is negligible, indicating that the forward premiums have little power in explaining the returns. However, the adjusted R^2 increases as the horizon lengthens, with an average of 0.49 at the 1-year horizon. This fact seems to suggest that over longer horizons, the forward premiums may contain more information for forecasting the returns.[12] This point will be further explored in the out-of-sample analysis.

Third, the point estimate of β tends to decrease as the horizon goes from 1-day to 1-year. As can be seen from Figure 1, for most currencies (except EURO and JY), the center line connecting point estimates shows a downward sloping trend with the horizons. For EURO and JY, $\hat{\beta}$ has unusually large size (in absolute

[11] The evidence in Section 3 indicates a discrepancy in the time series properties between the regressand and the regressors, over various horizons. This discrepancy may render unreliable conventional statistical tests (Maynard and Phillips, 2001). Nonetheless, since we aim to examine the horizon-sensitivity of the puzzle, we proceed with this regression method while taking some caution in interpreting the results.

[12] This finding is consistent with the evidence reported in Mark (1995). In his paper, Mark uses fundamentals instead of forward premiums as the regressors, and finds that the adjusted R^2increases from 0.01 to 0.64, when the horizon is lengthened from 1 to 16 quarters.

value) at 1-week and 2-week horizons, which obfuscates the downward trend. On average, however, $\hat{\beta}$ still tends to be more negative at long horizons than at short horizons for these two currencies.

Since the third finding is relatively unexplored in the literature, and may possibly be due to the overlapping feature in our data, we further check its robustness to non-overlapping cases as follows. For each horizon (except for 1-day), we create the non-overlapping data by sampling at the corresponding frequency from the original daily data. For example, for the 1-week horizon, we sample the 1-week foreign exchange returns and forward premiums at a weekly frequency. Since $k = 7$ for the 1-week horizon, we obtain seven weekly samples by changing the initial sample point. We then run the regression for each sample and report the lower quantile, upper quantile, median, mean, and standard deviation of the estimates (denoted as beta1w) . Similar methodology applies to other horizons. As seen from Table 3, the results of the non-overlapping case are very similar to those of the overlapping case. Take AD as an example. From the statistics, the empirical distribution of the estimates $\hat{\beta}$ shifts to the left as the horizon lengthens, which is in accordance with the downward trend of $\hat{\beta}$ in the overlapping case. Also, the mean of $\hat{\beta}$ at each horizon in the nonoverlapping case has similar magnitude to the corresponding point estimate in the overlapping estimation. Therefore, the third finding is not due to the overlapping sampling in our methodology.

In summary, we have the following in-sample findings. First, the unbiasedness hypothesis is often rejected, consistent with the literature. Second, the deviation from the unbiasedness hypothesis increases with the horizon. These two findings are partially in line with our conjecture in the introduction part, namely, the unbiasedness hypothesis may be contaminated by other factors, and the contamination may be lesser at shorter horizons. Third, without any restriction on the regression model, the explanatory power of the forward premiums increases with the horizon, as indicated by the adjusted R^2.

Out-of-sample forecast

Ever since the influential works of Meese and Rogoff (1983a, b), out-of-sample techniques have frequently been used in forecasting foreign exchange rates. The common finding in the literature is that the random walk model performs at least as well as any structural or time series exchange rate model. To our best knowledge, however, considerably less work has been done to examine the out-of-sample performance of the forward premium model.[13] Meanwhile, in terms of evaluation methods, statistical criteria, such as the mean squared forecast error (MSFE), the mean absolute forecast error (MAFE) or root mean squared error (RMSE) are often used, which may not be as important as economic measures, such as the direction-of-change statistics (Granger, 1999).[14]

In this section, therefore, we compare the out-of-sample performance of the forward premium model (FP)[15] with the random walk (RW) model, using both statistical and economic measures. To be more specific, we use four measures to evaluate the forecast performance: the MSFE, the MAFE, the MFTR (mean forecast trading return), and the MCFD (mean correct forecast direction), which are defined below. Also, we use the rolling scheme to construct out-of-sample forecasts.[16] Specifically, for each horizon k, we divide the whole sample (T observations of k-period foreign exchange returns and forward premiums) into first R and last P observations ($T \equiv R+P$). We then use the regression result from the first R observations to forecast the k-period-ahead returns at time $R+1$ ($\Delta\hat{s}_{R+1,k}$). The observations from 2 to $R+1$ are then used to forecast the k-period-ahead returns at time $R+2$ ($\Delta\hat{s}_{R+2,k}$). This procedure continues until the last forecasted return ($\Delta\hat{s}_{R+P,k}$) is obtained, yielding a total of P forecasts. We then construct the four measures as follows:

$$MSFE = P^{-1}\sum_{t=1}^{P}(\Delta s_{R+t,k} - \Delta\hat{s}_{R+t,k})^2 \ ,$$
$$MAFE = P^{-1}\sum_{t=1}^{P}|\Delta s_{R+t,k} - \Delta\hat{s}_{R+t,k}| \ ,$$
$$MFTR = P^{-1}\sum_{t=1}^{P} sign(\Delta\hat{s}_{R+t,k})\Delta s_{R+t,k} \ ,$$
$$MCFD = P^{-1}\sum_{t=1}^{P} 1(sign(\Delta\hat{s}_{R+t,k})sign(\Delta s_{R+t,k}) > 0) \ .$$

[13] Meese and Rogoff (1983a) compare the out-of-sample performance of the random walk model with that of the forward premium model at 1, 3, 6, and 12 months. The comparison, however is solely based on statistical measures, and no formal tests of whether the differences are statistically significant are employed. We also note that Clarida and Taylor (1997) and Clarida, Sarno, Taylor, and Valente (2003) examine the out-of-sample forecast performance of the forward premiums, relative to alternative models across different horizons. Their focus, however, is on the the term structure models of forward premiums.

[14] Cheung, et al. (2003) use a direction-of-change measure to compare the forward premium model with the random walk model. In contrast with our unrestricted model, however, they use the restricted coefficient ($\beta = 1$) to construct out-of-sample forecasts, and find that the restricted model cannot beat the random walk model.

[15] We also check the FP model with the restriction of $\alpha = 0, \beta = 1$, or just $\beta = 1$, and find the restricted models often perform worse than the RW model, which is consistent with the evidence in Cheung et al. (2003).

[16] There are three prevalent out-of-sample forecast schemes: recursive, rolling, and fixed. See MacCracken (2004) for a description and comparison of these three methods. In view of the large size of our data and a possible time-varying risk premium (e.g. Fama, 1984; Engel, 1996), we think that the rolling scheme may be more appropriate in our case.

$MSFE$ and $MAFE$ are traditional statistical measures, with smaller values equivalent to better forecast performance. $MFTR$ and $MCFD$ are economic measures, with the former evaluating the average rate of return from forecasting, and the latter relating to market timing. Larger values of these two measures implie better forecast performance. These four measures are also used in Hong and Lee (2003).

According to the RW model, the future exchange rates are unpredictable using publicly available information, that is,

$$\Delta s_{t,k} = s_{t+k} - s_t = \mu + \varepsilon_{t+k}.^{17} \tag{6}$$

For the RW model, the forecast at time $R+t$ is then obtained by:

$$\Delta \hat{s}_{R+t,k} = \hat{\mu},$$

where $\hat{\mu}$ is estimated from the in-sample observations. For the FP model, the forecast at time $R+t$ is obtained by

$$\Delta \hat{s}_{R+t,k} = \hat{\alpha} + \hat{\beta}(f_{R+t,k} - s_{R+t}),$$

where $\hat{\alpha}, \hat{\beta}$ are estimated from the in-sample observations, $(f_{R+t,k} - s_{R+t})$ is the forward premium observed at time $R+t$. To test the equivalence of predictive accuracy, we calculate the Diebold-Mariano (DM) statistic as:

$$DM = \sqrt{P}\frac{\bar{d}}{\sqrt{\hat{\Omega}_d}},$$

where $\bar{d}$ is the loss differential between the two models with respect to the four performance measures, that is,

$$\bar{d} = -(MSFE^1 - MSFE^0),$$
$$\bar{d} = -(MAFE^1 - MAFE^0),$$
$$\bar{d} = MFTR^1 - MFTR^0,$$
$$\bar{d} = MCFD^1 - MCFD^0,$$

[17] Actually, this is called the random walk with drift, different from the original version which restricts $\mu = 0$. We use the random walk with drift here, because the two economic measures of forecasting performance, MFTR and MCFD, are not directly applicable to the random walk without drift. Nonetheless, in the next section we also compare the forward premium model with the random walk without drift and find similar results.

and $\hat{\Omega}_d$ is the Newey-West HAC variance estimator of $\bar{d}$ with lag truncation parameter $k-1$. The superscript ^{0} and ^{1} denote the RW model and the FP model, respectively.[18] Under the null of equal forecast performance, we have $DM \rightarrow_d N(0,1)$ as $P \rightarrow \infty$ (Diebold and Mariano, 1995).

For the convenience of a visual check, we report the results in Figure 2.[19] In this figure, for each currency, we report the MSFE-, MAFE-, MCFD-ratio, and the MFTR-difference of the FP model relative to the RW model over different horizons. If the MSFE (MAFE) ratio is smaller than 1, or the MCFD ratio is greater than 1, or the MFTR difference is greater than 0, then a better performance of the FP model compared to the RW model is implied. Meanwhile, we use solid diamonds to denote those ratios (or differences) at a 15% significance level, and diamonds to denote the insignificant ratios (or differences).

Some patterns can be observed from Figure 2. First, for most of the currencies (except for JY), the FP model performs at least as well as the RW model over all the horizons. The outcome is recognized by comparing the ratio or the difference of forecast measures to the benchmark value (1 or 0). For JY, the RW model slightly outperforms the FP model at the 3-month horizon in terms of the MAFE ratio, and at the 1-, 2-, 3-week and 1-month horizons in terms of the MCFD ratio. Second, the longer the horizon is, the better the FP model performs relative to the RW model. This result can be seen from the downward slope of the MSFE- and MAFE-ratios, or the upward slope of MFTR differences and MCFD ratios. Generally, for horizons shorter than one month, both models have very similar forecast performance. When the horizon is beyond one month, however, the FP model outperforms the RW model in terms of the ratios/differences, and the superiority increases as the horizon is lengthened. The only exception is DM, where the FP model loses its superiority in terms of MCFD ratio at the 1-year horizon.

For a numerical illustration, Table 4 reports the average (of all seven currencies) MSFE-, MAFE-, MFTR-ratios, and the MCFD-differences of the FP model, relative to the RW model. The average value of the MSFE (MAFE) ratios is 1 for horizons within one month, then decreases all the way to 0.42 and 0.62, respectively, at the 1-year horizon. In terms of MFTR, the FP model shows no superiority over the RW model from one day to one month (the average MFTR difference = 0). Beyond one month, however, it generates 1% more profit at the 2-month and 3-month, 2% more profit at the 6-month, and 4% more

[18] To comply with the convention in the literature, we use the RW model as the benchmark model.

[19] The results in table form are available at request. Also, to check the robustness of the results, we vary the in/out sample ratio (R/P) from 2 to 1. The results are qualitatively similar.

profit at the 1-year horizons. In terms of correctly predicting the directions of changes in exchange rates (MCFD), the FP model does 1% less than the RW model at the 1-day horizon, but 3% more at the 1-week, 2-week and 3-week horizons, 10% more at the 1-month horizon, and, remarkably, 37% more at the 1-year horizon.

It is apprarent from the DM test that these ratios/differences are mostly significant, which can be noted from the number of the solid diamonds in Figure 2. However, we should notice that the RW model is nested in the FP model with restriction $\beta = 0$. According to Clark and McCracken (2001), the DM test of the equal MSFE of two nested models may have a non-standard limiting distribution unless $P/R \to 0$ as $T \to \infty$. Since in our case, P/R (1 or 1/2) is not negligible as $T \to \infty$, statistical inference based on the standard normal critical value may not be reliable. Therefore, we also compute the Chao, Corradi, and Swanson (2001) test statistic, which is designed for the nested case, for the null hypothesis of equal predictive ability,

$$CCS = P\bar{m}\hat{\Omega}_m\bar{m},$$

where $\bar{m} = P^{-1}\sum_{t=1}^{P}\hat{e}^0_{R+t,k}(f_{R+t,k} - s_{R+t})$, $\hat{e}^0_{R+t,k}$ is the forecast error of the RW model at time $R+t$, and $\hat{\Omega}_m$ is the HAC covariance estimator of $\bar{m}$ with lag truncation parameter $k-1$. Under the null hypothesis of equal MSFE, the test follows χ^2 distribution with 1 degree of freedom. As seen from plate 5 of Figure 2, although the statistical inferences may sometimes be different from those of the DM tests, the results are not qualitatively different.

Robustness checks

According to the results in Section 5, the FP model performs at least as well as the RW model in the out-of-sample forecast. More importantly, when the horizon is lengthened beyond one month, the former shows a systematic superiority over the latter. This result seems to be at odds with the common belief in the literature, that the RW model is unlikely to be outpredicted by either structural or statistical models. Therefore, in this section, we check the robustness of the FP model's out-of-sample superiority by comparing it with eight previously examined linear and nonlinear time series exchange rate models.[20] Specifically, the models to be examined for each horizon k are as follows:[21]

FP: forward premium model (benchmark model) $\Delta s_{t,k} = \alpha + \beta(f_{t,k} - s_t) + \varepsilon_{t+k}$;

L1: random walk model without drift $\Delta s_{t,k} = \varepsilon_{t+k}$;

L2: random walk model with drift $\Delta s_{t,k} = \mu + \varepsilon_{t+k}$;

L3: linear $AR(d)$ model $\Delta s_{t,k} = \beta_0 + \sum_{j=1}^{d} \beta_j \Delta s_{t-j\cdot k,k} + \varepsilon_{t+k}$;

N1: autoregressive polynomial model $\Delta s_{t,k} = \beta_0 + \sum_{j=1}^{d} \sum_{i=1}^{m} \alpha_{i,j} \Delta s_{t-j\cdot k,k}^{i} + \varepsilon_{t+k}$;

N2: functional coefficient model $\Delta s_{t,k} = a_0(U_t) + \sum_{j=1}^{d} a_j(U_t) \Delta s_{t-j\cdot k,k}$, where $U_t = S_{t-k} - L^{-1} \sum_{j=1}^{L} S_{t-j\cdot k}$, S denotes exchange rates in level;[22]

N3: the combined forecast of (L3-N2) $\Delta \hat{s}_{t,k} = \sum_{n=1}^{3} w_{nt} \Delta \hat{s}_{t,k}^{(n)}$, where $\Delta \hat{s}_{t,k}^{(n)}$ is the forecast by model m (one of L3-N2), and the weight:

$$w_{nt} = \frac{\exp[-\lambda_t \sum_{j=1}^{t-1} (\Delta s_{j,k} - \Delta \hat{s}_{j,k}^{(n)})^2]}{\sum_{n'=1}^{3} \exp[-\lambda_t \sum_{j=1}^{t-1} (\Delta s_{j,k} - \Delta \hat{s}_{j,k}^{(n')})^2]};$$

In addition to the above linear and nonlinear models, we also consider the following two models in forecasting the direction of exchange returns, which are used in practical trading:

N4: moving-average technical trading rule $sign(\Delta \hat{s}_{t,k}) = sign(U_{t+k})$, where U is defined in N2, $sign(x) = 1$ if $x > 0$, -1 if $x < 0$;

N5: buy & hold rule $sign(\Delta \hat{s}_{t,k}) = 1$ for all t.

The results in Section 5 are based on our daily, overlapping observations. To ensure that the results are robust to different frequencies, in this section, we use the constructed non-overlapping returns for each horizon.[23] Because of the limited number of non-overlapping observations over long horizons, we only examine horizons up to 1-month, which are less frequently examined in the literature. For each horizon, we calculate the MSFE, the MAFE, the MFTR and the MCFD of each competing model, and their ratios/differences relative to those of the FP model, as described in Section 5. We then test the following two null hypotheses:

[20] We do not consider the structural models here, because these models are found not to fit the data well, and generally are outperformed by the RW model (e.g. Meese and Rogoff, 1983a, b; Cheung, et al., 2003).

[21] Hong and Lee (2003) give a detailed discussion of the validity and estimation of these models.

[22] Following Hong and Lee (2003), we choose $d = 2$, $m = 5$, $L = 26$.

[23] The construction of non-overlapping observations has been discussed in Section 4, except that we now use only one sample path for simplicity. The results are similar for different sample paths.

H_0^1 : Model n is no better in the forecast than the benchmark model (FP);

H_0^2 : The best of the first n alternative models is no better in the forecast than the benchmark model (FP).

Note that the test of H_0^1 is similar to the binary comparison in Section 5. The second null involves comparisons among several models, which may lead to a 'data-snooping bias' if we use an individual test sequentially (White, 2000; Hong and Lee, 2003). To avoid the bias, we therefore use the White's reality check method, where the p-values are based on the bootstrap procedure.

The results are reported in Table 5.[24] Although specific rules may vary across currencies, some general patterns can be observed. First, consistent with the results in Section 5, according to the values of the MSFE (MAFE, MCFD) ratio and the MFTR difference, the forecast performance of the FP model relative to the competing models improves as the horizon lengthens. At short horizons such as 1 day or 1 week, some competing models may perform better than the FP model (that is, the MSFE (MAFE) ratio < 1, or the MFTR difference > 0, or the MCFD ratio>1).[25] At the horizon of 1 month, however, the FP dominates most of the alternatives. Take AD as an example. At the 1-day horizon, four of the alternative models, L1, L2, L3 and N3, outperform the FP model in terms of all measures. Meanwhile, all competing models demonstrate better performance than the FP model when economic measures (MFTR, MCFD) are considered. At the 1-month horizon, however, only model N5 beats the FP model, and the superiority is only in terms of economic measures (MFTR, MCFD).

Second, with respect to the statistical inference from White's reality check p-value P_{RC}^1, the null hypothesis H_0^1 is more easily rejected at short horizons than at long horizons. For example, at the 1-day horizon, the number of rejections of H_0^1 at the 15% level is 36 (9 for AD, 11 for BP, 2 for CD, 2 for JY, 12 for SF). At the 1-month horizon, the number decreases to 9 (2 for AD, 1 for BP, 0 for CD, 0 for JY, 6 for SF).

Third, based on P_{RC}^2, the null hypothesis H_0^2 is hardly rejected over all horizons. Equivalently, the best of the eight alternative models is unlikely to beat the FP model over all horizons. The exceptions include AD at the 3-week horizon (MFTR and MCFD), BP at the 1-day (MFTR) and 3-week horizons

[24] To save space, the results for R/P=2 are reported. The results for R/P=1 are similar and available upon request.

[25] Since the FP model is now the benchmark model, larger values of the first two measures, or smaller values of the last two measures, indicate better performance of the FP model.

(MFTR, MCFD), and SF at the 1-day (MCFD) and 3-week horizon (MFTR). It is easily seen that all these exceptions apply to economic measures only.[26]

In summary, the reality check results generally confirm the evidence in Section 5, that is, the out-of-sample forecast performance of the FP model is enhanced as the horizon lengthens. In addition, the results show that the FP model is unlikely to be outperformed by available statistical models of exchange rates, especially in terms of the statistical measures (MSFE and MAFE).

Conclusion

This paper checks the sensitivity of the forward premium puzzle to the time horizon, using the last decade of daily data covering nine horizons (ranging from 1 day to 1 year). The evidence shows that the forward premium model is sensitive to the time horizon in a systematic way. In terms of in-sample estimation, the point estimate of beta generally decreases with the horizon, while the adjusted R^2 goes in the opposite direction. These patterns are robust to most of the currencies examined and to both overlapping and non-overlapping cases. In the out-of-sample forecast, the performance of the forward premium model (with unrestricted coefficients) improves as the forecast horizon lengthens, in terms of both statistical and economic measures. More importantly, the random walk model, which is commonly believed to be unbeatable by either statistical or structural models, seems to be dominated by the forward premium model when the forecast horizon is longer than one month. The out-of-sample forecast superiority of the forward premium model and its horizon-sensitivity are further corroborated in the multiple comparisons between the forward premium model and eight linear/nonlinear time series models, using White's reality check procedure.

The forward premium puzzle remains unsolved, as the beta coefficient is found to be significantly negative over various horizons. However, the evidence here supports our view that the contamination of the unbiasedness hypothesis is sensitive to the horizon, as predicted by the equilibrium condition derived from a consumption-based asset pricing model. Meanwhile, it helps us to understand the interaction between the forward premium and the omitted variables (such as Jensen's Inequality Terms, and the risk

[26] It is possible that the insignificance P^2_{RC} may be due to low power of the reality check procedure in the finite samples, which remains to be investigated.

premiums) in determining future spot exchange rates. For example, as discussed in Fama (1984) and Engel (1996), the negative estimate of beta may be caused by a sufficiently large negative covariation between the forward premium and the omitted variables. If this is true, then the fact that the point estimate of beta decreases with the horizon indicates that the forward premium and the omitted variables are more negatively covariated as the horizon is lengthened. This relation could serve as a guideline for modeling the omitted variables. In addition, the in-sample inference might be reconciled with the out-of-sample forecasts. As found by Mark (1995), the fundamentals have more power in forecasting exchange rates at longer horizons. If this is the case, the increasing forecast power of the forward premiums over the time horizon may be due to two factors: the increasing (negative) covariation between the forward premiums and the omitted variables (fundamentals), and the increasing forecast power of the omitted variables (fundamentals). We leave formal investigation on this issue for future work.

Another possible extension is to examine the puzzle at longer horizons. Due to the difficulty in obtaining forward premiums (or interest rate differentials) with long-horizon maturities, the empirical results in this paper are restricted to horizons within one year. However, it would be interesting to extend our analysis to long-horizon data, as some recent evidence shows that the estimate of beta is positive and close to one at long horizons, such as 5 years and 10 years (Chinn and Meredith 2004). Although more work is needed to check the robustness of our findings over long horizons, we hope that this paper helps us understand more about the anomaly, and points us in the right direction toward explaining the puzzle.

REFERENCES

Baillie, R.T., Bollerslev, T., 1994. The long memory of the forward premium. Journal of International Money and Finance 13, 565–571.

Baillie, R.T., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. Journal of International Money and Finance 19, 471–488.

Bekaert, G., Hodrick, R.J., 1993. On biases in the measurement of foreign exchange risk premiums. Journal of International Money and Finance 12, 115–138.

Chaboud, A.P., Wright J.H., 2005. Uncovered interest parity: it works, but not for long. Journal of International Economics, 66, 349-362.

Chao, J., Corradi, V., Swanson, N.R., 2001. An out of sample test for Granger causality. Macroeconomic Dynamics 5, 598–620.

Cheung, Y., Chinn M., Pascual, A.G., 2003. Empirical exchange rate models of the nineties: Are any fit to survive? Working paper, Santa Cruz Center for International Economics.

Chinn, M., Meredith, G., 2004. Monetary policy and long-horizon uncovered interest parity. IMF Staff Papers 51, 409–430.

Clarida, R.H., Taylor, M.P., 1997. The term structure of forward exchange premiums and the forecastability of spot exchange rates: Correcting the errors. The Review of Economics and Statistics 3, 353–361.

Clarida, R.H., Sarno, L., Taylor, M.P., Valente, G., 2003. The out-of-sample success of term structure models as exchange rate predictors: A step beyond. Journal of International Economics 60, 61–83.

Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85–110.

Cornell, B., 1989. The impact of data errors on measurement of the foreign exchange risk premium. Journal of International Money and Finance 8, 147–157.

Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. Journal of Business and Economic Statistics 13, 134–144.

Engel, C., 1996. The forward discount anomaly and the risk premium: A survey of recent evidence. Journal of Empirical Finance 3, 123–191.

Evans, M.D., 1995. Peso problems: Their theoretical and empirical implications, in: Rao, C.R., Maddala, G.S. (eds.), Handbook of Statistics: Statistical Methods in Finance, Vol. 14. North-Holland, Amsterdam, 357–396.

Fama, E., 1984. Forward and spot exchange rates. Journal of Monetary Economics 14, 319–338.

Frankel, J.A., Andrew, K.R., 1994. An empirical characterization of nominal exchange rates, in: Grossman, G., Rogoff, K. (eds.), Handbook of International Economics, Vol. 3. Elsevier, Amsterdam, 1689-1729.

Gourinchas, P., Tornell A., 2003. Exchange rate puzzles and distorted beliefs. Mimeo, Princeton University.

Granger, C.W.J., 1999. Empirical Modeling in Economics: Specification and Evaluation. Cambridge University Press, London.

Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269–1286.

Hodrick, R.J., 1987. The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange markets. Harwood, Chur, Switzerland.

Hong, Y., Lee, T., 2003. Inference on via generalized spectrum and nonlinear time series models. The Review of Economics and Statistics 85, 1048–1062.

Huisman, R., Koedijka, K., Kool, C., Nissen, F., 1998. Extreme support for uncovered interest parity. Journal of International Money and Finance 17, 211–228.

Kim, C.S., Phillips, P.C.B., 1999. Modified log periodogram regression. Mimeo, Yale University.

McCracken, M.W., 2004. Asymptotics for out of sample tests of Granger Causality. Working paper, University of Missouri-Columbia.

Sarno, L., Valente, G., Leon, H., 2004. The forward bias puzzle and nonlinearity in deviations from uncovered interest parity: A new perspective. Working paper, University of Warwick.

Lewis, K., 1995. Puzzles in international financial markets, in: Grossman, G., Rogoff, K. (eds.), Handbook of International Economics, Vol. 3. Elsevier, Amsterdam, 1913-1971.

Lo, A.W., MacKinlay, A.C., 1999. A Non-random Walk down Wall Street. Princeton University Press, Priceton.

Lucas, R.E., 1978. Asset prices in an exchange economy. Econometrica 46, 1429–1446.

Lyons, R.K., 2001. The Microstructure Approach to Exchange Rates. MIT Press, Cambridge and London.

Mark, N.C., 1995. Exchange rates and fundamentals: Evidence on long-horizon predictability. American Economic Review 85, 201–218.

Maynard, A., Phillips, P.C.B., 2001. Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. Journal of Applied Econometrics 16, 671–708.

Meese, R.A., Rogoff, K., 1983a. Empirical exchange rate model of the seventies. Journal of International Economics 14, 3–24.

Meese, R.A., Rogoff, K., 1983b. The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? in: Frenkel, J.A. (Ed.), Exchange Rate and International Economics, University of Chicago Press, Chicago.

Neely, C.J., Sarno, L., 2002. How well do monetary fundamentals forecast exchange rates? Working paper, Federal Reserve Bank of St. Louis.

Obstfeld, M., Rogoff, K., 1996. Foundations of International Macroeconomics. MIT press, Cambridge.

Sarno, L., Taylor, M.P., 2002. The Economics of Exchange Rates. Cambridge University Press, Cambridge.

Walmsley, J., 2000. The Foreign Exchange and Money Markets Guide (Second Edition). John Wiley, New York.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

TABLE 1. Unit Root and Long Memory Results

currency	horizon	unit root in returns[2]	d of returns	ci_low	ci_up[2]	unit root in forward premiums	d of forward premiums	ci_low	ci_up
AD	1 day	***	(0.04)	(0.11)	0.03	***	0.05	(0.02)	0.12
	1 week	***	0.62	0.55	0.69	***	0.23	0.16	0.30
	2 week	***	0.66	0.59	0.73	***	0.23	0.16	0.30
	3 week	***	0.87	0.80	0.94	***	0.36	0.29	0.43
	1 month	***	0.91	0.84	0.98	***	0.43	0.36	0.50
	2 month	***	0.92	0.85	0.99	***	0.53	0.46	0.60
	3 month	***	0.87	0.80	0.94	*	0.64	0.57	0.71
	6 month		0.98	0.91	1.05		0.72	0.65	0.79
	1 year		0.95	0.88	1.02		0.92	0.85	0.99
BP	1 day	***	0.16	0.09	0.23	***	(0.07)	(0.14)	0.00
	1 week	***	0.60	0.53	0.67	***	0.19	0.12	0.26
	2 week	***	0.73	0.66	0.80	***	0.13	0.06	0.20
	3 week	***	0.90	0.83	0.97	***	0.54	0.47	0.61
	1 month	***	0.91	0.84	0.98	***	0.31	0.24	0.38
	2 month	***	0.95	0.88	1.03	***	0.68	0.61	0.75
	3 month	***	0.96	0.89	1.03		0.91	0.84	0.98
	6 month	**	0.98	0.91	1.05		1.04	0.97	1.11
	1 year		1.00	0.93	1.07		1.04	0.97	1.11
CD	1 day	***	0.24	0.17	0.31	***	0.30	0.23	0.37
	1 week	***	0.62	0.55	0.69	***	0.23	0.16	0.30
	2 week	***	0.68	0.61	0.75	***	0.28	0.21	0.35
	3 week	***	0.86	0.79	0.93	***	0.20	0.13	0.27
	1 month	***	0.94	0.87	1.01	***	0.66	0.59	0.73
	2 month	***	0.97	0.90	1.04	**	0.63	0.56	0.70
	3 month	***	1.00	0.93	1.07		0.96	0.89	1.03
	6 month		0.97	0.90	1.04		0.95	0.88	1.02
	1 year		1.00	0.93	1.07		0.99	0.92	1.06
DM	1 day	***	0.10	0.01	0.19	***	(0.23)	(0.32)	(0.14)
	1 week	***	0.83	0.74	0.92	***	0.34	0.25	0.43
	2 week	***	0.81	0.73	0.90	***	0.32	0.23	0.41
	3 week	***	0.84	0.75	0.93	***	0.46	0.37	0.55
	1 month	***	0.96	0.87	1.04	***	0.62	0.53	0.71
	2 month	**	0.97	0.88	1.06	***	0.64	0.55	0.73
	3 month		1.03	0.94	1.12	*	0.91	0.82	1.00
	6 month		1.00	0.91	1.09	***	0.86	0.77	0.95
	1 year		1.03	0.94	1.12		0.92	0.83	1.01
EURO	1 day	***	0.11	0.02	0.21	***	0.13	0.04	0.23
	1 week	***	0.76	0.67	0.86	***	0.35	0.26	0.45
	2 week	***	0.86	0.77	0.96	***	0.31	0.21	0.40
	3 week	***	0.85	0.75	0.94	***	0.19	0.09	0.28
	1 month	***	0.93	0.83	1.02		0.88	0.78	0.98
	2 month	**	0.88	0.78	0.97		0.96	0.86	1.05
	3 month	**	0.95	0.85	1.04		0.96	0.86	1.05
	6 month	*	0.96	0.86	1.05		0.99	0.89	1.08
	1 year		0.99	0.89	1.09		1.00	0.91	1.10
JY	1 day	***	0.06	(0.01)	0.13	***	0.07	0.00	0.14
	1 week	***	0.60	0.53	0.67	***	0.19	0.12	0.26
	2 week	***	0.69	0.62	0.76	***	0.21	0.14	0.28
	3 week	***	0.84	0.77	0.91	***	0.81	0.74	0.88
	1 month	***	0.90	0.83	0.97	**	0.60	0.53	0.67
	2 month	***	0.91	0.84	0.98		0.77	0.70	0.84
	3 month	**	0.94	0.87	1.01		0.89	0.82	0.96
	6 month		0.97	0.90	1.04		0.96	0.88	1.03
	1 year		0.95	0.88	1.02		0.94	0.87	1.01
SF	1 day	***	0.08	0.01	0.15	***	(0.01)	(0.08)	0.06
	1 week	***	0.59	0.52	0.66	***	0.20	0.13	0.27
	2 week	***	0.67	0.60	0.74	***	0.17	0.10	0.24
	3 week	***	0.86	0.79	0.93	***	0.14	0.07	0.21
	1 month	***	0.92	0.85	0.99	***	0.32	0.25	0.39
	2 month	***	0.93	0.86	1.00		0.70	0.63	0.77
	3 month	***	0.95	0.88	1.02		0.77	0.70	0.84
	6 month	*	0.97	0.90	1.04		0.83	0.76	0.90
	1 year		0.97	0.90	1.05		0.95	0.88	1.02

Note:

1 The unit root results are based on the Philips-Perron test statistics.
*, **, *** denote the rejection of unit root at 10%, 5% and 1%, respectively.

2 The memory parameter d is estimated from the modified log periodogram procedure by Kim and Phillips (1999).
ci_low and ci_up are the lower and upper bounds of the 95% confidence interval.

TABLE 2. In-Sample Estimation

regression equation: $\Delta s_{t,k} = \alpha + \beta(f_{t,k} - s_t) + \varepsilon_{t+k}$					sample frequency: daily				
currency	horizon[2]	$\hat{\alpha}$[3]	$\hat{\beta}$	$\overline{R}^2$	currency	horizon	$\hat{\alpha}$	$\hat{\beta}$	$\overline{R}^2$
AD	1 day	-0.0001	(0.01)	(0.00)	BP	1 day	0.00	0.18	0.00
		(0.0001)	(0.0861)				(0.0001)	(0.1306)	
	1 week	-0.0006	(1.43)	0.00		1 week	0.00	(0.72)	0.00
		(0.0007)	(0.4244)				(0.0005)	(0.6304)	
	2 week	-0.0011	(1.46)	0.01		2 week	0.00	0.08	(0.00)
		(0.0012)	(0.5114)				(0.0009)	(0.2973)	
	3 week	-0.0018	(2.16)	0.02		3 week	0.00	0.43	0.00
		(0.0017)	(0.721)				(0.0013)	(0.8965)	
	1 month	-0.0052	(5.91)	0.06		1 month	(0.00)	(1.32)	0.00
		(0.003)	(1.8275)				(0.002)	(1.292)	
	2 month	-0.0106	(6.42)	0.13		2 month	(0.00)	(2.96)	0.03
		(0.0051)	(1.7152)				(0.004)	(1.7)	
	3 month	-0.0164	(6.96)	0.23		3 month	(0.01)	(3.31)	0.06
		(0.0066)	(1.629)				(0.0057)	(1.6372)	
	6 month	-0.0325	(6.96)	0.44		6 month	(0.01)	(3.21)	0.12
		(0.0125)	(1.3489)				(0.0104)	(1.4415)	
	1 year	-0.0629	(7.10)	0.55		1 year	(0.02)	(4.13)	0.30
		(0.0266)	(1.4439)				(0.0161)	(1.2913)	
CD	1 day	0.0000	0.02	0.00	DM	1 day	0.00	0.13	0.00
		(0.0001)	(0.0301)				(0.0002)	(0.1389)	
	1 week	-0.0003	(0.42)	0.00		1 week	(0.00)	(0.35)	(0.00)
		(0.0004)	(0.4477)				(0.0011)	(0.876)	
	2 week	-0.0005	(0.18)	0.00		2 week	(0.00)	(0.55)	0.00
		(0.0006)	(0.224)				(0.0018)	(0.8894)	
	3 week	-0.0006	(0.54)	0.00		3 week	0.00	(2.62)	0.02
		(0.0008)	(0.3862)				(0.0027)	(1.8423)	
	1 month	-0.0004	(2.63)	0.03		1 month	0.01	(6.32)	0.08
		(0.0012)	(1.0972)				(0.0036)	(1.8754)	
	2 month	0.0000	(3.32)	0.09		2 month	0.01	(6.52)	0.15
		(0.0024)	(1.1624)				(0.0061)	(1.6034)	
	3 month	0.0008	(3.66)	0.15		3 month	0.02	(7.70)	0.27
		(0.0036)	(1.2725)				(0.0089)	(1.7134)	
	6 month	0.0014	(3.44)	0.25		6 month	0.05	(8.20)	0.50
		(0.0064)	(1.1977)				(0.0133)	(1.4144)	
	1 year	0.0040	(3.94)	0.47		1 year	0.10	(7.46)	0.53
		(0.0114)	(1.1874)				(0.0213)	(1.3604)	
EURO	1 day	-0.0001	(0.98)	(0.00)	JY	1 day	0.00	(0.03)	(0.00)
		(0.0002)	(1.2366)				(0.0002)	(0.147)	
	1 week	0.0003	(7.77)	0.03		1 week	0.00	(3.22)	0.01
		(0.001)	(2.9162)				(0.0012)	(1.1735)	
	2 week	0.0006	(6.29)	0.05		2 week	0.00	(1.60)	0.00
		(0.0017)	(2.0724)				(0.0018)	(0.7058)	
	3 week	0.0003	(4.54)	0.04		3 week	0.00	(0.32)	0.00
		(0.0025)	(1.9668)				(0.0024)	(0.6768)	
	1 month	0.0020	(6.79)	0.11		1 month	0.01	(3.75)	0.02
		(0.0035)	(2.2122)				(0.0064)	(1.7583)	
	2 month	0.0061	(7.06)	0.20		2 month	0.02	(3.29)	0.03
		(0.0068)	(1.9728)				(0.0125)	(1.7031)	
	3 month	0.0103	(7.10)	0.30		3 month	0.03	(3.24)	0.04
		(0.0089)	(1.7236)				(0.0178)	(1.659)	
	6 month	0.0242	(7.03)	0.63		6 month	0.07	(3.43)	0.08
		(0.0096)	(0.9598)				(0.0236)	(1.2576)	
	1 year	0.0699	(7.41)	0.82		1 year	0.16	(3.82)	0.21
		(0.0085)	(0.5471)				(0.0283)	(0.9233)	
SF	1 day	0.0000	0.09	(0.00)	SF	2 month	0.03	(6.01)	0.12
		(0.0001)	(0.152)				(0.0078)	(1.4967)	
	1 week	0.0008	(1.26)	0.00		3 month	0.04	(6.04)	0.17
		(0.0008)	(0.6122)				(0.0109)	(1.4486)	
	2 week	0.0024	(2.17)	0.01		6 month	0.08	(6.05)	0.35
		(0.0015)	(0.8535)				(0.0137)	(1.1601)	
	3 week	0.0028	(1.59)	0.01		1 year	0.17	(6.21)	0.54
		(0.002)	(0.6666)				(0.0185)	(1.0231)	
	1 month	0.0099	-4.2213	0.0416					
		(0.0036)	(1.3249)						

Note:
1 k=1, 7, 12, 17, 24, 46, 68, 134, 263 for 1 day, 1 week, 2 week, 3 week, 1 month, 2 month, 3 month, 6 month, 1 year.
2 The numbers in parentheses are White's standard errors for 1 day, and are Newey-West HAC standard errors for other horizons.

TABLE 3. Robustness of the In-sample Results to Sampling Frequencies

variable	25%	50%	75%	Mean	Std. Dev.	# of Obs.
			AD			
Beta1w	-2.55725	-2.16514	-0.88116	-1.71004	0.88419	7.00000
Beta2w	-3.49217	-2.13305	-1.10248	-1.93031	1.33152	12.00000
Beta3w	-3.65932	-2.84106	-2.19357	-2.91556	1.67007	17.00000
Beta1m	-6.87474	-6.39427	-5.84114	-6.24663	0.84216	24.00000
Beta2m	-7.54493	-7.01190	-6.43507	-6.93941	0.78259	46.00000
Beta3m	-7.71840	-7.40019	-6.99710	-7.29997	0.69266	68.00000
Beta6m	-7.90041	-7.59474	-7.20766	-7.53302	0.54774	134.00000
Beta1y	-8.26264	-7.67396	-6.88814	-7.67004	0.81889	263.00000
			BP			
Beta1w	-2.11978	-1.54900	-0.22255	-1.13561	0.99866	7.00000
Beta2w	-1.01147	-0.19430	0.45663	-0.01287	1.02439	12.00000
Beta3w	-0.73570	0.35829	0.83218	-0.18266	1.67253	17.00000
Beta1m	-2.60680	-2.19025	-1.25458	-1.81837	1.13246	24.00000
Beta2m	-4.06814	-3.43346	-2.90159	-3.46955	0.92657	46.00000
Beta3m	-3.91583	-3.57147	-3.03184	-3.45702	0.65915	68.00000
Beta6m	-4.67755	-3.75844	-3.14270	-3.85129	0.94529	134.00000
Beta1y	-6.89472	-5.94115	-4.25824	-5.65771	1.62713	263.00000
			CD			
Beta1w	-1.67495	-0.91657	-0.39487	-0.79227	0.88360	7.00000
Beta2w	-0.40733	-0.27167	-0.13704	-0.24478	0.27168	12.00000
Beta3w	-1.48175	-0.68173	-0.48054	-0.85438	0.63001	17.00000
Beta1m	-3.33504	-2.83739	-2.57188	-2.88894	0.55161	24.00000
Beta2m	-4.23606	-3.94932	-3.63200	-3.88646	0.54820	46.00000
Beta3m	-4.27671	-4.01703	-3.79887	-4.01552	0.47627	68.00000
Beta6m	-4.71733	-4.39958	-4.07278	-4.41963	0.44000	134.00000
Beta1y	-4.71573	-3.98812	-3.22654	-3.96659	0.79305	263.00000
			DM			
Beta1w	-2.35361	-1.34865	-0.15490	-0.75619	1.70347	7.00000
Beta2w	-1.71195	-0.33528	-0.01132	-0.88909	1.89881	12.00000
Beta3w	-5.46191	-4.09579	-1.96439	-3.32390	2.50205	17.00000
Beta1m	-6.33574	-5.72320	-5.21434	-5.47601	1.37563	24.00000
Beta2m	-7.11720	-6.10140	-5.04518	-5.96243	1.53515	46.00000
Beta3m	-7.43857	-6.93739	-6.55800	-7.08493	1.09548	68.00000
Beta6m	-8.92230	-7.70104	-6.98575	-7.97198	1.44999	134.00000
Beta1y	-9.55183	-8.19154	-6.87533	-9.39819	4.40824	263.00000

Note: This table checks the robustness of the in-sample results to the sampling frequencies.

TABLE 3(Continued):Robustness of the In-sample Results to Sampling Frequencies

variable	25%	50%	75%	Mean	Std. Dev.	# of Obs.
			EURO			
Beta1w	-8.84948	-8.39182	-7.79200	-7.93181	1.22921	7.00000
Beta2w	-7.13930	-6.91725	-6.70321	-6.61747	0.78204	12.00000
Beta3w	-7.01391	-6.36216	-5.37841	-5.60285	1.92867	17.00000
Beta1m	-7.67028	-7.35292	-7.20283	-7.41608	0.28441	24.00000
Beta2m	-7.98404	-7.80835	-7.56079	-7.79353	0.28046	46.00000
Beta3m	-7.52949	-7.19271	-7.01226	-7.25309	0.34413	68.00000
Beta6m	-7.80989	-7.09551	-6.68607	-7.17425	0.69709	134.00000
Beta1y	-7.68831	-7.20805	-6.70701	-7.19176	0.70949	263.00000
			JY			
Beta1w	-4.11838	-4.01682	-3.56984	-3.56691	0.90667	7.00000
Beta2w	-3.76203	-2.28167	-1.39654	-2.24384	1.56640	12.00000
Beta3w	-0.82192	-0.56894	-0.02512	-0.63283	0.92853	17.00000
Beta1m	-4.07493	-3.70878	-3.56448	-3.80717	0.40439	24.00000
Beta2m	-3.58319	-3.34975	-3.19187	-3.32414	0.35323	46.00000
Beta3m	-3.70042	-3.27111	-2.69278	-3.22284	0.67146	68.00000
Beta6m	-4.32758	-3.99471	-3.64949	-3.97793	0.46568	134.00000
Beta1y	-4.56363	-3.90920	-3.39810	-4.02291	0.77990	263.00000
			SF			
Beta1w	-2.10787	-1.49215	-1.21270	-1.26680	1.39011	7.00000
Beta2w	-5.33936	-2.92432	-1.24282	-2.89063	1.85095	12.00000
Beta3w	-3.02664	-2.05391	-0.84658	-2.10637	1.49651	17.00000
Beta1m	-5.84427	-5.42790	-4.10292	-4.66462	1.54208	24.00000
Beta2m	-6.38939	-6.22948	-5.98214	-6.12326	0.40024	46.00000
Beta3m	-6.24121	-5.96789	-5.83588	-5.97193	0.35247	68.00000
Beta6m	-6.86727	-5.77159	-5.24632	-5.96491	0.86926	134.00000
Beta1y	-6.51549	-5.95383	-5.40189	-5.96224	0.70452	263.00000

TABLE 4. Out-of-Sample Comparison between the FP Model and the RW Model

measure	1d	1w	2w	3w	1m	2m	3m	6m	1y
MSFE ratio	1.00	1.00	1.00	1.00	0.97	0.90	0.82	0.63	0.42
MAFE ratio	1.00	1.00	1.00	1.00	0.98	0.96	0.92	0.79	0.62
MFTR diff.	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.04
MCFD ratio	0.99	1.03	1.03	1.03	1.10	1.12	1.15	1.30	1.37

Note: This table gives the average (of all the seven currencies) MSFE-, MAFE-, MFTR-ratios and MCFD-differences of the FP model relative to the RW model.
The R/P ratio (R and P stand for the number of in-sample and out-of-smaple observations) is 2

TABLE 5. Reality check on predictive ability over different horizons

Panel A. Australian Dollars (AD); R/P = 2

		MSFE				MAFE				MFTR				MCFD			
k	Model	MSFE	Ratio	P^1_{RC}	P^2_{RC}	MAFE	Ratio	P^1_{RC}	P^2_{RC}	MFTR	Diff	P^1_{RC}	P^2_{RC}	MCFD	Ratio	P^1_{RC}	P^2_{RC}
								Horizon = 1 day; (R, P) = (1594,796)									
0	FP	0.502				0.542				-0.002				0.474			
1	L1	0.501	0.997	**0.136**	**0.137**	0.541	0.997	**0.025**	**0.015**	0.000	0.002	0.405	0.438	0.000	0.000	1.000	1.000
2	L2	0.502	0.999	0.236	0.153	0.542	0.999	**0.096**	**0.015**	0.001	0.003	0.298	0.578	0.477	1.008	0.212	0.491
3	L3	0.502	0.999	0.441	0.345	0.541	0.997	0.158	**0.109**	0.012	0.015	0.261	0.418	0.497	1.050	**0.070**	**0.119**
4	N1	0.504	1.004	0.734	0.502	0.542	1.000	0.482	0.258	0.025	0.027	**0.144**	0.309	0.500	1.056	**0.048**	**0.118**
5	N2	0.637	1.269	0.918	0.805	0.562	1.036	0.915	0.698	0.025	0.028	0.208	0.435	0.486	1.027	0.292	0.234
6	N3	0.501	0.997	0.213	0.863	0.541	0.997	0.043	0.764	0.046	0.049	**0.016**	0.155	0.511	1.080	**0.010**	**0.112**
7	N4									0.017	0.019	0.251	0.230	0.491	1.037	0.211	0.181
8	N5									-0.001	0.002	0.434	0.277	0.504	1.064	**0.136**	0.216
								Horizon = 1 week; (R, P) = (340,169)									
0	FP	2.264				1.201				0.039				0.491			
1	L1	2.271	1.003	0.655	0.650	1.205	1.004	0.767	0.789	0.000	-0.039	0.668	0.651	0.000	0.000	1.000	1.000
2	L2	2.306	1.019	0.969	0.766	1.220	1.016	0.997	0.886	-0.194	-0.233	0.932	0.751	0.408	0.831	0.951	0.988
3	L3	2.341	1.034	0.993	0.814	1.230	1.024	0.998	0.905	-0.257	-0.295	0.975	0.782	0.379	0.771	0.984	0.989
4	N1	2.366	1.045	0.993	0.854	1.232	1.026	0.995	0.928	-0.257	-0.296	0.973	0.808	0.426	0.867	0.923	0.988
6	N2	3.111	1.374	0.973	0.941	1.370	1.141	0.989	0.973	-0.144	-0.183	0.918	0.872	0.420	0.855	0.960	0.857
8	N3	2.278	1.006	0.661	0.960	1.209	1.007	0.669	0.983	-0.103	-0.142	0.809	0.881	0.420	0.855	0.932	0.875
10	N4									0.104	0.065	0.372	0.783	0.544	1.108	0.232	0.540
11	N5									0.192	0.153	0.241	0.563	0.574	1.169	0.152	0.322
								Horizon = 2 week; (R, P) = (161,80)									
0	FP	4.928				1.753				0.557				0.638			
1	L1	5.242	1.064	0.897	0.885	1.837	1.048	0.988	0.977	0.000	-0.557	0.925	0.921	0.000	0.000	1.000	1.000
2	L2	5.444	1.105	0.998	0.973	1.867	1.065	0.996	0.991	-0.450	-1.007	0.997	0.990	0.400	0.627	0.997	1.000
3	L3	5.689	1.154	0.999	0.988	1.895	1.081	1.000	0.998	-0.486	-1.043	1.000	0.994	0.438	0.686	0.998	1.000
4	N1	6.034	1.224	0.996	0.989	1.931	1.102	0.998	0.999	-0.605	-1.162	0.957	0.994	0.413	0.647	0.993	1.000
6	N2	10.658	2.163	0.937	0.998	2.289	1.305	0.974	1.000	-0.263	-0.820	0.981	0.997	0.463	0.725	0.961	1.000
8	N3	5.033	1.021	0.637	0.992	1.794	1.023	0.751	0.999	-0.638	-1.195	0.998	0.999	0.400	0.627	0.995	1.000
10	N4									0.238	-0.319	0.911	0.999	0.588	0.922	0.891	0.998
11	N5									0.536	-0.021	0.280	0.932	0.613	0.961	0.380	0.986
								Horizon = 3 week; (R, P) = (101,50)									
0	FP	6.660				2.120				-0.028				0.460			
1	L1	5.915	0.888	**0.090**	**0.078**	1.992	0.940	**0.081**	**0.070**	0.000	0.028	0.347	0.354	0.000	0.000	1.000	1.000
2	L2	6.315	0.948	0.251	**0.107**	2.087	0.984	0.307	**0.095**	-0.340	-0.312	0.888	0.600	0.400	0.870	0.788	0.899
3	L3	7.221	1.084	0.843	0.218	2.285	1.078	0.958	0.328	-0.613	-0.585	0.921	0.675	0.320	0.696	0.921	0.915
4	N1	8.143	1.223	0.968	0.321	2.463	1.162	0.990	0.411	-0.564	-0.536	0.864	0.728	0.320	0.696	0.903	0.933
6	N2	22.872	3.434	0.891	0.735	2.987	1.409	0.916	0.784	0.081	0.109	0.358	0.327	0.520	1.130	0.174	0.211
8	N3	5.904	0.886	**0.130**	0.740	1.987	0.937	0.185	0.784	0.041	0.069	0.437	0.368	0.480	1.043	0.377	0.238
10	N4									0.645	0.673	**0.071**	0.184	0.660	1.435	**0.000**	**0.026**
11	N5									0.726	0.754	**0.050**	**0.142**	0.680	1.478	**0.000**	**0.026**
								Horizon = 1 month; (R, P) = (72,35)									
0	FP	11.500				2.550				0.842				0.714			
1	L1	12.199	1.061	0.736	0.727	2.815	1.104	0.922	0.915	0.000	-0.842	0.973	0.960	0.000	0.000	1.000	1.000
2	L2	12.904	1.122	0.888	0.750	2.961	1.161	0.995	0.942	-0.692	-1.534	0.966	0.968	0.371	0.520	0.984	1.000
3	L3	14.058	1.222	0.976	0.784	3.160	1.239	1.000	0.961	-0.451	-1.293	0.925	0.973	0.371	0.520	0.997	1.000
4	N1	16.637	1.447	0.874	0.843	3.199	1.254	0.908	0.974	-0.086	-0.929	0.845	0.976	0.514	0.720	0.952	1.000
6	N2	37.592	3.269	0.969	0.964	4.429	1.737	0.997	0.993	0.335	-0.507	0.595	0.931	0.457	0.640	0.933	1.000
8	N3	11.979	1.042	0.595	0.923	2.789	1.094	0.871	0.994	0.824	-0.019	0.286	0.753	0.543	0.760	0.791	0.999
10	N4									0.666	-0.177	0.469	0.408	0.571	0.800	0.744	0.771
11	N5									1.122	0.280	**0.062**	0.408	0.743	1.040	**0.062**	0.716

Note:

1. The reality check results are based on non-overlapping observations constructed from the original daily data, that is, weekly data for a one-week horizon, monthly data for a one-month horizon, etc.
2. For the model description, please refer to page 13-14.
3. P^1_{RC} is the bootstrap p-values of White's (2000) test for the null that model n is no better in forecasting than the benchmark (FP) model, while P^2_{RC} is for the null that the best of the first n alternative models is no better in forecasting than the benchmark (FP) model.These p-values are based on 1,000 bootstrap replications and a bootstrap smoothing parameter a = 0.25. The results are similar when different values of a (0.5, 0.75) are used.
4. Redness denotes better performance of the alternative to the FP model in terms of ratio/difference, boldness denotes significance level at 15%.
5. The bootstrap program was generously provided by Tae-Hwy Lee.

TABLE 5 (Continued). Reality check on predictive ability over different horizons

Panel B. British Pound (BP); R/P = 2

		MSFE				MAFE				MFTR				MCFD			
k	Model	MSFE	Ratio	P^1_{RC}	P^2_{RC}	MAFE	Ratio	P^1_{RC}	P^2_{RC}	MFTR	Diff	P^1_{RC}	P^2_{RC}	MCFD	Ratio	P^1_{RC}	P^2_{RC}
Horizon = 1 day; (R, P) = (1594,796)																	
0	FP	0.251				0.379				-0.025				0.486			
1	L1	0.251	0.999	**0.027**	**0.028**	0.379	1.000	**0.141**	**0.144**	0.000	0.025	**0.055**	**0.054**	0.000	0.000	1.000	1.000
2	L2	0.251	1.000	**0.132**	**0.029**	0.379	1.000	0.247	0.179	-0.018	0.007	0.286	**0.077**	0.491	1.010	0.311	0.547
3	L3	0.251	1.000	0.505	0.386	0.380	1.003	0.796	0.504	0.003	0.028	**0.080**	**0.117**	0.497	1.023	0.254	0.432
4	N1	0.252	1.003	0.697	0.539	0.380	1.002	0.739	0.607	0.019	0.045	**0.044**	**0.047**	0.506	1.041	0.193	0.309
5	N2	0.256	1.020	0.852	0.780	0.385	1.015	0.959	0.808	0.015	0.041	**0.025**	**0.072**	0.514	1.057	**0.092**	0.251
6	N3	0.251	0.999	**0.014**	0.835	0.379	0.999	**0.060**	0.854	0.008	0.033	**0.060**	**0.044**	0.501	1.031	0.203	0.276
7	N4									-0.022	0.003	0.415	**0.062**	0.469	0.964	0.721	0.337
8	N5									-0.001	0.025	0.189	**0.082**	0.501	1.031	0.300	0.382
Horizon = 1 week; (R, P) = (340,169)																	
0	FP	1.348				0.923				0.110				0.544			
1	L1	1.362	1.010	0.843	0.835	0.930	1.008	0.849	0.849	0.000	-0.110	0.863	0.858	0.000	0.000	1.000	1.000
2	L2	1.364	1.012	0.892	0.873	0.930	1.008	0.865	0.879	-0.049	-0.159	0.883	0.917	0.515	0.946	0.689	0.825
3	L3	1.381	1.024	0.957	0.897	0.939	1.018	0.979	0.893	-0.100	-0.210	0.947	0.922	0.444	0.815	0.974	0.857
4	N1	1.384	1.027	0.944	0.933	0.935	1.013	0.857	0.917	-0.046	-0.156	0.917	0.936	0.527	0.967	0.639	0.831
6	N2	2.954	2.191	0.921	0.982	1.131	1.225	0.968	0.965	-0.007	-0.117	0.805	0.968	0.503	0.924	0.722	0.874
8	N3	1.359	1.008	0.832	0.984	0.928	1.006	0.853	0.978	0.095	-0.016	0.547	0.852	0.556	1.022	0.424	0.722
10	N4									0.033	-0.078	0.732	0.910	0.527	0.967	0.550	0.779
11	N5									0.130	0.019	0.316	0.836	0.533	0.978	0.580	0.811
Horizon = 2 week; (R, P) = (161,80)																	
0	FP	2.478				1.250				0.208				0.538			
1	L1	2.537	1.024	0.830	0.803	1.267	1.013	0.841	0.811	0.000	-0.208	0.869	0.842	0.000	0.000	1.000	1.000
2	L2	2.549	1.028	0.879	0.835	1.269	1.015	0.905	0.876	-0.054	-0.262	0.883	0.921	0.525	0.977	0.544	0.650
3	L3	2.675	1.079	0.908	0.863	1.267	1.013	0.591	0.884	0.013	-0.195	0.739	0.914	0.600	1.116	0.171	0.218
4	N1	2.765	1.116	0.938	0.898	1.328	1.062	0.827	0.900	-0.425	-0.633	0.969	0.924	0.413	0.767	0.789	0.275
6	N2	5.674	2.289	0.995	0.940	1.723	1.378	0.999	0.967	-0.399	-0.607	0.979	0.946	0.375	0.698	0.927	0.338
8	N3	2.581	1.042	0.688	0.966	1.285	1.028	0.644	0.975	-0.076	-0.284	0.820	0.953	0.513	0.953	0.497	0.382
10	N4									0.086	-0.123	0.815	0.903	0.525	0.977	0.612	0.402
11	N5									0.319	0.111	**0.111**	0.683	0.563	1.047	0.155	0.407
Horizon = 3 week; (R, P) = (101,50)																	
0	FP	4.212				1.703				-0.492				0.440			
1	L1	4.135	0.982	**0.143**	**0.125**	1.689	0.992	0.190	0.174	0.000	0.492	**0.052**	**0.038**	0.000	0.000	1.000	1.000
2	L2	4.205	0.998	0.475	**0.131**	1.703	1.000	0.465	0.189	-0.493	-0.001	0.512	**0.038**	0.400	0.909	0.779	0.885
3	L3	4.439	1.054	0.894	0.347	1.774	1.042	0.910	0.409	-0.481	0.011	0.507	**0.115**	0.380	0.864	0.698	0.904
4	N1	5.004	1.188	0.877	0.544	1.828	1.073	0.827	0.541	-0.420	0.072	0.439	0.152	0.400	0.909	0.541	0.919
6	N2	6.905	1.639	0.972	0.856	2.035	1.195	0.989	0.844	0.127	0.619	**0.039**	**0.096**	0.520	1.182	**0.083**	0.377
8	N3	4.203	0.998	0.425	0.889	1.703	1.000	0.351	0.880	-0.218	0.274	0.256	**0.111**	0.460	1.045	0.389	0.404
10	N4									0.246	0.738	**0.035**	**0.079**	0.560	1.273	**0.077**	0.282
11	N5									0.495	0.987	**0.002**	**0.024**	0.640	1.455	**0.015**	**0.062**
Horizon = 1 month; (R, P) = (72,35)																	
0	FP	5.640				1.760				0.628				0.657			
1	L1	5.866	1.040	0.748	0.784	1.822	1.035	0.836	0.847	0.000	-0.628	0.869	0.878	0.000	0.000	1.000	1.000
2	L2	5.992	1.062	0.887	0.818	1.840	1.045	0.917	0.868	-0.373	-1.001	0.900	0.899	0.429	0.652	0.879	0.985
3	L3	6.459	1.145	0.827	0.838	1.956	1.111	0.878	0.870	-0.208	-0.836	0.773	0.899	0.400	0.609	0.879	0.985
4	N1	7.795	1.382	0.821	0.877	2.058	1.169	0.768	0.899	-0.286	-0.914	0.797	0.901	0.457	0.696	0.778	0.985
6	N2	8.525	1.511	0.777	0.921	2.097	1.191	0.740	0.928	0.186	-0.442	0.598	0.830	0.543	0.826	0.581	0.837
8	N3	5.973	1.059	0.597	0.930	1.836	1.043	0.637	0.934	0.271	-0.356	0.569	0.790	0.486	0.739	0.723	0.839
10	N4									0.575	-0.053	0.397	0.668	0.629	0.957	0.385	0.679
11	N5									0.743	0.115	**0.133**	0.570	0.657	1.000	0.222	0.616

Note:

1 The reality check results are based on non-overlapping observations constructed from the original daily data, that is, weekly data for a one-week horizon, monthly data for a one-month horizon, etc.

2 For the model description, please refer to page 13-14.

3 P^1_{RC} is the bootstrap p-values of White's (2000) test for the null that model n is no better in forecasting than the benchmark (FP) model, while P^2_{RC} is for the null that the best of the first n alternative models is no better in forecasting than the benchmark (FP) model.These p-values are based on 1,000 bootstrap replications and a bootstrap smoothing parameter a = 0.25. The results are similar when different values of a (0.5, 0.75) are used.

4 Redness denotes better performance of the alternative to the FP model in terms of ratio/difference, boldness denotes significance level at 15%.

5 The bootstrap program was generously provided by Tae-Hwy Lee.

TABLE 5 (Continued). Reality check on predictive ability over different horizons

Panel C. Canadian Dollars (CD); R/P = 2

		MSFE				MAFE				MFTR				MCFD			
k	Model	MSFE	Ratio	P^1_{RC}	P^2_{RC}	MAFE	Ratio	P^1_{RC}	P^2_{RC}	MFTR	Diff	P^1_{RC}	P^2_{RC}	MCFD	Ratio	P^1_{RC}	P^2_{RC}
Horizon = 1 day; (R, P) = (1594,796)																	
0	FP	0.128				0.281				-0.003				0.480			
1	L1	0.127	0.998	**0.135**	**0.126**	0.280	0.999	**0.114**	**0.099**	0.000	0.003	0.401	0.403	0.000	0.000	1.000	1.000
2	L2	0.128	0.999	0.217	**0.128**	0.281	1.000	0.508	**0.102**	0.000	0.002	0.341	0.552	0.484	1.008	0.259	0.514
3	L3	0.127	0.998	0.226	0.317	0.281	1.000	0.546	0.276	0.004	0.007	0.297	0.459	0.487	1.016	0.329	0.529
4	N1	0.130	1.020	0.907	0.602	0.283	1.007	0.941	0.531	-0.010	-0.008	0.675	0.555	0.480	1.000	0.484	0.606
5	N2	0.166	1.302	0.953	0.817	0.299	1.065	0.986	0.835	-0.021	-0.018	0.880	0.626	0.470	0.979	0.664	0.408
6	N3	0.128	0.999	0.261	0.845	0.280	0.999	0.342	0.881	0.009	0.011	0.195	0.581	0.485	1.010	0.355	0.446
7	N4									-0.012	-0.010	0.689	0.648	0.480	1.000	0.461	0.505
8	N5									0.000	0.003	0.431	0.655	0.504	1.050	0.193	0.471
Horizon = 1 week; (R, P) = (340,169)																	
0	FP	0.966				0.807				-0.003				0.491			
1	L1	0.958	0.992	**0.122**	**0.103**	0.803	0.994	**0.093**	**0.076**	0.000	0.003	0.505	0.479	0.000	0.000	1.000	1.000
2	L2	0.964	0.998	0.321	**0.105**	0.807	0.999	0.268	**0.076**	-0.068	-0.065	0.885	0.703	0.462	0.940	0.886	0.951
3	L3	0.983	1.017	0.835	0.417	0.821	1.016	0.916	0.329	0.047	0.050	0.317	0.463	0.533	1.084	**0.150**	0.191
4	N1	0.983	1.018	0.711	0.578	0.811	1.005	0.594	0.475	-0.007	-0.003	0.500	0.515	0.485	0.988	0.429	0.233
6	N2	2.520	2.609	0.933	0.807	1.002	1.241	0.956	0.772	0.014	0.018	0.418	0.624	0.509	1.036	0.295	0.355
8	N3	0.955	0.988	**0.036**	0.864	0.804	0.996	0.226	0.839	0.092	0.095	**0.127**	0.399	0.533	1.084	**0.137**	0.392
10	N4									-0.060	-0.056	0.739	0.486	0.462	0.940	0.648	0.463
11	N5									0.069	0.073	0.258	0.506	0.527	1.072	0.232	0.484
Horizon = 2 week; (R, P) = (161,80)																	
0	FP	2.066				1.131				-0.089				0.475			
1	L1	2.026	0.981	**0.034**	**0.040**	1.122	0.991	**0.078**	**0.087**	0.000	0.089	0.190	0.204	0.000	0.000	1.000	1.000
2	L2	2.057	0.996	0.164	**0.040**	1.131	0.999	0.317	**0.087**	-0.179	-0.089	0.953	0.231	0.438	0.921	0.932	0.974
3	L3	2.081	1.007	0.546	**0.081**	1.127	0.996	0.213	0.217	-0.229	-0.140	0.711	0.300	0.425	0.895	0.614	0.981
4	N1	3.857	1.867	0.923	0.586	1.395	1.233	0.940	0.606	-0.185	-0.096	0.578	0.364	0.438	0.921	0.545	0.988
6	N2	33.500	16.21	0.904	0.814	2.016	1.782	0.902	0.799	-0.101	-0.011	0.535	0.546	0.475	1.000	0.479	0.841
8	N3	2.035	0.985	**0.087**	0.869	1.126	0.995	0.204	0.786	0.117	0.206	**0.063**	0.158	0.513	1.079	**0.075**	0.439
10	N4									0.080	0.170	0.241	0.219	0.513	1.079	0.306	0.543
11	N5									0.200	0.289	**0.118**	0.221	0.588	1.237	**0.099**	0.191
Horizon = 3 week; (R, P) = (101,50)																	
0	FP	2.786				1.270				0.182				0.560			
1	L1	2.756	0.989	0.326	0.313	1.256	0.988	0.318	0.316	0.000	-0.182	0.865	0.863	0.000	0.000	1.000	1.000
2	L2	2.811	1.009	0.808	0.379	1.275	1.004	0.579	0.375	-0.253	-0.434	0.939	0.910	0.440	0.786	0.976	0.994
3	L3	2.808	1.008	0.678	0.468	1.271	1.000	0.452	0.432	-0.181	-0.363	0.910	0.915	0.440	0.786	0.986	0.998
4	N1	12.337	4.427	0.922	0.801	1.751	1.378	0.921	0.802	-0.291	-0.472	0.963	0.940	0.420	0.750	0.996	1.000
6	N2	7.395	2.654	0.968	0.911	1.883	1.483	0.987	0.914	-0.185	-0.366	0.933	0.969	0.420	0.750	0.941	0.999
8	N3	2.737	0.982	0.218	0.951	1.246	0.981	0.215	0.933	0.088	-0.093	0.841	0.982	0.480	0.857	0.893	1.000
10	N4									0.298	0.117	0.254	0.842	0.600	1.071	0.229	0.801
11	N5									0.341	0.159	0.393	0.806	0.640	1.143	0.303	0.581
Horizon = 1 month; (R, P) = (72,35)																	
0	FP	3.858				1.590				0.142				0.543			
1	L1	3.919	1.016	0.630	0.583	1.621	1.019	0.684	0.654	0.000	-0.142	0.731	0.689	0.000	0.000	1.000	1.000
2	L2	4.044	1.048	0.710	0.627	1.667	1.048	0.797	0.691	-0.155	-0.297	0.743	0.740	0.429	0.789	0.825	0.907
3	L3	4.198	1.088	0.886	0.692	1.703	1.071	0.922	0.745	0.074	-0.067	0.656	0.739	0.457	0.842	0.775	0.877
4	N1	39.655	10.28	0.919	0.823	3.085	1.939	0.921	0.822	-0.045	-0.186	0.697	0.774	0.486	0.895	0.715	0.867
6	N2	24.815	6.432	0.916	0.914	3.010	1.892	0.949	0.899	-0.179	-0.320	0.769	0.827	0.400	0.737	0.881	0.886
8	N3	3.872	1.004	0.534	0.947	1.604	1.008	0.583	0.933	0.199	0.058	0.542	0.814	0.514	0.947	0.642	0.805
10	N4									0.366	0.224	0.379	0.588	0.571	1.053	0.450	0.591
11	N5									0.381	0.240	0.268	0.592	0.571	1.053	0.429	0.608

Note:

1. The reality check results are based on non-overlapping observations constructed from the original daily data, that is, weekly data for a one-week horizon, monthly data for a one-month horizon, etc.
2. For the model description, please refer to page 13-14.
3. P^1_{RC} is the bootstrap p-values of White's (2000) test for the null that model n is no better in forecasting than the benchmark (FP) model, while P^2_{RC} is for the null that the best of the first n alternative models is no better in forecasting than the benchmark (FP) model.These p-values are based on 1,000 bootstrap replications and a bootstrap smoothing parameter a = 0.25. The results are similar when different values of a (0.5, 0.75) are used.
4. Redness denotes better performance of the alternative to the FP model in terms of ratio/difference, boldness denotes significance level at 15%.
5. The bootstrap program was generously provided by Tae-Hwy Lee.

TABLE 5 (Continued). Reality check on predictive ability over different horizons

Panel D. Japanese Yen (JY); R/P = 2

		MSFE				MAFE				MFTR				MCFD			
k	Model	MSFE	Ratio	P^1_{RC}	P^2_{RC}	MAFE	Ratio	P^1_{RC}	P^2_{RC}	MFTR	Diff	P^1_{RC}	P^2_{RC}	MCFD	Ratio	P^1_{RC}	P^2_{RC}
									Horizon = 1 day; (R, P) = (1594,796)								
0	FP	0.367				0.460				-0.004				0.496			
1	L1	0.367	0.999	0.317	0.324	0.460	1.000	0.557	0.548	0.000	0.004	0.421	0.415	0.000	0.000	1.000	1.000
2	L2	0.367	1.000	0.186	0.324	0.460	1.000	0.218	0.601	-0.004	0.000	0.501	0.556	0.496	1.000	0.509	0.702
3	L3	0.368	1.002	0.953	0.560	0.460	1.001	0.915	0.783	0.004	0.008	0.374	0.545	0.505	1.018	0.325	0.524
4	N1	0.368	1.001	0.779	0.777	0.460	1.000	0.402	0.723	0.014	0.019	**0.110**	0.350	0.514	1.035	**0.142**	0.338
5	N2	0.380	1.035	0.999	0.926	0.467	1.016	0.994	0.910	-0.044	-0.039	0.925	0.508	0.476	0.959	0.806	0.475
6	N3	0.367	1.000	0.523	0.979	0.460	1.001	0.843	0.955	-0.011	-0.007	0.598	0.562	0.489	0.985	0.671	0.529
7	N4									0.013	0.017	0.247	0.451	0.501	1.010	0.359	0.602
8	N5									-0.015	-0.010	0.605	0.482	0.480	0.967	0.677	0.629
									Horizon = 1 week; (R, P) = (340,169)								
0	FP	1.847				1.107				0.155				0.533			
1	L1	1.854	1.004	0.685	0.665	1.102	0.996	0.368	0.354	0.000	-0.155	0.971	0.968	0.000	0.000	1.000	1.000
2	L2	1.852	1.003	0.595	0.684	1.102	0.996	0.374	0.405	0.068	-0.087	0.730	0.915	0.491	0.922	0.756	0.947
3	L3	1.862	1.008	0.681	0.696	1.104	0.998	0.440	0.425	-0.021	-0.177	0.878	0.921	0.479	0.900	0.817	0.947
4	N1	1.895	1.026	0.913	0.763	1.117	1.010	0.808	0.526	0.036	-0.119	0.851	0.950	0.509	0.956	0.688	0.903
6	N2	1.925	1.042	0.719	0.911	1.105	0.999	0.344	0.775	-0.018	-0.173	0.896	0.964	0.527	0.989	0.489	0.851
8	N3	1.870	1.013	0.573	0.921	1.104	0.998	0.331	0.798	-0.026	-0.181	0.933	0.967	0.509	0.956	0.670	0.857
10	N4									-0.102	-0.257	0.990	0.981	0.462	0.867	0.937	0.889
11	N5									0.072	-0.083	0.881	0.999	0.515	0.967	0.862	0.979
									Horizon = 2 week; (R, P) = (161,80)								
0	FP	3.335				1.431				0.014				0.463			
1	L1	3.239	0.971	**0.134**	**0.129**	1.396	0.976	**0.096**	**0.118**	0.000	-0.014	0.393	0.384	0.000	0.000	1.000	1.000
2	L2	3.268	0.980	0.162	0.156	1.409	0.985	**0.123**	**0.128**	-0.074	-0.088	0.679	0.622	0.438	0.946	0.677	0.816
3	L3	3.287	0.986	0.368	0.283	1.425	0.996	0.454	0.169	-0.031	-0.045	0.568	0.684	0.438	0.946	0.618	0.875
4	N1	3.374	1.012	0.606	0.354	1.457	1.018	0.737	0.257	-0.050	-0.064	0.616	0.730	0.463	1.000	0.429	0.719
6	N2	3.209	0.96	**0.106**	0.456	1.396	0.976	0.092	0.441	0.082	0.067	0.365	0.483	0.525	1.135	0.097	0.327
8	N3	3.247	0.974	0.451	0.679	1.401	0.979	0.374	0.660	0.003	-0.012	0.513	0.525	0.475	1.027	0.276	0.350
10	N4									0.156	0.142	0.170	0.572	0.488	1.054	0.177	0.396
11	N5									0.191	0.177	0.188	0.585	0.563	1.216	**0.070**	0.195
									Horizon = 3 week; (R, P) = (101,50)								
0	FP	5.851				1.963				-0.040				0.500			
1	L1	5.784	0.988	0.546	0.518	1.964	1.000	0.654	0.649	0.000	0.040	0.607	0.603	0.000	0.000	1.000	1.000
2	L2	5.803	0.992	0.340	0.528	1.953	0.995	0.333	0.577	0.225	0.265	**0.101**	0.411	0.580	1.160	0.028	0.299
3	L3	6.122	1.046	0.727	0.726	1.988	1.013	0.619	0.715	-0.319	-0.278	0.899	0.528	0.480	0.960	0.664	0.325
4	N1	6.492	1.110	0.860	0.860	2.034	1.036	0.789	0.866	0.015	0.055	0.617	0.694	0.540	1.080	0.504	0.443
6	N2	7.294	1.246	0.796	0.955	2.119	1.079	0.800	0.940	-0.293	-0.253	0.872	0.746	0.420	0.840	0.931	0.473
8	N3	5.954	1.018	0.524	0.985	1.979	1.008	0.530	0.963	-0.001	0.040	0.456	0.771	0.500	1.000	0.463	0.491
10	N4									0.346	0.386	**0.122**	0.647	0.540	1.080	**0.121**	0.543
11	N5									0.312	0.352	0.471	0.699	0.500	1.000	0.627	0.598
									Horizon = 1 month; (R, P) = (72,35)								
0	FP	6.805				2.169				0.060				0.486			
1	L1	6.441	0.947	0.430	0.394	2.107	0.971	0.390	0.358	0.000	-0.060	0.736	0.711	0.000	0.000	1.000	1.000
2	L2	6.467	0.950	0.278	0.395	2.099	0.968	0.193	0.326	0.017	-0.043	0.651	0.876	0.486	1.000	0.344	0.691
3	L3	6.668	0.980	0.339	0.491	2.108	0.972	0.269	0.441	0.405	0.344	0.319	0.480	0.543	1.118	0.314	0.530
4	N1	8.988	1.32	0.905	0.582	2.443	1.126	0.893	0.552	-0.255	-0.316	0.745	0.530	0.457	0.941	0.589	0.547
6	N2	7.008	1.030	0.499	0.709	2.208	1.018	0.591	0.672	-0.004	-0.065	0.539	0.599	0.457	0.941	0.606	0.608
8	N3	6.923	1.017	0.450	0.894	2.142	0.987	0.328	0.842	0.306	0.245	0.335	0.611	0.543	1.118	0.239	0.632
10	N4									0.186	0.125	0.069	0.613	0.543	1.118	0.042	0.641
11	N5									0.462	0.402	0.531	0.698	0.543	1.118	0.552	0.728

Note:

1 The reality check results are based on non-overlapping observations constructed from the original daily data, that is, weekly data for a one-week horizon, monthly data for a one-month horizon, etc.

2 For the model description, please refer to page 13-14.

3 P^1_{RC} is the bootstrap p-values of White's (2000) test for the null that model n is no better in forecasting than the benchmark (FP) model, while P^2_{RC} is for the null that the best of the first n alternative models is no better in forecasting than the benchmark (FP) model.These p-values are based on 1,000 bootstrap replications and a bootstrap smoothing parameter a = 0.25. The results are similar when different values of a (0.5, 0.75) are used.

4 Redness denotes better performance of the alternative to the FP model in terms of ratio/difference, boldness denotes significance level at 15%.

5 The bootstrap program was generously provided by Tae-Hwy Lee.

TABLE 5 (Continued). Reality check on predictive ability over different horizons

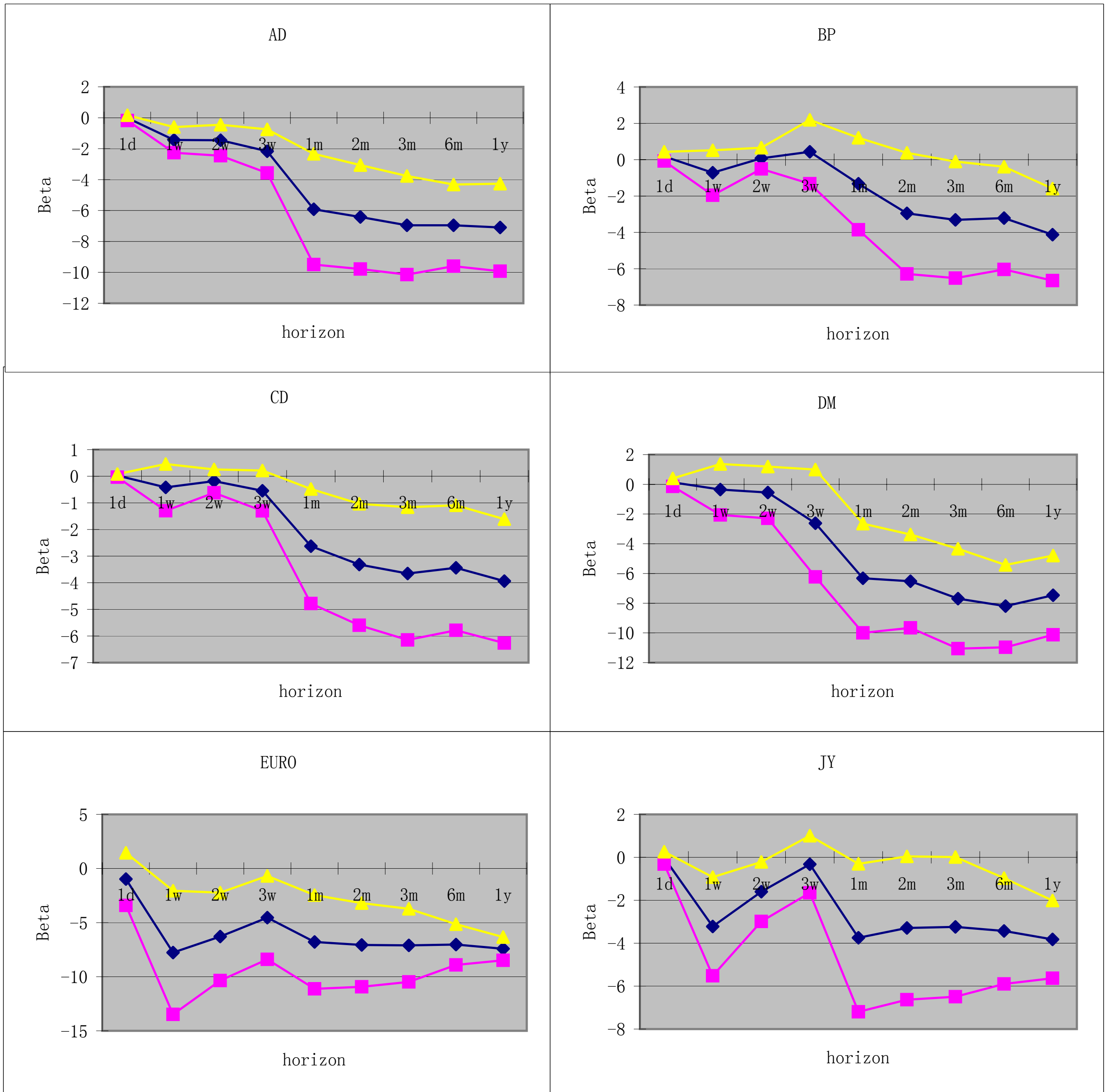
Panel E. Swiss Franc (SF); R/P = 2

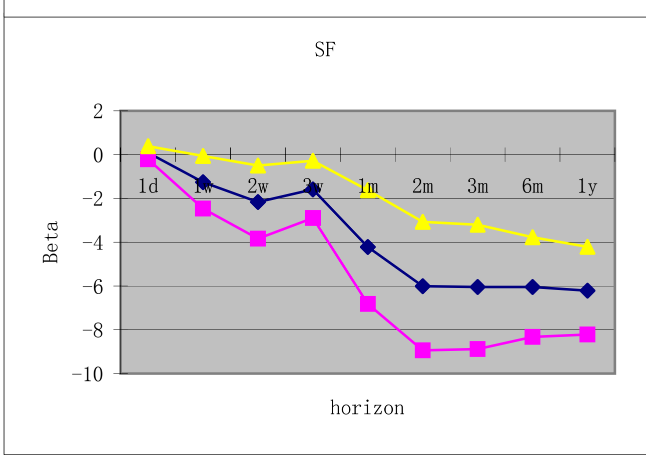
		MSFE				MAFE				MFTR				MCFD			
k	Model	MSFE	Ratio	P^1_{RC}	P^2_{RC}	MAFE	Ratio	P^1_{RC}	P^2_{RC}	MFTR	Diff	P^1_{RC}	P^2_{RC}	MCFD	Ratio	P^1_{RC}	P^2_{RC}
									Horizon = 1 day; (R, P) = (1594,796)								
0	FP	0.475				0.529				-0.024				0.481			
1	L1	0.474	0.997	**0.066**	**0.050**	0.528	0.999	0.161	**0.139**	0.000	0.024	**0.141**	0.151	0.000	0.000	1.000	1.000
2	L2	0.475	1.000	0.343	**0.050**	0.529	1.000	0.617	**0.139**	-0.028	-0.004	0.777	0.153	0.475	0.987	0.893	0.967
3	L3	0.474	0.997	0.258	0.258	0.527	0.996	**0.073**	**0.080**	0.023	0.047	**0.021**	**0.044**	0.524	1.089	**0.018**	**0.026**
4	N1	0.478	1.006	0.856	0.470	0.530	1.003	0.818	0.277	-0.001	0.023	**0.173**	**0.060**	0.496	1.031	0.213	**0.027**
5	N2	0.474	0.997	0.369	0.743	0.526	0.996	0.227	0.518	-0.013	0.011	0.379	0.155	0.482	1.003	0.516	**0.019**
6	N3	0.474	0.997	**0.059**	0.807	0.528	0.999	**0.136**	0.573	0.019	0.043	**0.065**	0.199	0.518	1.076	**0.043**	**0.022**
7	N4									-0.004	0.020	0.300	0.265	0.481	1.000	0.486	**0.042**
8	N5									0.024	0.048	**0.133**	0.287	0.518	1.076	**0.132**	**0.066**
									Horizon = 1 week; (R, P) = (340,169)								
0	FP	2.118				1.158				-0.004				0.515			
1	L1	2.118	1.000	0.540	0.503	1.161	1.003	0.769	0.769	0.000	0.004	0.515	0.518	0.000	0.000	1.000	1.000
2	L2	2.142	1.011	0.959	0.655	1.167	1.008	0.966	0.861	-0.086	-0.082	0.769	0.667	0.485	0.943	0.767	0.909
3	L3	2.160	1.020	0.990	0.729	1.169	1.010	0.966	0.909	-0.237	-0.233	0.965	0.712	0.432	0.839	0.961	0.926
4	N1	2.318	1.094	0.971	0.869	1.210	1.045	0.967	0.955	-0.138	-0.134	0.914	0.755	0.479	0.931	0.848	0.940
6	N2	2.273	1.073	0.958	0.956	1.195	1.033	0.957	0.988	-0.167	-0.163	0.929	0.845	0.450	0.874	0.942	0.963
8	N3	2.107	0.995	0.218	0.914	1.155	0.998	0.349	0.941	-0.073	-0.069	0.734	0.688	0.491	0.954	0.754	0.387
10	N4									-0.069	-0.065	0.707	0.750	0.462	0.897	0.908	0.486
11	N5									0.144	0.148	0.182	0.502	0.550	1.069	0.335	0.512
									Horizon = 2 week; (R, P) = (161,80)								
0	FP	4.393				1.748				0.223				0.513			
1	L1	4.378	0.997	0.498	0.523	1.746	0.999	0.535	0.536	0.000	-0.223	0.870	0.872	0.000	0.000	1.000	1.000
2	L2	4.504	1.025	0.799	0.612	1.744	0.998	0.515	0.592	-0.304	-0.527	0.952	0.918	0.525	1.024	0.513	0.597
3	L3	4.499	1.024	0.778	0.681	1.745	0.999	0.504	0.651	-0.155	-0.378	0.889	0.941	0.538	1.049	0.344	0.593
4	N1	4.735	1.078	0.905	0.744	1.784	1.021	0.648	0.707	-0.444	-0.667	0.971	0.947	0.438	0.854	0.789	0.627
6	N2	4.914	1.119	0.988	0.837	1.811	1.037	0.885	0.813	-0.427	-0.650	0.974	0.950	0.513	1.000	0.545	0.666
8	N3	4.353	0.991	0.533	0.868	1.722	0.985	0.404	0.768	0.083	-0.140	0.731	0.777	0.563	1.098	0.348	0.501
10	N4									0.005	-0.218	0.890	0.814	0.525	1.024	0.264	0.514
11	N5									0.357	0.134	0.159	0.791	0.525	1.024	0.240	0.516
									Horizon = 3 week; (R, P) = (101,50)								
0	FP	6.345				2.007				-0.306				0.480			
1	L1	6.040	0.952	**0.025**	**0.017**	1.964	0.978	**0.024**	**0.022**	0.000	0.306	**0.127**	**0.100**	0.000	0.000	1.000	1.000
2	L2	6.332	0.998	**0.141**	**0.017**	2.002	0.998	**0.080**	**0.022**	-0.358	-0.052	0.612	**0.107**	0.480	1.000	0.361	0.512
3	L3	6.436	1.014	0.691	**0.017**	2.037	1.015	0.804	**0.024**	-0.166	0.139	0.322	0.251	0.500	1.042	0.261	0.433
4	N1	6.192	0.976	0.232	**0.142**	1.983	0.988	0.264	0.215	0.215	0.520	**0.030**	**0.113**	0.540	1.125	**0.118**	0.205
6	N2	7.224	1.138	0.966	0.476	2.176	1.084	0.964	0.549	-0.238	0.068	0.346	0.163	0.500	1.042	0.256	0.321
8	N3	6.053	0.954	**0.086**	0.676	1.968	0.980	0.265	0.788	-0.033	0.273	**0.108**	0.188	0.520	1.083	0.202	0.430
10	N4									0.277	0.583	**0.104**	0.248	0.480	1.000	0.509	0.569
11	N5									0.621	0.926	**0.025**	**0.073**	0.500	1.042	0.261	0.579
									Horizon = 1 month; (R, P) = (72,35)								
0	FP	7.852				2.355				0.180				0.400			
1	L1	7.920	1.009	0.611	0.593	2.230	0.947	**0.033**	**0.030**	0.000	-0.180	0.715	0.703	0.000	0.000	1.000	1.000
2	L2	8.309	1.058	0.854	0.660	2.280	0.968	**0.078**	**0.045**	-0.248	-0.427	0.831	0.819	0.400	1.000	0.402	0.559
3	L3	8.977	1.143	0.970	0.802	2.389	1.014	0.675	0.201	-0.550	-0.730	0.938	0.897	0.429	1.071	0.478	0.630
4	N1	9.860	1.256	1.000	0.851	2.541	1.079	0.930	0.473	-0.640	-0.820	0.963	0.910	0.429	1.071	0.505	0.680
6	N2	22.046	2.808	0.930	0.912	2.977	1.264	0.824	0.660	0.416	0.236	0.506	0.734	0.600	1.500	**0.075**	0.180
8	N3	7.873	1.003	0.589	0.911	2.203	0.935	**0.103**	0.628	-1.091	-1.271	0.990	0.767	0.371	0.929	0.650	0.183
10	N4									0.667	0.487	**0.149**	0.329	0.514	1.286	**0.094**	0.199
11	N5									0.823	0.643	0.198	0.335	0.514	1.286	0.179	0.203

Note:

1. The reality check results are based on non-overlapping observations constructed from the original daily data, that is, weekly data for a one-week horizon, monthly data for a one-month horizon, etc.
2. For the model description, please refer to page 13-14.
3. P^1_{RC} is the bootstrap p-values of White's (2000) test for the null that model n is no better in forecasting than the benchmark (FP) model, while P^2_{RC} is for the null that the best of the first n alternative models is no better in forecasting than the benchmark (FP) model.These p-values are based on 1,000 bootstrap replications and a bootstrap smoothing parameter a = 0.25. The results are similar when different values of a (0.5, 0.75) are used.
4. Redness denotes better performance of the alternative to the FP model in terms of ratio/difference, boldness denotes significance level at 15%.
5. The bootstrap program was generously provided by Tae-Hwy Lee.

Figure 1. The Beta over Different Horizons.

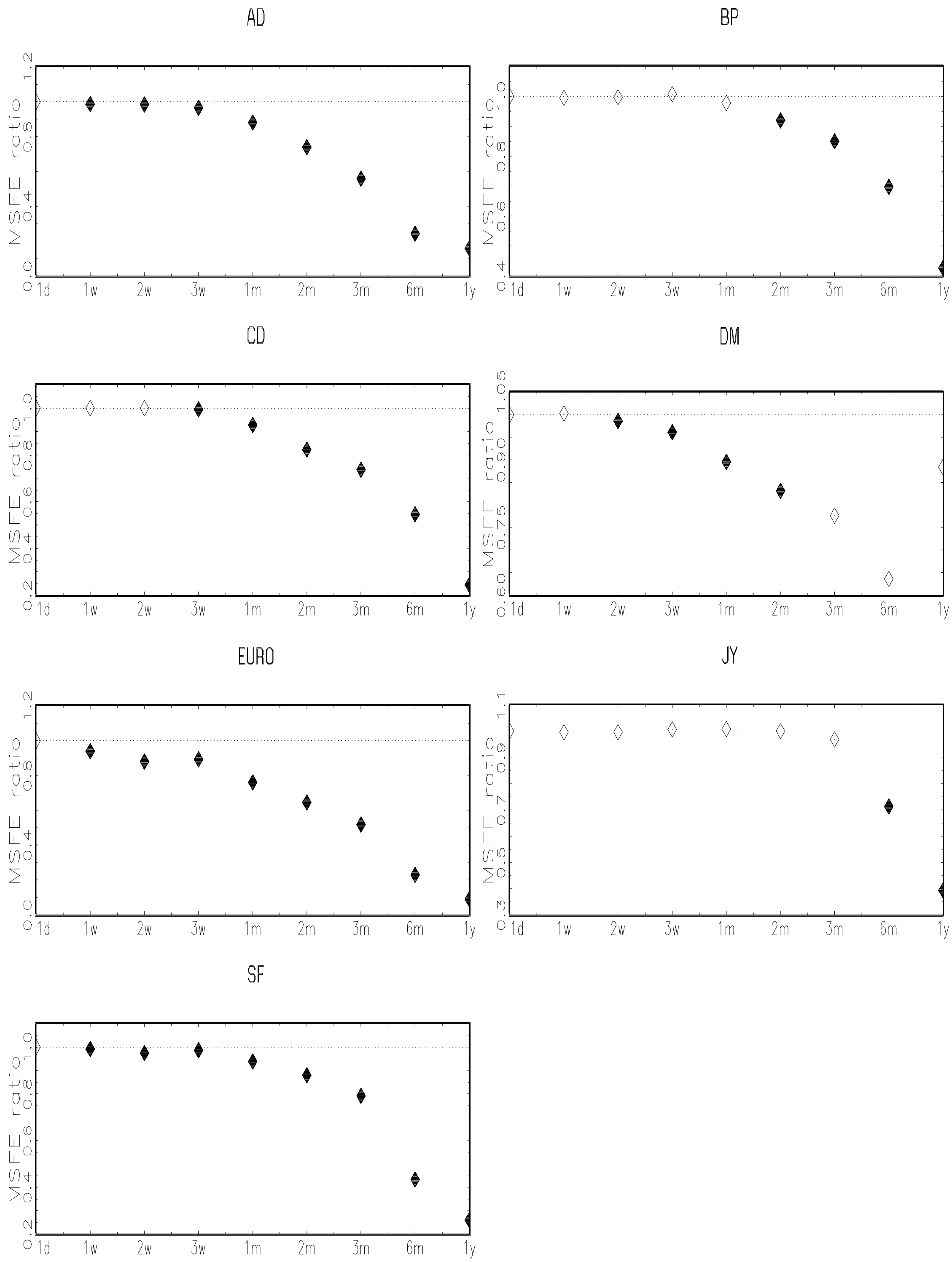




Note: The central line connects the point estimates of beta, while the lower and upper lines connect the lower and upper bounds of the 95% confidence interval.

Figure 2. Out-of-Sample Comparison (FP v.s. RW).

Plate 1. MSFE Ratio (R/P =2).

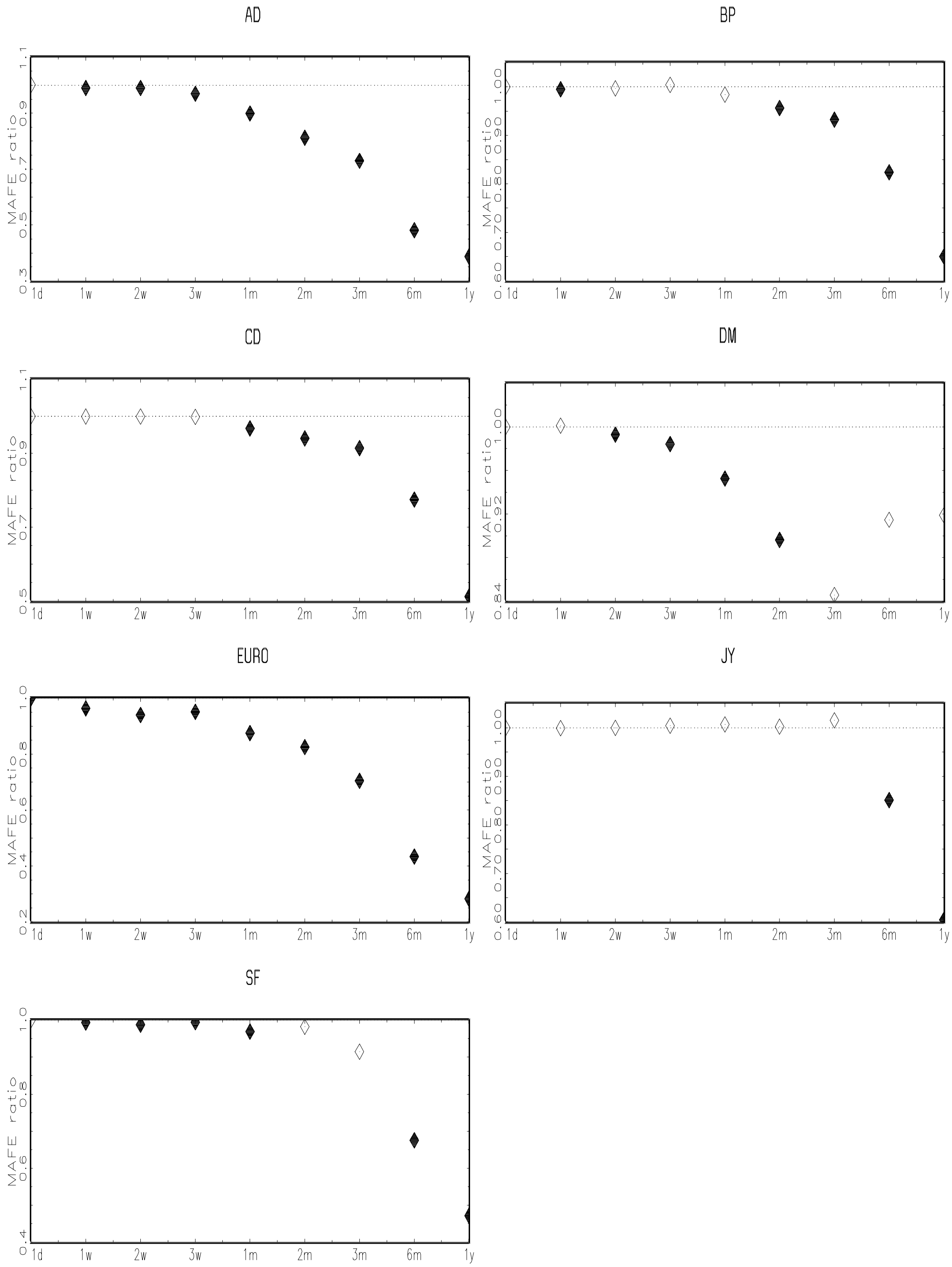


Note: MSFE ratio = $MSFE^1/MSFE^0$, where ^{1} denotes the FP model and ^{0} denotes the RW model.

The solid diamonds denote the ratios significant at the 15% level in the Diebold-Mariano test, while the diamonds denote the insignificant cases. The dotted line gives the benchmark ratio (=1).

Figure 2 (Continued). Out-of-Sample Comparison (FP v.s. RW).

Plate 2. MAFE Ratio (R/P =2).

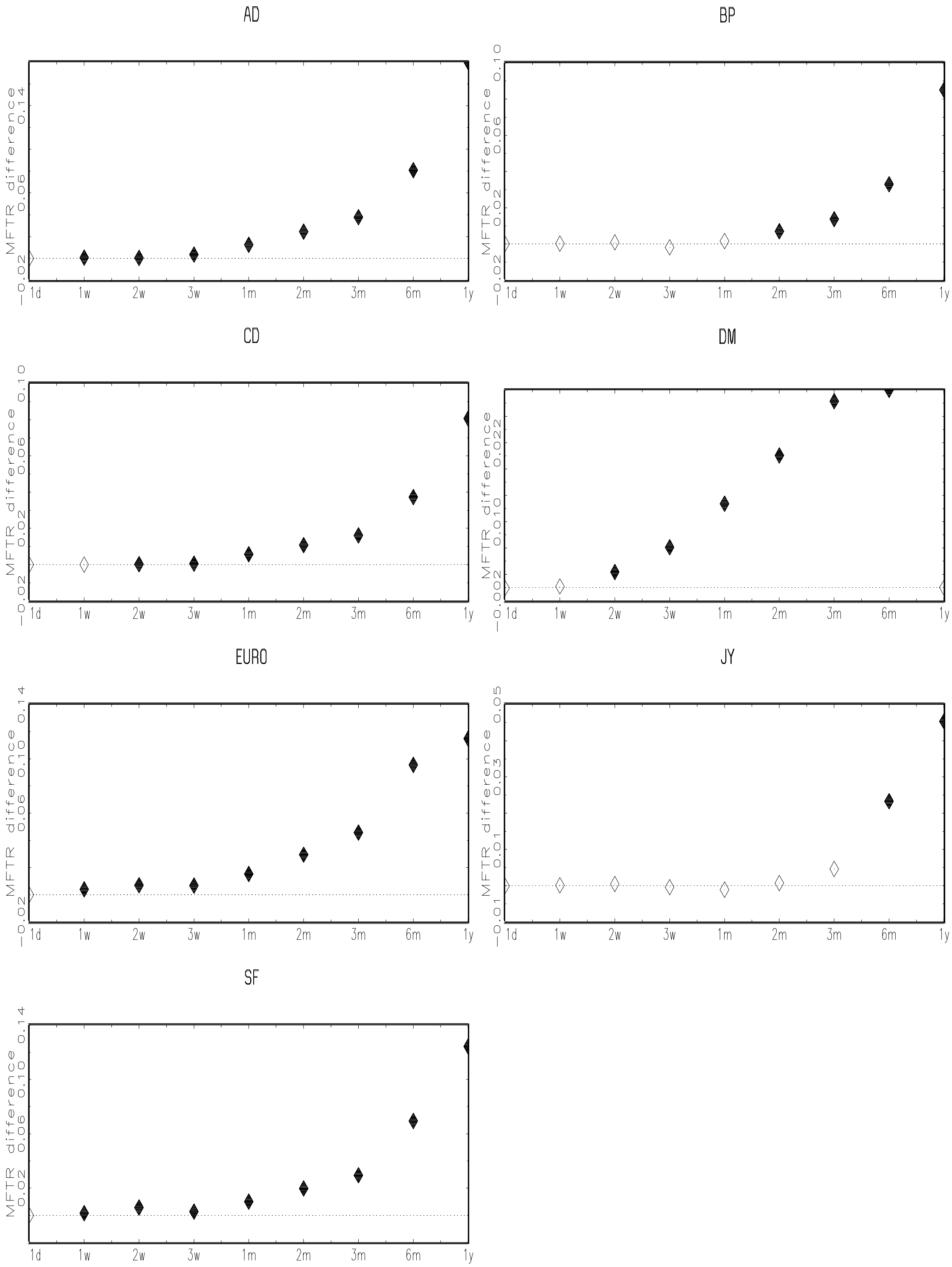


Note: MAFE ratio = $MAFE^1/MAFE^0$, where ^{1} denotes the FP model and ^{0} denotes the RW model.

The solid diamonds denote the ratios significant at the 15% level in the Diebold-Mariano test, while the diamonds denote the insignificant cases. The dotted line gives the benchmark ratio (=1).

Figure 2 (Continued). Out-of-Sample Comparison (FP v.s. RW).

Plate 3. MFTR Difference (R/P =2).

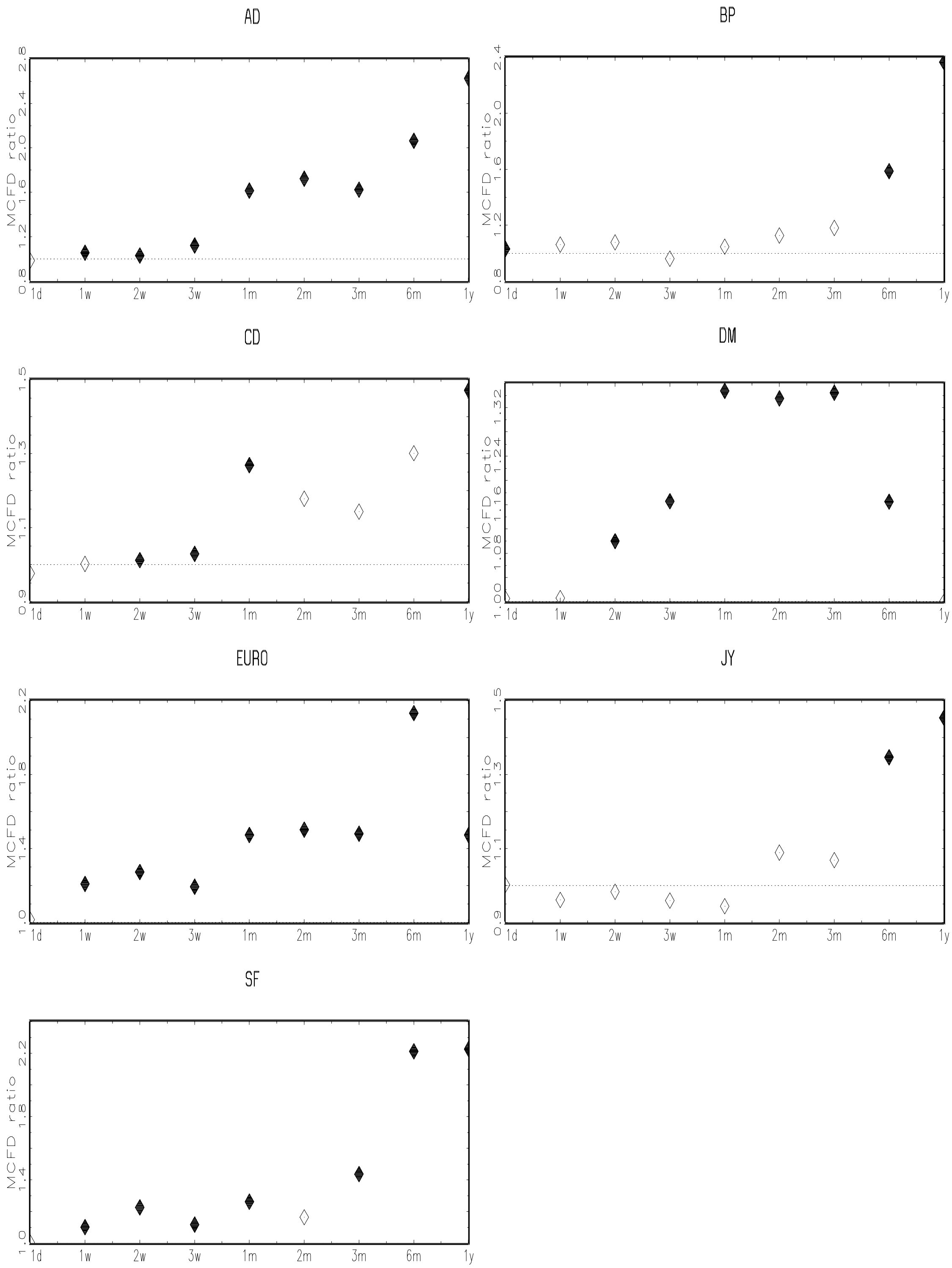


Note: MFTR difference = $MFTR^1$-$MFTR^0$, where ^{1} denotes the FP model and ^{0} denotes the RW model.

The solid diamonds denote the differences significant at the 15% level in the Diebold-Mariano test, while the diamonds denote the insignificant cases. The dotted line gives the benchmark difference (=0).

Figure 2 (Continued): Out-of-Sample Comparison (FP v.s. RW).

Plate 4. MCFD Ratio (R/P =2).

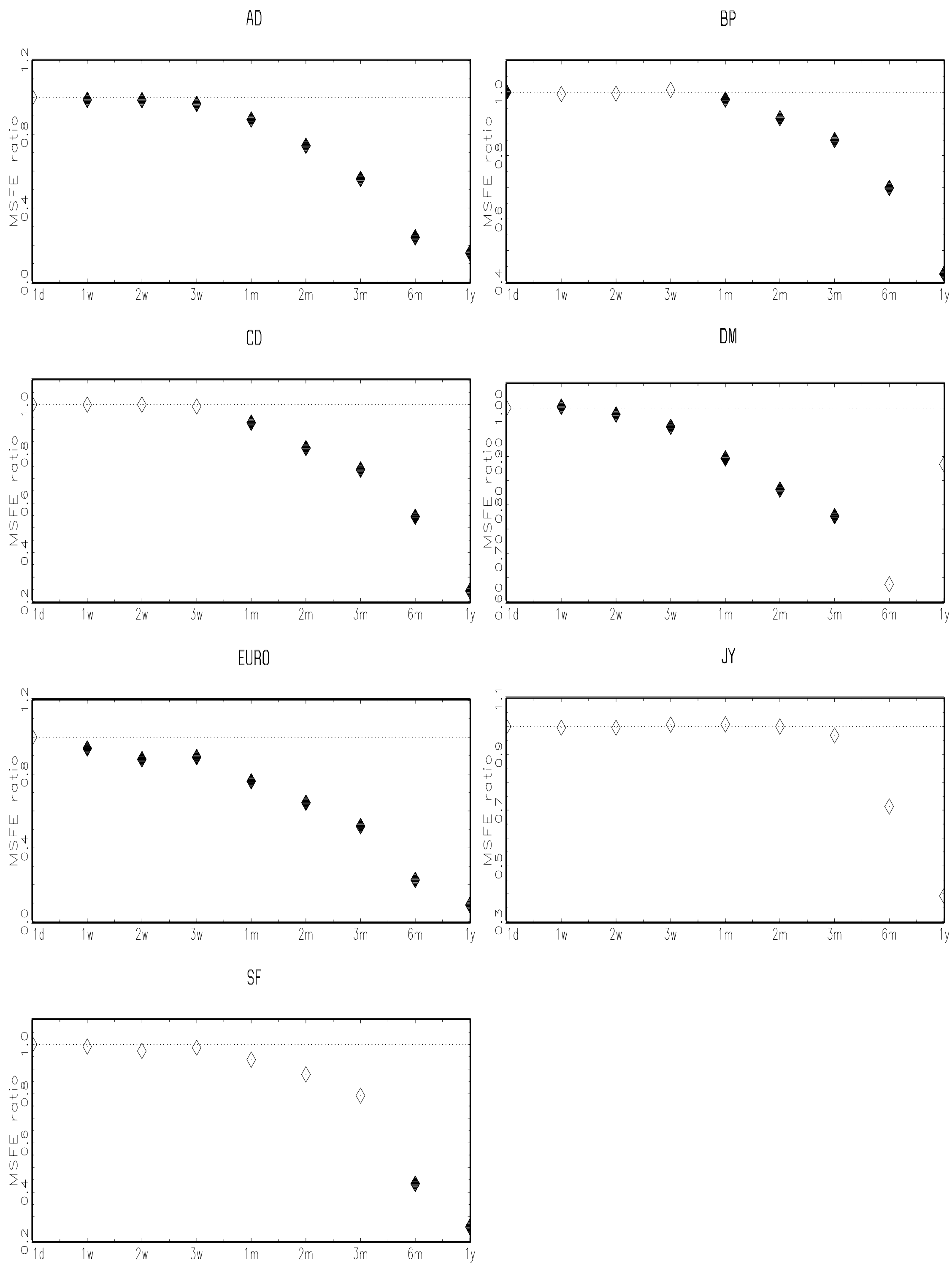


Note: MCFD ratio = $MCFD^1/MCFD^0$, where ^{1} denotes the FP model and ^{0} denotes the RW model.

The solid diamonds denote the ratios significant at the 15% level in the Diebold-Mariano test, while the diamonds denote the insignificant cases. The dotted line gives the benchmark ratio (=1).

Figure 2 (Continued). Out-of-Sample Comparison (FP v.s. RW).

Plate 5. MSFE Ratio with CCS Test (R/P =2).



Note: MSFE ratio = $MSFE^1/MSFE^0$, where ^{1} denotes the FP model and ^{0} denotes the RW model.

The solid diamonds denote the ratios significant at the 15% level in the Chao, Corradi and Swanson's (CCS) test, while the diamonds denote the insignificant cases. The dotted line gives the benchmark ratio (=1).

CHAPTER II

INTER-MARKET INFORMATION TRANSMISSIONS: EVIDENCE FROM HIGH-FREQUENCY INDEX FUNDS DATA

Introduction

Important to understanding international financial integration, the linkages across international stock markets have received much attention over the past two decades. Most of the research relies on parametric GARCH models based on low-frequency data, and often finds conflicting results about causality. For example, Hamao, Masulis, and Ng (1990) studied the return and volatility spillovers for Tokyo, London, and New York, using a GARCH-in-mean model based on daily open and closing prices of three major stock indices. They found uni-directional volatility spillovers from New York to Tokyo and London for the pre-October 1987 period. On the other hand, Lin, Engle, and Ito (1994) found bi-directional volatility spillovers between Tokyo and New York for the same period. In contrast to Hamao, Masulis, and Ng (1990), they used a signal-extraction GARCH model to separate the global factor that affected stock returns globally from a local factor that affected stock returns locally. And their model was based on the first 15-minute or 30-minute index prices in order to mitigate the "stale quotes" problem. Both the GARCH-in-mean and the signal-extraction GARCH approaches, however, did not take into account the asymmetric relation between stock returns and volatility changes. In view of this problem, Susmel and Engle (1994) used an asymmetric GARCH model to study the interrelations between New York and London. They found weak evidence of bi-directional volatility spillovers between these two markets. Other related works include Bae and Karolyi (1994), Karolyi (1995), Koutmos and Booth (1995), and Ng (2000). Despite extensive studies in this literature, the issues of which model is the most appropriate, and in which direction the true causalities go remains unresolved.

In this study, we examine the short-term linkages in stock returns and volatilities by utilizing high-frequency, intra-day exchange-traded funds (ETFs) data. As a new type of investment tool developed in the mid-1990s, the ETFs are aimed at achieving the same return as a particular market index, such as the

S&P 500 index and the Morgan Stanley Capital International (MSCI) equity indices.[27] Moreover, unlike most of the international equity indices, the ETFs are traded continuously in stock exchanges, such as the American Stock Exchange, and their real-time transaction prices are available at tick-by-tick frequencies.[28] These two features enable us to examine the cross-market linkages from a more microscopic view, which in turn provides us with higher statistical precision than do the lower-frequency data employed in the previous literature.

Based on the high-frequency feature of our data, we apply the realized volatility method to estimate daily return volatilities and other quantities of interest (daily standardized returns and daily correlations). The realized volatility method has three major advantages. (1) In contrast to the parametric GARCH framework employed in previous studies, the realized volatility method does not assume any specific functional form for the data-generating process, and thus is free of model misspecification (Andersen, Bollerslev, and Diebold, 2005). This will avoid the complication of searching a correctly specified parametric volatility measurement model. (2) According to the theory of quadratic variation, the realized volatilities measure the latent volatilities approximately free of error, as long as the prices are sampled at an appropriately high frequency (Andersen, Bollerslev, Diebold, and Labys, ABDL, 2001, 2003; Barndorff-Nielsen and Shephard, 2004). Hence, the analysis based on these measures has a high level of statistical precision. The GARCH approach, on the other hand, may provide very noisy estimates of daily volatilities, even though the model is correctly specified (ABDL, 2003). (3) Our method is easy to implement even in a high-dimensional multivariate environment. This feature is important to the study of international stock markets, as a broad range of assets are involved. In contrast, the estimation of GARCH models becomes very difficult when the dimension of analysis extends to three or above.

The second extension of this paper is that we examine not only the in-sample, cross-market linkages, but also how these linkages help to forecast the relevant quantities out-of-sample. Most studies in this literature rely on in-sample analysis with little emphasis on out-of-sample forecasts, though forecasts are important in financial practice. The main reason that the literature examines the out-of-sample forecasts less frequently lies in the fact that volatilities and correlations are not directly observable. As a result, the measure of

[27] MSCI indices represent a broad aggregation of national equity markets, and are the leading benchmarks for global portfolio managers.

[28] For more detailed information about ETFs, see www.amex.com and www.ishares.com.

out-of-sample forecast errors (such as mean squared errors) is not easy to calculate. Since the volatilities and correlations studied in this paper are constructed from high-frequency data approximately free of measurement error, they may be treated as directly observable. Therefore, the out-of-sample technique can easily be accommodated.

Lastly, we study a broader range of assets over a much longer period of time than in those cases examined in the previous literature. As mentioned above, the stock markets examined in earlier studies are normally confined to two or three, and the sample period spans confined within five years. In contrast, the international ETFs used in this study track fourteen international stock markets. The sample period spans nearly a decade, ranging from May 1996 to December 2004. The extensive sample examined in this paper provides us with a higher degree of freedom and more precise estimators.

We note that all the ETFs studied in this paper are traded in the US market. Therefore, the ETFs' return volatilities should be interpreted as the information revealed during the opening time of the US market. We do not think, however, that this qualification causes a problem for the purpose of this paper for at least two reasons. (1) Even though the underlying stock market (for example, the Tokyo market) is closed, news that has impact on economic fundamentals (for example, an earthquake in Tokyo) is still released, and may affect the price of the corresponding ETF traded in the US market. (2) The US news released during the opening time of the US market may have impact on investors' expectations about international stock markets, which in turn causes the ETFs' prices to change. For example, an announcement of an increased trade deficit with the UK may cause investors to expect a booming UK economy, and hence the UK ETF's price is likely to increase. Indeed, there is an additional benefit using the ETFs instead of national stock indices: it avoids the problem of non-synchronous trading in international stock markets. This outcome enables us to examine the contemporaneous correlations between different stock market returns.[29]

Our main goal is to detect the cross-market linkages using the ETFs prices. The linkages, if any, could be generated through different channels, such as the macro news (e.g., Becker, Finnerty, and Friedman, 1995), or the contagion effects (e.g., King and Wadhwani, 1990). In addition, the ETFs prices, though generally tracking the stock indices very well, may significantly deviate from their fair values for a short

[29] Similar methodology was employed in Karolyi and Stulz (1996) where they used the NYSE-traded American Depository Receipts (ADRs) as a proxy for the Japanese stock index.

period of time (e.g., Jares and Lavin, 2004). We leave further exploration along these lines for future work.

The remainder of the paper is organized as follows. In Section 2, we describe our high-frequency, intra-day ETFs data. In Section 3, we provide a brief discussion of the theory underpinning our realized volatility measures, along with a discussion of the volatility and correlation calculations. We then analyze the unconditional distributions and time series properties of the calculated measures in Section 4. Using these properties, we examine the in-sample return, volatility and correlation spillovers, and the associated out-of-sample forecastability in Section 5. In Section 6, we extend our analysis to investigate additional related hypotheses, such as the Monday effects, the leverage effects, and the contagion effects. We conclude in Section 7 with a brief summary of our main findings and some suggestions for future research.

Data

We obtain high-frequency intra-day transaction data between 9:30 and 16:00 Eastern Standard Time (EST) from the Trade and Quote (TAQ) database for the S&P 500 index fund (US), as well as fourteen international ETFs that track the MSCI indices of the following countries: Austria, Belgium, France, Germany, Hong Kong, Italy, Japan (JP), Malaysia, the Netherlands, Singapore, Spain, Sweden, Switzerland, and the United Kingdom.[30] The S&P 500 index fund was introduced on January 29, 1993, and all the international ETFs were introduced on March 12, 1996. We examine the period from May 1, 1996, to December 31, 2004, to avoid thin trading during the introductory period. To see how closely the ETFs track their underlying stock market indices, we also obtain the corresponding daily MSCI indices denominated in US Dollars from Datastream Inc. over the same period.

We calculate the correlations of daily, weekly, and monthly returns between the ETFs and their underlying MSCI indices. The results are reported in Table 1. Generally, the correlations of daily returns are low, especially for Asian countries (with the minimum of 0.50 for Hong Kong). These are likely due to non-synchronized daily data, as the ETFs data are based on New York trading time, while the MSCI indices are based on local trading time.[31] As the horizon extends to one week or one month, the correla-

[30] For more information on the TAQ database, see Andersen, Bollerslev, Diebold, and Ebens (ABDE, 2001).

[31] As detailed in Hamao, Masulis, and Ng (1990), the New York trading time does not overlap with the Asian market trading time and partly overlaps with the European market trading time.

tions become remarkably high (with the minimum of 0.75 and 0.86 for Malaysia at weekly and monthly frequencies, respectively). The US fund has the highest correlations among all at 0.95, 0.97, and 0.96 for daily, weekly, and monthly returns, respectively. In summary, the ETFs track the underlying stock market indices reasonably well and can serve as good instruments for international asset allocation and risk hedging.

Table 2a gives the list of ETFs ticker symbols as well as a summary of their daily transactions. From Table 2a, we can see that the transaction intensity varies among the ETFs, with the maximum of 3877 trades per day for the US, and the minimum of 6 trades per day for the Netherlands. Therefore, to achieve a certain degree of comparability and to qualify the high-frequency technique to be used later, we focus our investigation on the two most active ETFs, the US and JP, as well as two equally-weighted portfolios: the Asia ex-Japan portfolio (AS), which includes Hong Kong, Malaysia and Singapore; the Europe portfolio(EU), which includes Austria, Belgium, France, Germany, Italy, the Netherlands, Spain, Sweden, Switzerland, and the United Kingdom. Similar international portfolio construction is also used in Guidolin and Timmermann (2004).

The components of ETF portfolios and their transaction activities are summarized in Table 2b. As can be seen from Table 2b, the portfolios are active enough for high-frequency analysis (with the minimum of 122 trades per day, which amounts to the inter-trade duration of 3 minutes). Meanwhile, focusing on portfolios rather than individual assets has the following additional benefits: 1) it reduces the dimension of analysis and facilitates the statistical estimation and testing; and 2) it extracts the common movement in regional markets, which is more relevant in international asset pricing and risk hedging than the individual movement. The second point will be further illustrated below.

To achieve high statistical precision while avoiding possible market microstructure noise, we construct artificial five-minute prices for each individual ETF using the previous-tick interpolation method, that is, using the prices recorded at or immediately before the corresponding five-minute marks.[32] We then calculate the individual ETF returns as the logarithmic differences between adjacent prices, and construct

[32] For a detailed description of the interpolation methods, we refer to Dacorogna et al. (2001, chapter 3). We note that Zhang, Mykland, and Aït-Sahalia (2003), and Hansen and Lunde (2004) have proposed incorporating the microstructure noise to utilize higher frequency data. In view of our data properties (the two portfolios are not highly active), we think that the five-minute frequency may be the upper bound of the sampling frequencies. To see the robustness of our results, we also experiment with the half-hour frequency and get qualitatively similar results, which are available upon request.

the portfolio returns as the arithmetical means of the individual component returns.[33] In Figure 1, we plot a randomly selected sub-sample of the historical five-minute trade and quote returns for the studied ETFs/portfolios. Visual inspection shows that the quote data are very noisy, possibly due to discrete clustering and bid-ask bounce effects (Dacorogna et al., 2001, chapter 5). We therefore focus our analysis on trade data. Table 3 gives the summary statistics of five-minute trade returns. The means of all five-minute return series are approximately zero. In terms of five-minute volatilities, JP and AS are the most volatile series, with sample standard deviations around 0.3, followed by the US (0.16) and then EU (0.1). The US and EU are skewed to the left, while JP and AS are skewed to the right. All the return series are leptokurtic. These are generally in line with the stylized facts of high-frequency returns (e.g., Dacorogna, 2001, chapter 5).

In addition, we calculate the first principal components of the individual ETFs trade returns and plot them in Figure 2. Comparing Figure 2 with Figure 1a, we see that the portfolio returns behave almost identically to the first principal components of their individual ETFs returns, allowing for a different scale. This confirms our previous claim that the portfolio returns extract the common components in regional movement.

Measurement of volatilities and correlations

In view of the high-frequency property of our data, we apply a recently developed method, the realized volatility and correlation, to measure the ETFs return volatilities and correlations. This method assumes that the multivariate asset return process is a special semi-martingale (which is justified if the asset price process is arbitrage-free and has a finite instantaneous mean; e.g., Back, 1991, Meheswaran and Sims, 1993). Under this assumption as well as some other mild assumptions, the realized volatility and correlation are unbiased estimators of the conditional volatility and correlation, without further assuming any specific return generating model. ABDL (2001, 2003) present formal derivations and proofs, while Barndorff-Nielsen and Shephard (2004) provide the asymptotic distribution theory by adding the assumption that the (logarithmic) asset price process is a continuous stochastic volatility semi-martingale. Below we briefly

[33] All the returns are expressed in percentage. And the first five-minute return of a day is equal to the logarithmic difference between the first five-minute price and the previous day's last five-minute price, thus is quivalently the overnight return.

review the theory and empirical findings in the literature, and discuss the calculations of our volatility and correlation measures. For a thorough review and comparison between the realized volatility and other volatility measurements, we refer to Andersen, Bollerslev, and Diebold (2005).

Theory

The sample period is denoted as $[0, T]$. Let an n-dimensional vector of arbitrage-free logarithmic prices at any time point t be p_t, and the associated cumulative returns are defined as:

$$r^*(t) \equiv p_t - p_0 \ . \tag{7}$$

Under regular conditions (such as a finite mean in the asset price process, the information filtration satisfying the usual conditions of P completeness and right continuity, etc.), the cumulative return process is a special semi-martingale, and has the following unique canonical decomposition:

$$r^*(t) = \mu_t + M_t = \mu_t + M_t^c + \Delta M_t, \tag{8}$$

where μ_t (the mean return process) is a locally integrable and predictable process of finite variation, M_t (the return innovation process) is a local martingale, M_t^c and ΔM_t are the continuous part and the compensated jump part of M_t. Meanwhile, the quadratic variation or covariation (QV) of the return process at time t is well defined as:

$$[r, r]_t \equiv \text{plim}_{J\to\infty} \sum_{j=0}^{J-1} [r^*(t_{j+1}) - r^*(t_j)][r^*(t_{j+1}) - r^*(t_j)]' \tag{9}$$

for any sequence of partitions $t_0 = 0 < t_1 < \cdots < t_J = t$ with $\sup_j\{t_{j+1} - t_j\} \to 0$ as $J \to \infty$, and $\text{plim}_{J\to\infty}$ refers to convergence in probability.[34] In short, the QV process measures the realized sample-path variation of the squared return process. It immediately follows from the definition that for a time interval $(t, t+h]$ within the sample $[0, T]$, the increment of QV has the property:

[34] For a more rigorous definition of the quadratic variation or covariation process, we refer to Protter (1990, chapter 2).

$$[r,r]_{t+h} - [r,r]_t = \text{plim}_{M\to\infty} \sum_{m=0}^{M-1} [r^*(t_{m+1}) - r^*(t_m)][r^*(t_{m+1}) - r^*(t_m)]' \tag{10}$$

for any sequence of partitions $t_0 = t < t_1 < \cdots < t_M = t + h$ with $\sup_m\{t_{m+1} - t_m\} \to 0$ as $M \to \infty$. The sum on the right hand side of (10) is referred to as realized variation or covariation (RV) in the literature. Furthermore, the increment of QV is related to the conditional return covariance matrix (which is highly relevant in economic modeling and financial practice) by the following theorem (ABDL, 2003):

Theorem: Assume 1) the arbitrage-free logarithmic price process is square-integrable; and conditional on information at time t: 2) the mean return process $\{\mu_s - \mu_t\}_{s\in(t,t+h]}$ *is independent of the return innovation process* $\{M_u\}_{u\in(t,t+h]}$*; 3) the mean process* $\{\mu_s - \mu_t\}_{s\in(t,t+h]}$ *is a predetermined function over* $(t, t+h]$. *Then the increment of QV is an unbiased estimator of the return covariance matrix conditional on information at time t, that is,*

$$Cov[r^*(t+h) - r^*(t)|\mathcal{F}_t] = E\{[r,r]_{t+h} - [r,r]_t|\mathcal{F}_t\}, \tag{11}$$

where $\mathcal{F}_t$ *is the information filtration at time t.*

Although the assumptions in the theorem are somehow restrictive, they accommodate a variety of situations in the literature, such as a constant mean in the return process, deterministic intra-period variation in the conditional mean process, the asymmetric relation between returns and volatilities, etc. The drawback of the assumptions is that they exclude the feedback effects from the return innovation to the mean return. However, these feedback effects seem to be of trivial magnitude in practice, as discussed in ABDL (2003).

Combining (10) with (11), we can see that the RV serves as a desirable measurement of the conditional return covariance matrix, as long as the sampling frequency is high enough and the market microstructure noise is controlled.

Calculations

The above theory justifies using the RV in the construction of conditional volatilities and correlations. Specifically, in our case, $n = 4$ (assets), $h = 1$ (day), $M = 78$ (five-minute intervals), $T = 2178$ (days). The daily return over $(t-1, t]$ is calculated as $r_t = r^*(t) - r^*(t-1)$. And the RV at the daily interval $(t-1, t]$ is calculated as:

$$RV_{t-1,t} = \sum_{m=0}^{M-1} [r^*(t_{m+1}) - r^*(t_m)][r^*(t_{m+1}) - r^*(t_m)]', t = 1, 2, \cdots T, \tag{12}$$

where $t_0 = t - 1 < t_1 < \cdots < t_M = t$, and $t_{m+1} - t_m = \dfrac{1}{M}, \forall m = 0, 1, \cdots M - 1$. Furthermore, following ABDL (2001, 2003), for $t = 1, 2, \cdots, T$, we define daily realized variance $v_{i,t}^2 \equiv [RV_{t-1,t}]_{ii}$, standard deviation $v_{i,t} \equiv \sqrt{[RV_{t-1,t}]_{ii}}$, logarithmic standard deviation $lv_{i,t} \equiv \ln \sqrt{[RV_{t-1,t}]_{ii}}$, covariance $cov_{ij,t} \equiv [RV_{t-1,t}]_{ij}$, and correlation $cor_{ij,t} = \dfrac{cov_{ij,t}}{v_{i,t}v_{j,t}}$, where the subscripts $_{ii}$ and $_{ij}$ refer to the (i,i) and (i,j) element of a matrix. Obviously, $v_{i,t}^2$, $v_{i,t}$, and $lv_{i,t}$ provide appropriate measures of the conditional variances, standard deviations and logarithmic standard deviations of asset i returns, while $cov_{ij,t}$ and $cor_{ij,t}$ measure the conditional covariances and correlations between asset i and asset j returns.

Empirically, these measures have been applied to high-frequency foreign exchange rates (ABDL, 2001, 2003), and actively traded stocks such as the Dow Jones Industrial Average (DJIA) stocks (Andersen, Bollerslev, Diebold, and Ebens, ABDE, 2001). The common findings can be summarized as follows. In terms of unconditional distributions, $v_{i,t}^2$, $v_{i,t}$, $cov_{ij,t}$ are right-skewed and fat-tailed, while $lv_{i,t}$, $cor_{ij,t}$, and daily returns standardized by daily realized standard deviations $(r_{i,t}/v_{i,t})$ are approximately Gaussian. Regarding time series properties, $lv_{i,t}$ and $cor_{ij,t}$ are stationary but have strong persistence, which can largely be captured by a fractionally-integrated long memory process.

Data analysis of daily returns, volatilities, and correlations

Unconditional distributions

The upper panel in Table 4 summarizes the unconditional distributions of daily returns (r_t), as well as daily standardized returns (r_t/v_t) for the four ETFs/portfolios.[35] Generally, the means of r_t are around zero, and the standard deviations are larger than one. The daily US and EU returns are (slightly) left-skewed while the daily JP and AS returns are right-skewed. All the daily returns have fatter tails than the normal distribution with the mean kurtosis equal to 6.569. These properties are consistent with those reported in the literature (e.g., ABDE, 2001; ABDL, 2003).

On the contrary, the unconditional distributions of daily standardized returns are similar across the four ETFs/portfolios and are approximately standard normal. With the means still around zero, the standard deviations are now close to one. Meanwhile, the mean of skewness coefficients is reduced from 0.105 to -0.018, and the mean of kurtosis coefficients is remarkably reduced from 6.569 to 2.558. Allowing for certain sampling variation, these statistics suggest that the daily standardized returns approximately follow the standard normal distribution.[36]

The middle panel in Table 4 summarizes the unconditional distributions of daily realized standard deviations (v_t) and logarithmic standard deviations (lv_t). Generally, v_t is skewed to the right (with the minimal skewness coefficient equal to 1.566) and is very leptokurtic (with the minimal kurtosis coefficient equal to 9.720). In contrast, lv_t is only slightly skewed (with the maximal absolute skewness coefficient equal to 0.355) and much less leptokurtic than v_t (three out of four kurtosis coefficients are between 3 and 4). Therefore, although v_t is strikingly different from a normal distributed random variable, lv_t is approximately normally distributed.

The bottom panel in Table 4 reports the unconditional distribution statistics for daily realized covariances (cov_t) and correlations (cor_t). We report only the covariances and correlations between the US and other ETF/portfolios, as these may be of main interest to US investors. Other covariances and correla-

[35] For notational simplicity, we omit the subscript i (or j) when the symbol is self-evident.

[36] Under the null hypothesis that the returns are i.i.d. normally distributed, the sample skewness and kurtosis are asymptotically normal with means of 0 and 3, and variances of $\frac{6}{T}$ and $\frac{24}{T}$. Since $T = 2178$ in our case, the two standard errors are 0.052 and 0.105.

tions share similar patterns. From Table 4, the daily realized covariances are extremely skewed to the right (with the minimal skewness coefficient equal to 4.564) and extremely fat-tailed (with the minimal kurtosis coefficient equal to 42.320). In contrast, the corresponding daily realized correlations (cor_t) are approximately normal, with the skewness and kurtosis coefficients close to 0.5 and 3.5.

In summary, the unconditional distribution properties found in our study are in line with those found in foreign exchange rates and DJIA individual stocks. Specifically, the daily standardized returns, realized logarithmic standard deviations, and realized correlations are close to being normally distributed. For convenience of analysis, we thereafter focus our attention on these three quantities.

Time series properties

We now turn to the dynamic properties of the daily returns, volatilities, and correlations. Specifically, we examine the following three properties: temporal dependence, stationarity, and long memory. For temporal dependence, we test serial correlations in the time series using the Ljung-Box Q statistic. Under the null of no serial correlations up to lag k, Q_k is asymptotically Chi-square distributed with k degrees of freedom. We choose $k = 20$ (approximately one month) to take into account possible weekly and monthly seasonalities. With respect to stationarity, we employ both the augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) tests. In the ADF test, we include a constant term, a linear time trend, and k $(= 20)$ lagged difference terms of the dependent variable in the standard Dickey-Fuller regression. Similarly, in the PP test, we include a constant term, a linear time trend in the standard Dickey-Fuller regression, and calculate the residual spectrum at frequency zero using the Bartlett kernel with bandwidth $k + 1$ $(= 21)$. Under the null that the series contain a unit root, both tests follow a nonstandard distribution, with the critical values given by simulation results.

In terms of long memory properties, we estimate the fractionally integrated model:

$$(1 - L)^d(y_t - \mu) = \varepsilon_t, t = 1, 2, \cdots T, \tag{13}$$

where $\{y_t\}$ is the time series of interest, and $\{\varepsilon_t\}$ is a general short memory time series. The parameter d determines the memory properties of $\{y_t\}$: when $d \leq 0$, $\{y_t\}$ is a short memory stationary process with no serial correlations or with quick-decaying autocorrelations; when $0 < d < 0.5$, $\{y_t\}$ is a long memory stationary process with slow-decaying autocorrelations; when $d \geq 0.5$, $\{y_t\}$ is a long memory non-stationary process.[37]

We apply two methods to estimate d. The first is the modified log periodogram regression (MLP) by Kim and Phillips (1999a), which is a modified version of the log periodogram regression (LP) originally proposed by Geweke and Porter-Hudak (1983). Specifically, the LP estimator of d (denoted as d_{LP}) is based on the least squares regression:

$$\ln(I(\lambda_j)) = c_{LP} - d_{LP} \ln|1 - e^{i\lambda_j}|^2 + u_j, \tag{14}$$

where $I(\lambda_j)$ is the periodogram of $\{y_t\}$ at the the fundamental frequencies $\lambda_j = 2\pi j/T$ $(j = 1, \cdots, T^a)$, and c_{LP}, d_{LP} are the parameters to be estimated. And the MLP involves a similar regression:

$$\ln(I^*(\lambda_j)) = c_{MLP} - d_{MLP} \ln|1 - e^{i\lambda_j}|^2 + u_j, \tag{15}$$

where $I^*(\lambda_j) = v(\lambda_j)\overline{v(\lambda_j)}$, $v(\lambda_j) = w(\lambda_j) + \dfrac{e^{i\lambda_j}}{1 - e^{i\lambda_j}} \dfrac{y_T}{\sqrt{2\pi T}}$, $w(\lambda_j)$ is the discrete Fourier transform of $\{y_t\}$ at frequency λ_j, and $\overline{v(\lambda_j)}$ is the complex conjugate of $v(\lambda_j)$. Although both $\hat{d}_{LP}$ and $\hat{d}_{MLP}$ are similar in the stationary case ($d < 0.5$), $\hat{d}_{LP}$ compares less favorably to $\hat{d}_{MLP}$ when $0.5 \leq d < 1$ and is not consistent when $d > 1$ (Phillips 1999, Kim and Phillips 1999b). Therefore, we use the MLP (with $a = 0.75$), and report the point estimates as well as the 95% confidence intervals.[38]

With respect to the second method, we make use of a scaling law as observed in the fractionally integrated time series (e.g., Diebold and Lindner, 1996). Specifically, when y_t is fractionally integrated as modeled by (13), the h-fold partial sums, $[y_t]_h = \sum_{j=1,\cdots,h} y_{h(t-1)+j}$, obey a scaling law $var([y_t]_h) = c \cdot h^{2d+1}$.

Therefore, we run the regression:

$$\ln[var([y_t]_h)] = \alpha + \beta \ln(h) + u_h, h = 1, \cdots, H, \tag{16}$$

[37] For a detailed discussion of long memory processes, we refer to Beran (1994) and Robinson (2003).

[38] Nevertheless, we also use the log periodogram regression method formalized by Robinson (1995) and obtain similar estimation results for d. Details are available upon request.

and estimate the long memory parameter, $\hat{d}_S = (\hat{\beta} - 1)/2$. In our case, we choose $H = 30$ and report the point estimates and the adjusted R-squared.

Returns

The time series properties of daily standardized returns are summarized in the upper panel of Table 5. As shown by the Ljung-Box test statistics, the daily standardized returns are not serially correlated, which is in line with the efficient market hypothesis that the asset returns are generally not forecastable. The exception is the Europe portfolio return process, which has some evidence of serial correlations (at a 5% significance level). All the return series are stationary, as the unit root hypothesis is rejected by both the ADF test and the PP test at a 1 % significance level.

The above results are further corroborated by the long memory parameter estimation result. The point estimates of d_{MLP} and d_S are around zero for all ETFs/portfolios returns, indicating that the return series are short memory stationary. Moreover, the fact that $\bar{R}_s^2$s are nearly equal to one shows a remarkable goodness of fit in the scaling law regression. Combined with the Ljung-Box tests and the unconditional distribution properties (zero mean), the evidence shows that the daily standardized returns approximately follow white noise processes.

Volatilities and correlations

In contrast with the returns, the daily realized logarithmic standard deviations and realized correlations have very strong dynamic dependence, as seen from the middle and bottom panel of Table 5. The Q_{20} statistics are extremely large, with the means of 13041 and 282.2 for lv_t and cor_t, respectively. As a result, the null of no serial correlations up to lag order of 20 has uniformly been rejected. Despite this strong evidence of temporal dependence, however, the unit root is rejected for almost all daily realized volatilities and correlations (except for lv_t of the Asia ex-Japan Portfolio). These together point to stationary long memory processes, as further supported by the long memory estimation results. The estimates of d_{MLP} and d_S generally lie in the range $(0.3, 0.5)$ for lv_t (with the means of 0.448 and 0.425, respectively) and in the range $(0.1, 0.3)$ for cor_t (with the means of 0.117 and 0.171, respectively). Therefore, both lv_t and cor_t seem to follow fractionally integrated processes, and the former tends to be more persistent than the latter. Note that a larger value of d means more persistence.

To further illustrate the long memory behavior of the volatilities and correlations, we plot in Figure 3 the sample autocorrelation functions (ACF) up to lag order of 120 for lv_t and cor_t, as well as for the filtered series, $(1-L)^{d_{MLP}} lv_t$ and $(1-L)^{d_{MLP}} cor_t$.[39] The solid lines in the figure represent the ACF of the original series, while the dashed lines are the ACF of the filtered series. The dotted lines give the 95% confidence bands of an i.i.d. Gaussian process. As seen from the figure, the ACF of lv_t and cor_t are significantly positive even at the lag of 120 days, with lv_t showing much stronger persistence than cor_t. In contrast, the ACF of the filtered series are within the 95% confidence bands most of the time. This result shows that the fractional differencing operator has eliminated much of the temporal dependence in the daily realized logarithmic standard deviations and realized correlations. However, there are still significant autocorrelations, at least at the first one or two lags in the filtered series.

In summary, the daily standardized returns are found to approximately follow white noise processes, while the daily realized logarithmic standard deviations and realized correlations are long memory stationary. Applying the fractionally differenced operator $(1-L)^d$, where d is the long memory parameter estimated by either the MLP or the scaling law method, we can eliminate much but not all of the serial dependence in the realized logarithmic standard deviations and realized correlations.

Cross-market linkages in returns, volatilities and correlations

Taking into account the data properties observed in Section 4, we now investigate the linkages of daily returns, volatilities, and correlations across the United States (US), Japan (JP), Asia ex-Japan (AS), and Europe (EU) markets/regions. The investigation is carried out in two ways. First, we examine whether there are any significant in-sample return, volatility, and correlation spillovers across the four markets/regions. Second, we examine whether the in-sample spillovers help to forecast daily returns, volatilities, and correlations out-of-sample .

[39] Similar results are found for the filtered series $(1-L)^{d_S} lv_t$ and $(1-L)^{d_S} cor_t$.

While in-sample analysis has been frequently applied to investigate the cross-market linkages with relatively low frequency data (e.g., Hamao, Masulis, and Ng, 1990; Lin, Engle, and Ito, 1994; Craig, David, and Richardson, 1995), the out-of-sample forecastability has been less examined although it may be more relevant in a practical sense. Since the volatilities and correlations in our study are constructed from high-frequency, intra-day data approximately free of measurement error (as discussed in Section 3), we may treat them as directly observable and hence evaluate the out-of-sample forecastability accordingly. For a good reference of using realized volatility measures in out-of-sample forecasts, we refer to ABDL (2003).

In-sample spillovers

In view of the differences in time series properties, we model and estimate the in-sample return, volatility and correlation linkages separately. Meanwhile, we note that our sample period covers the Southeast Asia Crisis (1997) and the recent technology bubble (1996-2000).[40] Also, the trading of the ETFs was not highly active during the early period after the inception. These factors could lead to potential structural changes in the sample and affect the estimation results. Therefore, in addition to the full sample analysis, we do the same analysis over the sub-period from September 2000 to December 2004 (approximately half the size of the entire sample period), during which there was a dramatic decline in the US market.

Returns

Although the daily returns have fatter tails than the normal distribution, the daily standardized returns are approximately Gaussian (as described in Section 4). Meanwhile, the daily standardized returns are contemporaneously related, based on their sample correlation matrix (with the minimal correlation coefficient equal to 0.44; the table is available upon request). Therefore, we model the daily standardized returns as a Gaussian Vector Autoregressive (VAR) system, that is:

$$y_t = \alpha + \sum_{j=1}^{5} \beta_j y_{t-j} + \varepsilon_t, \tag{17}$$

[40] See Forbes and Rigobon (2002), and Ball and Torous (2004) for analysis of the cross-market linkages during the Southeast Asia Crisis. See also Brooks and Negro (2002), and Hon, Strauss, and Yong (2003) for effects of the technology bubble on the cross-market comovement.

where $y_t = \{r_{us,t}/v_{us,t}, r_{jp,t}/v_{jp,t}, r_{as,t}/v_{as,t}, r_{eu,t}/v_{eu,t}\}'$.[41] All the inverse roots of the characteristic AR polynomial lie inside the unit circle, indicating covariance stationarity. In addition, the VAR residuals reveal no significant autocorrelations based on the portmanteau tests (Lutkepohl, 1991, p. 150), which suggests that the lag order of 5 is adequate to accommodate the dynamics in the standardized returns system. To save space, we report only the coefficients for the first lags, which can be seen as measuring the overnight return spillovers, and the adjusted R-squared ($\bar{R}^2$). Meanwhile, we carry out pairwise Granger causality tests to see whether there are significant return spillovers (up to 5 lags) from one endogenous variable to another. Under the null of no significant spillovers, the test statistic is asymptotically Chi-square distributed with 5 degrees of freedom. In addition, we test to determine if there are significant spillovers (up to 5 lags) from all other lagged endogenous variables, in which case the Granger causality test statistic is asymptotically Chi-square distributed with 15 ($= 5 \times 3$) degrees of freedom. To save space, we report only the p-values for the Granger causality tests.

The results are reported in Table 6a. Three specific patterns are observed. First, the standardized returns generally demonstrate reversals over a one-day horizon, as can be inferred from the negative sign of the AR(1) coefficients. (The only exception is the EU returns, which have positive but insignificant AR(1) coefficients.) The reversals in asset returns have been documented in the literature and may be explained by the "stock market overreaction" hypothesis (e.g., DeBondt and Thaler, 1985; Lo and MacKinlay, 1999). According to this explanation, investors are subject to waves of optimism and pessimism and tend to overreact to unexpected and dramatic news. As an empirical implication, the asset returns must be negatively autocorrelated for some holding period.[42]

Second, there is some evidence of overnight return spillovers across markets/regions, though the signs might differ. On one hand, positive spillovers are observed between the US and EU markets/regions. For example, over the full sample, the coefficient for the first lag of EU (US) in the US (EU) return equation is 0.05 (0.21) at a 1% significance level. In an economic sense, it means that a 1% increase in the EU (US) standardized return leads to a 0.05% (0.21%) increase in the following day's US (EU) standardized return

[41] The Akaike and Schwarz criteria choose the lag order of 5 and 1, respectively. Considering our large sample size, we choose the lag order of 5 (approximately one week) to maintain conservatism.

[42] See Lo and MacKinlay (1999, chapter 5) for a more detailed discussion.

(all other things being equal). On the other hand, negative spillovers are found from JP to the US and EU. For example, over the full sample, the coefficient for the first lag of JP in the US (EU) return equation is -0.08 (-0.16) at a 5% (1%) significance level. That is, a 1% increase in the JP standardized returns leads to a 0.08% (0.16%) decrease in the following day's US (EU) standardized returns (all other things being equal). It remains uncertain what causes the difference in the signs of return spillovers.

Third, the in-sample return spillovers (up to 5 lags) generally disappear in the sub-sample period. This can be seen from the results of pairwise Granger causality tests. Over the full sample, there are significant spillovers from all other lagged returns to the US returns (at a 2% level), of which the spillovers from the AS returns play an important role (at a 3% level). Similarly, significant spillovers from all other lagged returns to the EU returns (at a 1% level) are observed, where the spillovers from the US and JP returns (at a 1% level) play the dominant role. Over the sub-sample, however, no significant spillovers are found.

The change of results from the full sample to the sub-sample is consistent with a basic prediction of efficient markets. Specifically, during the initial period after the inception of ETFs, the investors knew little about these new products and therefore might not have taken positions in the ETFs. The thin trading led to market inefficiency in that the ETF returns might be forecastable from cross-market linkages (and its own history). However, as the investors were well informed about the ETFs, they were more involved in trading the ETFs and exploring possible arbitrage opportunities. These activities eventually weakened or annihilated the cross-market linkages in returns.[43]

Despite these three patterns, the daily standardized returns of the four markets/regions are hardly explained by either the spillover effects or its own lags, as reflected by the negligible magnitude of adjusted R-squared (with the maximum of 0.02). It is consistent with our finding in Section 4 that daily standardized returns are short memory stationary (approximately white noise processes). Combined with the out-of-sample evidence below, our result is in line with previous findings in the literature that the returns are generally not forecastable.

Volatilities

Based on the time series properties discussed in Section 4, we model daily realized logarithmic standard

[43] Lo and MacKinlay (1999, chapter 1) have provided a more detailed explanation along this line.

deviations as a long memory Gaussian VAR system, that is:

$$(1-L)^d y_t = \alpha + \sum_{j=1}^{5} \beta_j (1-L)^d y_{t-j} + \varepsilon_t, \tag{18}$$

where $y_t = \{lv_{us,t}, lv_{jp,t}, lv_{as,t}, lv_{eu,t}\}'$, and d is the long memory parameter vector for y_t estimated from the MLP method, i.e., $d = \{d_{us}, d_{jp,} d_{as,} d_{eu}\}'_{MLP}$, and the product of $(1-L)^d$ and y_t is calculated through element-by-element multiplication.[44] The fractional difference filter $(1-L)^d$ is used to capture the long memory in the $\{lv_t\}$ process, while the VAR captures any remaining dynamics.[45] The long memory Gaussian VAR model is validated by the characteristic AR polynomial results and the portmanteau test results. Again, we report only the coefficients for the first lags (measuring the overnight volatility spillovers) and the pairwise Granger causality results.

The results are reported in Table 6b. Several patterns are observed, although the full-sample results are slightly different from the sub-sample results. First, self memory (both long and short memory) dominates the volatility dynamics. The long memory parameter captures the main dynamics in the volatility process (see Section 4, Figure 3). And as indicated by the magnitude of the AR(1) coefficient (relative to that of the spillover coefficients) in each equation, the short memory plays the dominant role in the filtered volatility processes. Note that negative AR(1) coefficients do not conflict with the volatility clustering observed in the data, since they are applied to the filtered instead of the original volatility series (ABDE, 2001).

Second, there are asymmetric overnight volatility spillovers, which are mainly driven by the US information. Specifically, there are significant overnight volatility spillovers from the US market to all other three markets/regions, but not in the opposite directions. For example, over the full sample, the coefficients for the first lag of the US in the JP, AS, EU volatility equations are 0.07 (at a 1% significance level), 0.06 (at a 5% significance level), 0.08 (at a 5% significance level), respectively, with the magnitude only smaller than that of the corresponding AR(1) coefficients.[46] In the US volatility equation, however, no significant

[44] Similar results are obtained using the long memory parameters estimated from the scaling law method (d_S).

[45] The Akaike and Schwarz criteria choose the lag order of 4 and 1, respectively. Considering our large sample size, we employ VAR(5) to maintain conservatism.

[46] The values of AR(1) and spillover coefficients should be interpreted carefully since they are applied to the filtered, instead of the original volatilities.

overnight spillovers are found from any other markets/regions. This evidence corroborates the finding in Hamao, Masulis, and Ng (1990), as summarized in the introduction. In addition, there are significant overnight volatility spillovers between AS and JP in the full sample (with both spillover coefficients equal to 0.05 at a 5% significance level), as well as from EU to AS (with spillover coefficients equal to 0.04 and 0.06 at a 5% significance level in the full and sub-sample, respectively).

The interpretation of the asymmetry in the volatility spillovers is uncertain at this stage. It could be explained by the public information hypothesis, as discussed in Becker, Finnerty, and Friedman (1995, p. 1192). According to this explanation, "because the US is the dominant producer of goods and services in the world economy, the US is also the most important producer of information. In addition, US investors will possess a more provincial view and ignore information from other countries." Therefore, the volatilities in the US stock markets (as caused by the US news) tend to lead those in other markets/regions. Alternatively, the asymmetry may be due to market frictions in the international ETFs. As reported in Table 2, the international ETFs are traded much less frequently than the US ETF. Accordingly, the bid-ask spreads of the international ETFs are much wider than those of the US ETF (generally more than ten-fold in magnitude, not reported here). These market frictions may cause international ETFs to slowly respond to the US news, say, with one day lag or longer. On the other hand, the high liquidity and small bid-ask spreads in the US ETF market enable the US ETF to immediate incorporate the news from other markets, say, within a day or even an hour. Thus, we find significant inter-day information spillovers from the US to other ETFs/ETF portfolios, while no inter-day spillovers in the opposite direction. Formal investigation is needed of the causes of these volatility spillovers. Not surprisingly, the bilateral overnight spillovers between AS and JP could be attributed to the close tie between these two economies.

Third, cross-market spillovers provide strong forecastability in the JP and AS volatilities, as indicated by the Granger causality results. Specifically, the lagged volatilities of all other markets/regions (up to 5 lags) Granger cause the JP and AS volatilities at a 1% level in both full and sub-sample. This evidence suggests the importance of appreciating the cross-market linkages in volatility forecasts. We will further explore this point in the out-of-sample analysis.

Correlations

Correlations have frequently been utilized to examine international transmission mechanisms using low-frequency return data (e.g., King and Wadhwani, 1990; Lee and Kim, 1993; Calvo and Reinhart, 1996; See Claessens and Forbes, 2001, for an excellent survey of recent empirical papers). The realized correlation method using high-frequency data, however, is relatively less employed in this literature. A related study is done by ABDE (2001), where they use the daily realized correlations to examine whether past volatilities have a larger impact on asset return correlations when the markets are in a downturn.

As in the volatility modeling, we model the daily realized correlations (cor_t) as a long memory Gaussian VAR system, that is:

$$(1-L)^d y_t = \alpha + \sum_{j=1}^{5} \beta_j (1-L)^d y_{t-j} + \varepsilon_t, \tag{19}$$

where $y_t = \{cor_{us,jp,t}, cor_{us,as,t}, cor_{us,eu,t}\}'$, $d = \{d_{us,jp}, d_{us,as}, d_{us,eu}\}'_{MLP}$.[47] Again, the model is validated by the characteristic AR polynomial results and the portmanteau test results. We report the coefficients for the first lags (measuring overnight correlation spillovers) and the pairwise Granger causality results.

These results are reported in Table 6c. A large part of the correlation dynamics is captured by their own long memory (as discussed in Section 4). The VAR system explains only a small portion of the correlation variations, as reflected by the marginal magnitude of $\bar{R}^2$ (with the mean of 0.016). Nonetheless, the VAR system successfully removes the remaining dynamics in the filtered correlation series, as the VAR residuals reveal no significant autocorrelations based on the portmanteau tests.

With regard to correlation linkages, we observe positive spillovers from the US-EU to the US-AS correlations. Specifically, the overnight spillover coefficients are 0.06 (at a 1% significance level) and 0.09 (at a 5% significance level) in the full and sub-sample, respectively. This lead-lag relation is further confirmed by the Granger causality results, where the US-EU correlations Granger cause the US-AS correlations at a 5% and a 1% significance level in the full and sub-sample, respectively. As a result, there is a significant predictive enhancement in the US-AS correlation from all cross-market spillovers (a 1% level in both the full and sub-sample). Besides, we also observe bilateral causality between the US-JP and the US-AS in the sub-sample.

[47] Both the Akaike and Schwarz criteria choose the lag order of 1 and we employ VAR(5) to maintain conservatism.

The correlation spillovers from the US-EU to the US-AS could possibly be due to timezone difference in the underlying stock markets. As documented in Hamao, Masulis, and Ng (1990, Figure 1), the US market opens in late afternoon for the European (London) market, resulting in concurrent trading in both markets for about two-and-a-half hours (ignoring the differences in daylight savings time). The Asian (Tokyo) market, on the other hand, is closed during the US trading hours. Therefore, when some common shock occurs during the US trading period, the US-EU correlations could be immediately impacted, while the effect on the US-AS correlations might not be observed until the next day. As a result, we observe a lead-lag relation from the US-EU to the US-AS. The drawback for this interpretation is that it could not explain why the US-EU correlations do not lead the US-JP correlations. Formal investigation may need to take into account the trading activity of each market, investors' expectations, etc.

Out-of-sample forecastability

In the out-of-sample analysis, we examine whether the in-sample cross-market linkages help to improve the forecastability of returns, volatilities and correlations out-of-sample. There has been a growing amount of literature discussing the pros and cons of the in-sample and out-of-sample analysis (e.g., Granger, 1990; Inoue and Kilian, 2005). We view here the in-sample linkages and out-of-sample forecastability as mutually complementary, if not two sides of the same coin. On one hand, the discovered in-sample linkages may not enhance the forecasts significantly as their explanatory power may be marginal (as indicated by small $\bar{R}^2$). In this case, although the in-sample linkages help us understand the market relations, they may not be of much value in financial practice where forecasts are of more concern. On the other hand, even though some in-sample linkages are not significant, they may be relevant in improving forecasts if they contain useful information of missing variables in the true forecast model.

With respect to the out-of-sample methodology, there are three prevalent schemes: recursive, rolling, and fixed.[48] We use the recursive scheme in this paper, and qualitatively similar results are found for the other two methods. Specifically, we divide the whole sample $\{y_t\}$ of size T into two sub-samples of size R and P, where $R = P = \frac{T}{2}$. For each model (described below), we use the estimation result from

[48] See McCracken (2004) for description and comparison of these three schemes.

the first sub-sample to form a one-day-ahead forecast, denoted as $\hat{y}_{R+1}$. We calculate the forecast error at $T+1$ as $\hat{e}_{R+1} = y_{R+1} - \hat{y}_{R+1}$. Then we expand the estimation window by including y_{R+1} and run the regression on the increased sample and form the forecast error $\hat{e}_{R+2} = y_{R+2} - \hat{y}_{R+2}$. We repeat these steps until the last forecast error, $\hat{e}_{R+P} = y_{T} - \hat{y}_{T}$, is obtained. We then calculate the mean squared errors as $MSE = P^{-1} \sum_{t=1}^{P} \hat{e}_{R+t}^2$.

In accordance with our in-sample analysis, the models used in out-of-sample forecasts are as follows:

$$
\begin{aligned}
H_0 &: y_{i,t} = \alpha + \sum_{j=1}^{5} \beta_j^i y_{i,t-j} + \varepsilon_{i,t}; \\
H_a^k &: y_{i,t} = \alpha + \sum_{j=1}^{5} \beta_j^i y_{i,t-j} + \sum_{j=1}^{5} \beta_j^k y_{k,t-j} + \varepsilon_{i,t}, k \neq i, k = 1, \cdots, n; \\
H_a^{all} &: y_{i,t} = \alpha + \sum_{k=1}^{n} \sum_{j=1}^{5} \beta_j^k y_{k,t-j} + \varepsilon_{i,t},
\end{aligned}
$$

where y_t refers to r_t/v_t, $(1-L)^{d_{MLP}} lv_t$, $(1-L)^{d_{MLP}} cor_t$, accordingly; $n = 4$ for r_t/v_t, $(1-L)^{d_{MLP}} lv_t$, and $n = 3$ for $(1-L)^{d_{MLP}} cor_t$. In other words, we use H_0 (AR(5) model) as the benchmark model for return, volatility and correlation forecasts. We then examine the forecastability from the kth market spillovers by comparing the MSE of H_0 and H_a^k. In addition, we examine the forecastability from all possible cross-market linkages by comparing the MSE of H_0 and H_a^{all}. Therefore, there are four alternatives to the benchmark model in each return (or volatility) forecast, and three alternatives in each correlation forecast.

To test the significance of cross-market forecastability, we construct the Diebold-Mariano (DM) t-statistic as:

$$
DM = \sqrt{P} \frac{\bar{d}}{\sqrt{\hat{\omega}_d}},
$$

where $\bar{d}$ is the difference between the MSE of H_0 and H_a^k (or H_a^{all}), and $\hat{\omega}_d$ is the heteroscedasticity consistent variance estimator of $\bar{d}$. Under the null that the alternative model does not improve the out-of-sample forecastability over the benchmark model, DM t-statistic follows an asymptotically normal distribution (Diebold and Mariano, 1995). However, since here H_0 is nested in H_a^k (or H_a^{all}), the asymptotic distribution of DM t-statistic is nonstandard unless $P/R \to 0$ as $T \to \infty$ (see Clark and McCracken, 2001; McCracken, 2004). In view of this fact, we use the critical values calculated numerically by McCracken (2004). As shown by the Monte Carlo and empirical evidence in McCracken (2004), the critical values provide accurately sized and powerful tests for forecast comparison among nested models.

The results are reported in Table 7. For daily standardized returns, there is no significant out-of-sample forecastability from cross-market linkages. Indeed, including the cross-market information usually does worse than the benchmark model, as reflected by the negative sign of DM t-statistics. Note that negative DM t-statistic means larger MSE of H_a^k (H_a^{all}) than that of H_0. This finding confirms our previous result that daily standardized returns are generally not forecastable.

In contrast, there is strong evidence of predictive enhancement from cross-market linkages in volatilities. Specifically, the volatility spillovers from US improve both the JP and AS volatility forecasts at a 5% significance level. Meanwhile, the information from JP (AS) enhances the forecastability in AS (JP) volatility at a 1% (5%) significance level. As a result, the information from all other markets/regions yields superior forecast performance in the JP and AS volatilities at a 1% significance level. No significant forecastability is found for other cross-market volatility linkages.

With regard to correlations, we find that the US-AS correlations improve the forecastability in the US-JP correlations at a 1% significance level. In addition, for the US-JP correlation forecasts, the model taking account of all cross-market information outperforms the benchmark model at a 5% level. Meanwhile, the lagged US-JP correlations improve the US-EU forecasts at a 5% level, but all the cross-market information together does not yield a significant improvement in forecasting the US-EU correlations.

In summary, the out-of-sample results are partially consistent with the in-sample evidence. The weak in-sample return spillovers do not help to forecast daily standardized returns out-of-sample, consistent with the common wisdom that returns are generally not forecastable. The in-sample volatility spillovers from the US, as well as bilateral spillovers between JP and AS, yield a significant enhancement in forecasting the JP and AS volatilities. The in-sample correlation spillovers from the US-AS to the US-JP improve the predictive performance in the US-JP correlations. Other discovered in-sample volatility and correlation linkages, however, do not provide significant improvement in out-of-sample forecasts.

Extensions

The cross-market linkages discussed in the preceding section are examined in a simplified framework, that is, without considering possible exogenous variables. In this section, additional tests of related hypotheses are performed.

The first set of tests concerns the Monday effects in daily returns, volatilities, and correlations. On one hand, negative mean returns are documented for US stocks on Mondays (e.g., French, 1980; Gibbons and Hess, 1981). Similar evidence is also found for international stocks (e.g., Jaffe and Westerfield, 1985; Condoyanni, O'Hanlon, and Ward, 1988). On the other hand, the return volatilities of US stocks are found to be higher on Mondays (Fama, 1965; Godfrey, Granger, and Morgenstern, 1964). Since correlations are closely related to volatilities, it is natural to suspect that similar seasonality might also exist in daily correlations. Therefore, we add a dummy variable D_t for the day following a weekend or holiday in all return, volatility, and correlation equations to examine possible seasonality. [49]

The second set of tests deals with the leverage effects of daily returns on daily volatilities. It is a well-known stylized fact that negative returns have a larger impact on future volatilities than do positive returns of similar magnitude, as first discussed by Black (1976).[50] In a recent work, ABDE (2001) used daily realized logarithmic standard deviations to examine this effect and find statistically (but not economically) significant volatility asymmetry for most DJIA stocks. Motivated by their methodology, we include in each equation in the volatility VAR system the following extra terms $(1-L)^d x_{t-1}$, where $x_t = \{lv_{us,t} I(r_{us,t} < 0), lv_{jp,t} I(r_{jp,t} < 0), lv_{as,t} I(r_{as,t} < 0), lv_{eu,t} I(r_{eu,t} < 0)\}'$, $d = \{d_{us}, d_{jp}, d_{as}, d_{eu}\}'_{MLP}$, and $I(\cdot)$ is the indicator function. Note that we here allow for more general leverage effects, that is, not only the asset's own returns but also other assets' returns may have asymmetric effects on the asset's volatilities.

Finally, we check whether there are contagion effects in daily correlations. As defined in Forbes and Rigobon (2002), contagion refers to "a significant increase in cross-market linkages after a shock to one country (or group of countries)." To test this hypothesis, cross-market return correlations are often calculated for a stable period and then compared with the return correlations calculated after a shock. If there

[49] We also study the Monday effects by isolating the Monday effects from the day-after-holiday effects and obtain similar results. The tables are available upon request.

[50] It is under discussion whether this volatility asymmetry is due to the leverage effect as explained by Black, or due to a volatility feedback as discussed by Campbell and Hentschel (1992).

was a significant rise in correlations after the shock, then contagion occurred (e.g., King and Wadhwani, 1990; Lee and Kim, 1993; Calvo and Reinhart, 1996). One drawback for this methodology is that it does not allow for the time-varying property in return correlations. In fact, as with return volatilities, the return correlations are stochastic and change over time (e.g., Dacorogna et al., 2001, chapter 10). Ignoring the stochastic property in correlations, therefore, may lead to an over-rejection of the null of no contagion, as discussed in Ball and Torous (2004). The realized correlation, by construction, takes into account the stochastic nature in correlations. In addition, the high-frequency feature in this measure enables us to detect contagion (if there is any) within a much shorter period.

Specifically, we include in each equation in the correlation VAR system the following extra terms z_{t-1}, where $z_t = \{I(r_{us,t} < r_{us,(q)}), I(r_{jp,t} < r_{jp,(q)}), I(r_{as,t} < r_{as,(q)}), I(r_{eu,t} < r_{eu,(q)}\}'$, the subscript $_{(q)}$ denotes the qth quantile. Therefore, the extra terms z_{t-1} are the dummy variables for those markets/regions experiencing a dramatic downturn. To our knowledge, there is no economic theory to guide the choice of the threshold quantile q. We therefore experiment with various values from the $\frac{5}{1000}th$ quantile to the $\frac{1}{2}th$ quantile (median). The results are similar and we report the results with $q = \frac{1}{100}$.

Table 8 shows the results for the three sets of tests.[51] With respect to daily standardized returns, there are generally no significant Monday effects. The only exception is the AS market in the full sample, with the coefficient equal to -0.08 at a 5% significance level. The difference between our result and previous findings may be due to the high-frequency data and methodology employed here. Alternatively, it could be that the Monday effects in returns have weakened or disappeared as practitioners implement strategies to take advantage of this anomaly, as argued in Schwert (2002).

With respect to daily volatilities, significant Monday effects are observed in all markets/regions. Interestingly, in contrast with other markets/regions, where significantly positive Monday effects are found, the US market has significantly lower volatilities on Mondays.[52] In addition, we find significant leverage effects of the US returns on the US, AS, and EU volatilities in the full sample. That is, negative US returns are likely to increase the following day's return volatilities in the US, AS, and EU markets/regions. Significant leverage effects are also observed for the AS returns on the US and AS volatilities (full sample), as well as

[51] To save space, we report only the coefficients measuring the Monday, leverage, and contagion effects. Detailed estimation results are available upon request.

[52] This is contrary to findings of Fama (1965) and Godfrey, Granger, and Morgenstern (1964), but is in accordance with evidence in Halil Kiymaz and Hakan Berument (2002), where the lowest return volatility is observed on Mondays for the US.

for the EU returns on the US (sub-sample) and AS (full sample) volatilities.

In terms of daily correlations, the evidence of Monday effects is mixed. Specifically, significantly positive Monday effects are found in the US-AS correlations (sub-sample) as well as the US-EU correlations (full and sub-sample), but not in other cases. In addition, we find no evidence of contagion in all markets/regions, as all the contagion coefficients are insignificant. The lack of contagion evidence could be due to the fact that there are no extreme shocks such as the 1987 US market crash during the sample period we examine.[53] Alternatively, it could be the case that there is indeed no contagion effect at all, as discussed in Forbes and Rigobon (2002).

Conclusion

In this paper we examine the cross-market linkages in returns and volatilities over the period 1996-2004, using high-frequency intra-day transaction data of the exchange-traded funds (ETFs) that track the S&P 500 index and fourteen international stock indices. To overcome inactive intra-day trading of the international ETFs, we focus on the linkages across the United States, Japan, and two regions: Asia ex-Japan and Europe. The high-frequency feature in the data enables us to construct model-free estimates of daily volatilities and correlations with statistically high precision, as suggested by the theory of quadratic variation. This allows us to analyze the properties of daily returns, volatilities, and correlations as they are directly observable and use these quantities to model and test the in-sample cross-market spillovers and out-of-sample forecastability.

We find that our constructed measures (of daily returns, volatilities, and correlations) share very similar properties to those constructed from high-frequency exchange rates or Dow-Jones Industrial Average individual stock prices, as documented in Andersen, Bollerslev, Diebold, and Labys (2001, 2003), and Andersen, Bollerslev, Diebold, and Ebens (2001). In terms of unconditional distributions, daily realized variances and covariances are right-skewed and leptokurtic, while daily returns standardized by daily realized standard deviations, daily realized logarithmic standard deviations, and daily realized correlations are

[53] Although the 1997 Southeast Asia crisis is covered, the crisis may not affect the countries/regions outside that area (e.g., Ball and Torous, 2004).

approximately Gaussian. With respect to time series properties, daily realized logarithmic standard deviations and realized correlations are long memory stationary, which may be largely captured by fractionally integrated processes.

We observe weak cross-market linkages in daily standardized returns and correlations, as well as strong linkages in daily realized volatilities. Specifically, there is some evidence of in-sample return spillovers over the entire sample period (May 1996-December 2004). However, the discovered cross-market return linkages either disappeared or diminished over the sub-period (September 2000-December 2004). Moreover, the cross-market return spillovers, if there are any, do not help to forecast the daily returns out-of-sample. We view this evidence as consistent with a basic prediction of efficient markets. In contrast, there are significant in-sample volatility spillovers from the US market to the other markets/regions, but not in the opposite direction. This finding corroborates the evidence in Hamao, Masulis, Ng (1990), but is contrary to the finding in Lin, Engle, and Ito (1994). Moreover, the discovered volatility spillovers significantly improve the out-of-sample forecastability in the Japan and Asia ex-Japan volatilities. This suggests the importance of incorporating cross-market linkages in volatility forecasting. In terms of correlation linkages, we find the US-Europe daily correlations lead the US-Asia ex-Japan correlations. The discovered correlation linkages, however, do not significantly improve the forecastability in the US-Asia ex-Japan correlations.

We further check the Monday effects (in daily returns, volatilities, and correlations), the leverage effects (in daily volatilities), and the contagion effects (in daily correlations). We find no significant Monday effects in daily returns, weak evidence of positive Monday effects in daily correlations, and strong evidence of Monday effects in daily volatilities. Interestingly, in contrast with other three markets/regions, the US volatilities are found to be lower on Monday or the day after a holiday. Regarding the leverage effects, negative US returns are likely to increase the following day's volatilities in three out of the four markets/regions examined. Finally, we find no contagion effects in daily correlations; that is, there is no significant rise in correlations after a shock to a market. This could be seen as evidence supporting Forbes and Rigobon (2002).

The results in this paper suggest at least three avenues of future research. First, our volatility measures are based on the realized volatility method, since this method is free of model-misspecification. It is interesting to see if our results hold under parametric GARCH models as employed in the previous literature. By directly comparing the results from both methods, we may get a better sense of the gains from high-frequency sampling. This approach is pursued in Yang (2006).

Second, although we detect significant cross-market linkages, we remain uncertain about the forces behind them. It is conceivable that the US macro news plays an important role because all the ETFs are traded on the US market. However, since the ETFs are special instruments that track national stock indices, the underlying countries' economic fundamentals, the exchange rate dynamics, as well as the investors' expectations, all may contribute to the observed patterns as well. To disentangle these intervening factors, we need to add actual information flows (e.g., macroeconomic news) to sort out what the markets are responding to. Again, the high-frequency feature in our data is essential for such analysis.

Lastly, our study in this paper mainly concerns the short-run cross-market linkages. It is of interest to examine the cross-market linkages at a longer horizon, such as one week or one month. These may help us to detect the trend in international financial integration. And these linkages may provide useful information to investors of medium and long horizons. In addition, since the number of observations increases within each week or month, we could apply the realized volatility and correlation method to a wider range of individual markets (such as the UK, Germany, Hong Kong, Singapore, etc.), that may be of interest to a particular group of theorists and investors. Moreover, we could incorporate economic variables that are only available at low frequencies, such as industrial production, inflation, and international trade account, into the model as to examine the macro factors behind the cross-market linkages. We leave these for future work.

REFERENCES

Aggarwal, R., and R. Pietra, 1989, "Seasonal and Day-of-the-Week Effects in Four Emerging Stock Markets," *Financial Review,* 24 (4), 541-550.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics,* 61, 43-76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association,* 96, 42-55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica,* 71, 579-625.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega, 2004, "Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets," working paper, National Bureau of Economic Research.

Andersen, T. G., T. Bollerslev, and F. X. Diebold, 2005, "Parametric and Nonparametric Measurement of Volatility," in Yacine Aït-Sahalia and Lars Peter Hansen (eds.), *Handbook of Financial Econometrics*, North Holland, Amsterdam.

Back, K., 1991, "Asset Prices for General Processes," *Journal of Mathematical Economics,* 20, 317-395.

Ball, C. A., and W. N. Torous, 2004, "Contagion in the Presence of Stochastic Interdependence," working paper, Vanderbilt University.

Bandi, F. M., and J. Russell, 2003, "Volatility or Microstructure Noise?" working paper, Graduate School of Business, The University of Chicago.

Barndorff-Nielsen, O. E., and N. Shephard, 2004, "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica,* 72*,* 885-925.

Bae, K., and G. A. Karolyi, 1994, "Good News, Bad News and International Spillovers of Stock Return Volatility between Japan and the U.S.," *Pacific-Basin Finance Journal*, 2, 405-438.

Becker, K. G., J. E. Finnerty, and J. Friedman, 1995, "Economic News and Equity Market Linkages between the U.S. and U.K.," *Journal of Banking and Finance,* 19, 1191-1210.

Beran, J., 1994, *Statistics for Long Memory Processes*, Chapman and Hall, New York.

Black, F., 1976, "Studies of Stock Market Volatility Changes," *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 177-181.

Brooks R., and M. D. Negro, 2002, "The Rise in Comovement across National Stock Markets: Market Integration or IT Bubble?" working paper, Federal Reserve Bank of Atlanta.

Calvo, S., and C. M. Reinhart, 1996, "Capital Flows to Latin America: Is There Evidence of Contagion Effects?" in G. A. Calvo, M. Goldstein and E. Hochreiter (eds), *Private Capital Flows to Emerging Markets After the Mexican Crisis,* Institute for International Economics, Washington, DC.

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, 281-318.

Chao, J., V. Corradi, and N. R. Swanson, 2001, "An Out of Sample Test for Granger Causality," *Macroeconomic Dynamics,* 5, 598-620.

Clark, T. E., and M. W. McCracken, 2001, "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics,* 105, 85-110.

Claessens, S., and K. J. Forbes, 2001, *International Financial Contagion*, Kluwer Academic Publishers, Norwell, MA.

Condoyanni, L., J. O'Hanlon, and C. Ward, 1988, "Weekend Effects in Stock Market Returns: International Evidence," in E. Dimson (ed.), *Stock Market Anomalies,* Cambridge University Press, Cambridge.

Craig, A., A. David, and M. Richardson, 1995, "Market Efficiency around the Clock: Some Supporting Evidence Using Foreign-Based Derivatives," *Journal of Financial Economics,* 39, 161-180.

Dacorogna, M. M., R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet, 2001, *An Introduction to High-Frequency Finance*, Academic Press, San Diego.

DeBondt, W., and R. Thaler, 1985, "Does the Stock Market Overreact?" *Journal of Finance,* 40, 793-805.

Diebold, F. X., and P. Lindner, 1996, "Fractional Integration and Interval Prediction," *Economics Letters,* 50, 305-313.

Diebold, F. X., and R. S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* 13, 134-144.

Fama, E., 1965, "The Behavior of Stock Market Prices," *Journal of Business,* 38, 34-105.

Forbes, K., and R. Rigobon, 2002, "No Contagion, Only Interdependence: Measuring Stock Market Comovement," *Journal of Finance,* 43, 2223-2261.

French, K., 1980, "Stock Returns and the Weekend Effect," *Journal of Financial Economics,* 8, 55-69.

Geweke, J. and S. Porter-Hudak, 1983, "The Estimation and Application of Long-Memory Time Series Models," *Journal of Time Series Analysis,* 4, 221-238.

Gibbons, M. R., and P. Hess, 1981, "Day of the Week Effects and Asset Returns," *Journal of Business,* 54, 579-596.

Godfrey, M., C. Granger, and O. Morgenstern, 1964, "The Random Walk Hypothesis of Stock Market Behavior," *Kylos,* 17, 1-30.

Granger, C. W. J., 1990, *Modeling Economic Time Series: Readings in Econometric Methodology*, Oxford University Press, Oxford, UK.

Guidolin, M., and A. Timmermann, 2004, "International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences," working paper, Federal Reserve Bank of St. Louis.

Hamao, Y., R. W. Masulis, and V. Ng, 1990, "Correlations in Price Changes and Volatilities across International Stock Markets," *Review of Financial Studies,* 3, 281-307.

Hansen, P. R., and A. Lunde, 2004, "An Unbiased Measure of Realized Variance," working paper, Stanford University.

Hon, M., J. Strauss, and S. Yong, 2003, "Deconstructing the Technology Bubble: An Analysis of International Stock Comovements," working paper, Saint Louis University.

Inoue, A. and L. Kilian, 2005, "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?", *Econometric Reviews*, 23(4), 371-402.

Jaffe, J. and R. Westerfield, 1985, "The Weekend Effect in Common Stock Returns: The International Evidence," *Journal of Finance,* 40, 433–454.

Jares, T. E., and A. M. Lavin, 2004, "Japan and Hong Kong Exchange-Traded Funds (ETFs): Discounts, Returns, and Trading Strategies," *Journal of Financial Services Research,* 25, 57-69.

Karolyi, G. A., 1995, "A Multivariate GARCH Model of International Transmission of Stock Returns and Volatility: The Case of the United States and Canada," *Journal of Business & Economic Statistics,* 13, 11-25.

Kim, C. S., and P. C. B. Phillips, 1999a, "Modified Log Periodogram Regression," mimeo, Yale University.

Kim, C. S., and P. C. B. Phillips, 1999b, "Log Periodogram Regression: The Nonstationary Case," mimeo, Yale University.

King, M. A., and S. Wadhwani, 1990, "Transmission of Volatility between Stock Markets," *Review of Financial Studies,* 3, 5-33.

King, M. A., E. Sentana, and S. Wadhwani, 1994, "Volatility and Links between National Stock Markets," *Econometrica,* 62, 901-933.

Kiymaz, H., and H. Berument, 2002, "The Day of the Week Effect and Stock Market Volatility: Evidence from Developed Markets," working paper, University of Houston-Clear Lake.

Koutmos, G., and G. G. Booth, 1995, "Asymmetric Volatility Transmission in International Stock Markets," *Journal of International Money and Finance,* 14, 747-762.

Lee, S. B., and K. J. Kim, 1993, "Does the October 1987 Crash Strengthen the Comovements among National Stock Markets?" *Review of Financial Economics,* 3, 89-102.

Lin, W., R. F. Engle, and T. Ito, 1994, "Do Bulls and Bears Move across Borders? International Transmission of Stock Returns and Volatilities," *Review of Financial Studies,* 7, 507-538.

Lo, A. W., and A.C. MacKinlay, 1999, *A Non-Random Walk down Wall Street.* Princeton University Press, Princeton.

Lutkepohl, H., 1991, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, New York.

McCracken, M. W., 2004, "Asymptotics for Out of Sample Tests of Granger Causality," working paper, University of Missouri-Columbia.

Martens, M., D. V. Dijkz, and M. D. Pooterx, 2003, "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," working paper, Erasmus University Rotterdam.

Meese, R. A., and K. Rogoff, 1983a, "Empirical Exchange Rate Model of the Seventies," *Journal of International Economics,* 14, 3-24.

Meese, R. A., and K. Rogoff, 1983b, "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" in Frenkel JA (ed.), the *Exchange Rate and International Economics,* University of Chicago Press, Chicago.

Meheswaran, S., and C. A. Sims, 1993, "Empirical Implication of Arbitrage-Free Asset Markets," In P.C.B. Phillips (eds.), *Models, Methods and Applications of Econometrics: Essays in Honor of A.R. Bergstrom*, Blackwell Publishers, Cambridge, MA.

Ng, A., 2000, "Volatility Spillover Effects from Japan and the US to the Pacific–Basin," *Journal of International Money and Finance,* 19, 207-233.

Phillips P. C. B., 1999, "Unit Root Log Periodogram Regression," working paper, Yale University.

Protter, P., 1990, *Stochastic Integration and Differential Equations: A New Approach*, Springer-Verlag, New York.

Robinson, P. M., 1995, "Log-Periodogram Regression of Time Series with Long Range Dependence," *The Annals of Statistics,* 23, 1048-1072.

Robinson, P. M., 2003, *Time Series with Long Memory*, Oxford University Press, Oxford.

Schwert, G. W., 2002, "Anomalies and Market Efficiency," working paper, University of Rochester.

Susmel, R., and R. F. Engle, 1994, "Hourly Volatility Spillovers between International Equity Markets," *Journal of International Money and Finance,* 13, 3-25.

Yang, K., 2006, "How Much Do We Gain from High-Frequency Sampling: Information Transmission, Risk Hedging, and Portfolio Optimization," working in progress, Vanderbilt University.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2003, "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data," working Paper, National Bureau of Economic Research.

Table 1. Correlations between the Exchange-Traded Funds and their Underlying Market Indices' Returns

Country	Daily correlaton	Weekly correlation	Monthly correlation
Austria	0.57	0.83	0.90
Belgium	0.67	0.86	0.94
France	0.77	0.92	0.96
Germany	0.74	0.92	0.96
Hong Kong	0.50	0.81	0.93
Italy	0.80	0.93	0.97
Japan	0.64	0.88	0.93
Malaysia	0.52	0.75	0.86
The Neitherlands	0.75	0.90	0.96
Singapore	0.57	0.84	0.93
Spain	0.78	0.92	0.96
Sweden	0.73	0.93	0.95
Switzerland	0.65	0.89	0.95
United Kingdom	0.65	0.87	0.94
United States	0.95	0.97	0.96

Note: The exchange-traded funds (ETFs) prices are taken from TAQ database, while their underlying market indices (MSCI national market indices) are from Datastream Inc. Daily (weekly, monthly) correlations refer to the correlations of daily (weekly, monthly) returns between the ETFs and their underlying MSCI indices. The sample period is from May 1996 to December 2004.

Table 2a. Individual Exchange-Traded Funds

Country	Ticker	Trades per day	Quotes per day
Austria	EWO	13	68
Belgium	EWK	7	39
France	EWQ	10	101
Germany	EWG	26	157
Hong Kong	EWH	60	166
Italy	EWI	9	81
Japan	EWJ	248	516
Malaysia	EWM	34	128
The Neitherlands	EWN	6	46
Singapore	EWS	35	108
Spain	EWP	8	57
Sweden	EWD	9	54
Switzerland	EWL	10	58
United Kingdom	EWU	24	113
United States	SPY	3877	39605

Table 2b. Equally-Weighted Portfolios of Exchange-Traded Funds

Portfolios	Components	Trades per day	Quotes per day
United States	-	3877	39605
Japan	-	248	516
Asia ex-Japan	Hong Kong, Malaysia, Singapore	129	402
Europe	Austria, Belgium, France, Germany, Italy, The Netherlands, Spain, Sweden, Switzerland, United Kingdom	122	774

Note: The trades and quotes per day refer to the average number of trades and quotes from 9:30 EST until 16:00 EST. The calculation is based on the transaction records of the American Stock Exchange (AMEX) and the Nasdaq National Market System (NMS) over the period May 1996 to December 2004. The trades and quotes per day for the ETF portfolios are calculated as the sum of their individual components' trades and quotes per day.

Table 3. Summary Statistics of Five-Minute Trade Returns

	Mean	Std.Dev.	Skewness	Kurtosis	Obs.
United States	0.0004	0.1592	-0.2421	90.9677	169884
Japan	-0.0002	0.2999	0.3810	50.5501	169884
Asia ex-Japan	-0.0003	0.2866	0.9770	69.0208	169884
Europe	0.0003	0.0992	-0.8276	81.2505	169884

Note: The sample covers the period May 1996 to December 2004. The number of working days (when all the four ETF portfolios are traded) is 2178. With 78 five-minute intervals per day, we thus have a total of 169884 (= 2178×78) observations for each series. The portfolio returns are calculated as the arithmetic mean of the individual component returns.

Table 4. Unconditional Distribution Summary of Daily Returns, Volatilities and Correlations

	r_t				r_t/v_t			
	Mean	St.Dev.	Skew.	Kurt.	Mean	St.Dev.	Skew.	Kurt.
United States	0.028	1.234	-0.084	5.923	0.054	0.907	0.035	2.697
Japan	-0.019	1.702	0.391	6.228	0.002	0.662	0.023	2.621
Asia ex-Japan	-0.022	1.841	0.290	8.435	-0.005	0.709	-0.054	2.620
Europe	0.020	1.207	-0.179	5.690	0.065	1.279	-0.077	2.292
Mean	0.002	1.496	0.105	6.569	0.029	0.890	-0.018	2.558
St.Dev.	0.026	0.323	0.278	1.263	0.036	0.281	0.056	0.181

	v_t				lv_t			
	Mean	St.Dev.	Skew.	Kurt.	Mean	St.Dev.	Skew.	Kurt.
United States	1.258	0.630	3.034	22.539	0.135	0.424	0.301	3.814
Japan	2.391	1.146	1.566	9.720	0.762	0.479	-0.274	3.066
Asia ex-Japan	2.169	1.309	2.197	11.125	0.629	0.529	0.190	3.260
Europe	0.778	0.403	2.227	12.428	-0.363	0.476	-0.355	7.164
Mean	1.649	0.872	2.256	13.953	0.291	0.477	-0.034	4.326
St.Dev.	0.760	0.426	0.602	5.830	0.512	0.043	0.328	1.918

	cov_t				cor_t			
	Mean	St.Dev.	Skew.	Kurt.	Mean	St.Dev.	Skew.	Kurt.
US-JP	0.321	1.137	6.312	80.688	0.091	0.210	0.563	3.936
US-AS	0.291	1.214	5.772	92.008	0.071	0.199	0.500	3.765
US-EU	0.139	0.393	4.564	42.320	0.094	0.191	0.428	3.289
Mean	0.250	0.914	5.549	71.672	0.085	0.200	0.497	3.664
St.Dev.	0.098	0.453	0.895	26.042	0.013	0.009	0.067	0.335

Note: The sample covers the period May 1996 to December 2004, altogether 2178 observations per series. The daily returns (r_t) are calculated using daily open and close prices, while the daily realized standard deviations (v_t), logarithmic standard deviations (lv_t), covariances (cov_t), and correlations (cor_t) are calculated from five-minute intraday returns as described in Section 3.2.

Table 5. Dynamic Dependence of Daily Returns, Volatilities and Correlations

	r_t/v_t								
	Q_{20}	$p-value$	ADF_{20}	PP	d_{MLP}	ci_low	ci_up	d_S	$\bar{R}_s^2$
United States	24.82	0.208	-9.587	-47.90	0.019	-0.054	0.092	-0.006	0.991
Japan	22.06	0.337	-10.64	-49.79	0.078	0.005	0.151	-0.028	0.990
Asia ex-Japan	26.44	0.152	-9.199	-48.74	0.061	-0.012	0.134	0.033	0.993
Europe	31.54	0.049	-9.167	-43.60	0.052	-0.021	0.125	0.031	0.998
Mean	26.22	0.186	-9.649	-47.51	0.052	-0.021	0.125	0.007	0.993
St.Dev.	3.980	0.120	0.690	2.722	0.025	0.025	0.025	0.030	0.003

	lv_t								
	Q_{20}	$p-value$	ADF_{20}	PP	d_{MLP}	ci_low	ci_up	d_S	$\bar{R}_s^2$
United States	12046	0.000	-4.265	-24.70	0.554	0.481	0.627	0.416	1.000
Japan	15254	0.000	-4.003	-25.18	0.387	0.314	0.460	0.440	1.000
Asia ex-Japan	18400	0.000	-3.253	-20.75	0.407	0.334	0.480	0.453	1.000
Europe	6465	0.000	-3.819	-32.03	0.445	0.372	0.518	0.393	0.999
Mean	13041	0.000	-3.835	-25.66	0.448	0.375	0.522	0.425	1.000
St.Dev.	5094	0.000	0.429	4.684	0.075	0.075	0.075	0.027	0.001

	cor_t								
	Q_{20}	$p-value$	ADF_{20}	PP	d_{MLP}	ci_low	ci_up	d_S	$\bar{R}_s^2$
US-JP	688.6	0.000	-7.851	-44.73	0.144	0.071	0.217	0.254	0.994
US-AS	79.05	0.000	-9.230	-47.70	0.135	0.062	0.208	0.128	0.993
US-EU	78.89	0.000	-9.408	-45.52	0.071	-0.002	0.144	0.133	0.997
Mean	282.2	0.000	-8.830	-45.99	0.117	0.044	0.190	0.171	0.995
St.Dev.	352.0	0.000	0.852	1.540	0.040	0.040	0.040	0.071	0.002

Note: The table summarizes the time-series dependence in the daily realized standardized returns (r_t/v_t), logarithmic standard deviations (lv_t), and correlations (cor_t), over the period May 1996 to December 2004. The Ljung-Box Q-statistic, Q_{20}, is a test statistic for the null hypothesis of no autocorrelation up to order 20. The unit root hypothesis is tested by both the augmented Dickey-Fuller statistic with 20 augmentation lags, ADF_{20}, and the Phillips-Perron statistic, PP, with the 1% and 5% critical values given by –3.9676 and -3.4145. The long memory parameter is estimated by two methods, the modified log periodogram estimation (Kim and Phillips, 1999a) and the scaling law regression (Diebold and Lindner, 1996). The point estimates, and the lower and upper bounds of the 95% confidence intervals estimated from the former method, are denoted as d_{MLP}, ci_low and ci_up. The point estimates and the adjusted R-squared from the latter method are denoted as d_S and $\bar{R}_s^2$.

Table 6a. In-Sample Spillovers

	r_t/v_t							
	United States		Japan		Asia ex-Japan		Europe	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
C	0.06^a	0.01	-0.00	0.02	-0.01	0.01	0.06^b	0.05
	(3.04)	(0.44)	(-0.05)	(0.90)	(-0.61)	(0.61)	(2.05)	(1.22)
US(-1)	-0.06^b	-0.06	-0.02	-0.06	0.02	-0.00	0.21^a	0.15^b
	(-2.12)	(-1.35)	(-0.98)	(-1.74)	(0.95)	(-0.08)	(5.19)	(2.40)
JP(-1)	-0.08^b	-0.08	-0.07^a	-0.07	-0.04	-0.02	-0.16^a	-0.15^b
	(-2.35)	(-1.41)	(-2.82)	(-1.62)	(-1.25)	(-0.51)	(-3.09)	(-2.00)
AS(-1)	0.06	0.08	0.01	0.07	-0.06^b	-0.11^a	0.06	0.10
	(1.85)	(1.30)	(0.31)	(1.54)	(-2.11)	(-2.66)	(1.35)	(1.29)
EU(-1)	0.05^a	0.03	0.03	0.03	0.02	0.02	0.01	0.01
	(2.60)	(1.10)	(1.82)	(1.14)	(1.30)	(1.05)	(0.30)	(0.12)
$\bar{R}^2$	0.008	0.007	0.002	0.001	0.006	0.011	0.020	0.009

Pairwise Granger Causality

	United States		Japan		Asia ex-Japan		Europe	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
US	-	-	[0.62]	[0.24]	[0.51]	[0.51]	$[0.00]^a$	[0.10]
JP	[0.13]	[0.22]	-	-	[0.34]	[0.73]	$[0.01]^a$	[0.13]
AS	$[0.03]^b$	[0.26]	[0.82]	[0.32]	-	-	[0.34]	[0.65]
EU	[0.07]	[0.08]	[0.08]	[0.11]	[0.10]	[0.24]	-	-
All	$[0.02]^b$	[0.13]	[0.49]	[0.32]	[0.11]	[0.44]	$[0.00]^a$	[0.14]

Note: The table reports the results for the in-sample return VAR(5) $y_t = \alpha + \sum_{j=1}^{5} \beta_j y_{t-j} + \varepsilon_t$, where $y_t = \{r_{us,t}/v_{us,t}, r_{jp,t}/v_{jp,t}, r_{as,t}/v_{as,t}, r_{eu,t}/v_{eu,t}\}'$. The Akaike and Schwarz criteria choose the lag order of 5 and 1, respectively. Considering our large sample size, we employ VAR(5) to maintain conservatism. All the inverse roots of the characteristic AR polynomial lie inside the unit circle, indicating covariance stationarity. The VAR residuals reveal no significant autocorrelations based on the portmanteau tests. We report only the coefficients for the first lags, which can be seen as measuring overnight return spillovers. The full sample covers May 1996 to December 2004, while the sub-sample covers September 2000 to December 2004. The numbers in parenthesis (brackets) are t-statistics (p-values). The symbols ^{a}, ^{b} denote significance level at 1% and 5% , respectively.

Table 6b. In-Sample Spillovers (continued)

	lv_t							
	United States		Japan		Asia ex-Japan		Europe	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
C	-0.01	-0.01	0.04^a	0.00	0.03^a	0.01	-0.03^a	-0.02
	(-0.83)	(-1.22)	(5.95)	(0.35)	(4.03)	(0.67)	(-3.07)	(-1.81)
US(-1)	-0.30^a	-0.34^a	0.07^a	0.08^b	0.06^b	0.06	0.08^b	0.06
	(-12.82)	(-10.17)	(2.75)	(2.29)	(2.21)	(1.78)	(2.45)	(1.48)
JP(-1)	0.04	0.03	-0.21^a	-0.22^a	0.05^b	0.05	0.03	0.04
	(1.77)	(0.89)	(-9.06)	(-6.75)	(2.00)	(1.75)	(1.10)	(0.99)
AS(-1)	-0.01	-0.01	0.05^b	0.03	-0.19^a	-0.23^a	-0.02	-0.01
	(-0.68)	(-0.22)	(2.12)	(0.86)	(-8.36)	(-7.02)	(-0.77)	(-0.15)
EU(-1)	0.01	0.01	0.01	0.04	0.04^b	0.06^b	-0.34^a	-0.35^a
	(0.42)	(0.50)	(0.50)	(1.27)	(2.05)	(2.20)	(-15.10)	(-10.75)
$\bar{R}^2$	0.076	0.096	0.037	0.037	0.041	0.052	0.100	0.094

Pairwise Granger Causality

	United States		Japan		Asia ex-Japan		Europe	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
US	-	-	[0.06]	[0.14]	$[0.00]^a$	$[0.00]^a$	$[0.03]^b$	[0.45]
JP	[0.27]	[0.31]	-	-	$[0.00]^a$	[0.17]	[0.26]	[0.90]
AS	[0.34]	[0.90]	$[0.00]^a$	[0.11]	-	-	[0.24]	[0.59]
EU	[0.84]	[0.74]	[0.76]	[0.77]	[0.23]	[0.16]	-	-
All	[0.51]	[0.80]	$[0.00]^a$	$[0.01]^a$	$[0.00]^a$	$[0.00]^a$	$[0.02]^b$	[0.66]

Note: The table reports the results for the long-memory Gaussian VAR(5) $(1-L)^d y_t = \alpha + \sum_{j=1}^{5} \beta_j (1-L)^d y_{t-j} + \varepsilon_t$, where $y_t = \{lv_{us,t}, lv_{jp,t}, lv_{as,t}, lv_{eu,t}\}'$,and $d = \{d_{us}, d_{jp}, d_{as}, d_{eu}\}'_{MLP}$. The product of $(1-L)^d$ and y_t is calculated through element-by-element multiplication. The Akaike and Schwarz criteria choose the lag order of 4 and 1, respectively. Considering our large sample size, we employ VAR(5) to maintain conservatism. All the inverse roots of the characteristic AR polynomial lie inside the unit circle, indicating covariance stationarity. We report only the coefficients for the first lags, which can be seen as measuring overnight volatility spillovers. The full sample covers May 1996 to December 2004, while the sub-sample covers September 2000 to December 2004. The numbers in parenthesis (brackets) are t-statistics (p-values). The symbols ^{a}, ^{b} denote significance level at 1% and 5% , respectively.

Table 6c. In-Sample Spillovers (continued)

	cor_t					
	US-JP		US-AS		US-EU	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
C	0.04^a	0.06^a	0.03^a	0.05^a	0.06^a	0.07^a
	(6.56)	(6.71)	(5.71)	(5.30)	(10.82)	(8.69)
US-JP(-1)	-0.09^a	-0.09^a	0.00	-0.04	-0.01	-0.01
	(-3.97)	(-2.62)	(-0.08)	(-1.43)	(-0.49)	(-0.39)
US-AS(-1)	-0.04	-0.07^b	-0.17^a	-0.19^a	0.00	0.00
	(-1.88)	(-1.97)	(-7.35)	(-5.78)	(0.01)	(-0.09)
US-EU(-1)	0.01	0.01	0.06^a	0.09^b	-0.04^b	-0.07^b
	(0.28)	(0.17)	(2.79)	(2.51)	(-1.97)	(-2.26)
$\bar{R}^2$	0.011	0.012	0.027	0.045	0.002	0.002

Pairwise Granger Causality

	US-JP		US-AS		US-EU	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
US-JP	-	-	[0.06]	$[0.01]^a$	[0.06]	[0.25]
US-AS	[0.06]	$[0.04]^b$	-	-	[0.97]	[0.72]
US-EU	[0.84]	[0.99]	$[0.03]^b$	$[0.01]^a$	-	-
All	[0.20]	[0.30]	$[0.01]^a$	$[0.00]^a$	[0.24]	[0.34]

Note: The table reports the results for the long-memory Gaussian VAR(5) $(1-L)^d y_t = \alpha + \sum_{j=1}^{5} \beta_j (1-L)^d y_{t-j} + \varepsilon_t$, where $y_t = \{cor_{us,jp,t}, cor_{us,as,t}, cor_{us,eu,t}\}'$, and $d = \{d_{us,jp}, d_{us,as}, d_{us,eu}\}'_{MLP}$. The product of $(1-L)^d$ and y_t is calculated through element-by-element multiplication. The Akaike and Schwarz criteria choose the lag order of 4 and 1, respectively. Considering our large sample size, we employ VAR(5) to maintain conservatism. All the inverse roots of the characteristic AR polynomial lie inside the unit circle, indicating covariance stationarity. We report only the coefficients for the first lags, which can be seen as measuring overnight correlation spillovers. The full sample covers May 1996 to December 2004, while the sub-sample covers September 2000 to December 2004. The numbers in parenthesis (brackets) are t-statistics (p-values). The symbols ^{a}, ^{b} denote significance level at 1% and 5% , respectively.

Table 7. Out-of-Sample Forecastability

Model	r_t/v_t United States	Japan	Asia ex-Japan	Europe
H_a^1. AR(5)+$\sum_{p=1}^5$US($-p$)	–	-1.469	-0.981	-0.383
H_a^2. AR(5)+$\sum_{p=1}^5$JP($-p$)	-1.748	–	-0.772	-0.523
H_a^3. AR(5)+$\sum_{p=1}^5$AS($-p$)	0.054	-1.996	–	-0.392
H_a^4. AR(5)+$\sum_{p=1}^5$EU($-p$)	-0.540	0.014	-0.435	–
H_a^{all}. VAR(5)	-0.771	-1.040	-1.322	-0.384
Cri. Value for $H_a^1 - H_a^4$	0.995 (1%)	0.386 (5%)	0.062 (10%)	
Cri. Value for H_a^{all}	<0.621 (1%)	<0.043 (5%)	<-0.248 (10%)	

Model	lv_t United States	Japan	Asia ex-Japan	Europe
H_a^1. AR(5)+$\sum_{p=1}^5$US($-p$)	–	0.656^b	0.771^b	0.055
H_a^2. AR(5)+$\sum_{p=1}^5$JP($-p$)	-0.634	–	0.651^b	-0.565
H_a^3. AR(5)+$\sum_{p=1}^5$AS($-p$)	-1.206	2.062^a	–	-1.361
H_a^4. AR(5)+$\sum_{p=1}^5$EU($-p$)	-0.869	-0.311	-0.093	–
H_a^{all}. VAR(5)	-1.405	0.836^a	0.562^b	-1.441
Cri. Value for $H_a^1 - H_a^4$	0.995 (1%)	0.386 (5%)	0.062 (10%)	
Cri. Value for H_a^{all}	<0.621 (1%)	<0.043 (5%)	<-0.248 (10%)	

Model	cor_t US-JP	US-AS	US-EU
H_a^1. AR(5)+$\sum_{p=1}^5$US-JP($-p$)	–	-0.140	0.676^b
H_a^2. AR(5)+$\sum_{p=1}^5$US-AS($-p$)	1.349^a	–	-1.718
H_a^3. AR(5)+$\sum_{p=1}^5$US-EU($-p$)	-1.187	0.002	–
H_a^{all}. VAR(5)	0.299^b	0.032	-0.728
Cri. Value for $H_a^1 - H_a^3$	0.995 (1%)	0.386 (5%)	0.062 (10%)
Cri. Value for H_a^{all}	0.621 (1%)	0.043 (5%)	-0.248 (10%)

Note: The table reports the Diebold-Mariano (DM) t-statistics for the out-of-sample comparison between the benchmark $AR(5)$ model H_0: $y_{i,t} = \alpha + \sum_{j=1}^5 \beta_j y_{i,t-j} + \varepsilon_{i,t}$, where y_t refers to r_t/v_t, $(1-L)^{d_{MLP}} lv_t$, $(1-L)^{d_{MLP}} cor_t$; and alternative models H_a^k, where the AR(5) model is augmented by adding five lags of another variable; and H_a^{all}, where the AR(5) model is augmented by adding five lags of all other variables (VAR(5) model). We use the recursive scheme in constructing DM t-statistics. The sample covers May 1996 to December 2004, with the in/out sample ratio (R/P) equal to 1. Since H_a^k and H_a^{all} nest H_0, the DM t-statistics follow nonstandard distributions under the null of no superior forecastability from H_a^k (H_a^{all}) relative to H_0. Hence, we report the asymptotically valid critical values from McCracken (2004). The superscript ^{a} and ^{b} indicate significance at 1% and 5% level, respectively.

Table 8. Tests of Related Hypotheses

	r_t/v_t							
	United States		Japan		Asia ex-Japan		Europe	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
Monday	0.04	0.01	-0.05	-0.02	-0.08^b	-0.00	-0.04	0.01
	(0.77)	(0.20)	(-1.58)	(-0.38)	(-2.16)	(-0.06)	(-0.64)	(0.12)

	lv_t							
	United States		Japan		Asia ex-Japan		Europe	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
Monday	-0.08^a	-0.07^a	0.03^b	0.03	0.06^a	0.05^b	0.10^a	0.11^a
	(-5.13)	(2.46)	(2.05)	(1.40)	(3.42)	(2.10)	(4.93)	(4.02)
Leverage_US	0.14^a	0.06	0.06	0.00	0.12^a	0.12^b	0.13^b	0.12
	(3.33)	(0.96)	(1.22)	(0.05)	(2.62)	(1.96)	(2.29)	(1.49)
Leverage_JP	-0.02	-0.03	0.01	0.01	-0.01	0.05	0.09	-0.05
	(-0.62)	(-0.52)	(0.27)	(0.18)	(-0.18)	(0.92)	(-1.63)	(-0.64)
Leverage_AS	0.08^b	0.07	0.05	0.06	0.10^b	0.09	0.06	0.07
	(2.06)	(1.28)	(1.13)	(0.98)	(2.40)	(1.55)	(1.05)	(0.96)
Leverage_EU	0.06	0.11^b	0.04	0.06	0.07^b	0.07	0.00	0.03
	(1.86)	(2.46)	(1.26)	(1.16)	(1.98)	(1.44)	(0.04)	(0.57)

	cor_t					
	US-JP		US-AS		US-EU	
	full sample	sub-sample	full sample	sub-sample	full sample	sub-sample
Monday	0.00	0.02	0.02	0.04^b	0.02^b	0.03^b
	(0.39)	(0.91)	(1.65)	(2.63)	(2.32)	(2.31)
Contagion_US	-0.02	0.00	-0.01	-0.03	0.04	0.05
	(-0.44)	(-0.01)	(-0.14)	(-0.51)	(0.84)	(0.84)
Contagion_JP	0.05	0.10	0.00	0.07	0.03	0.08
	(1.22)	(1.06)	(-0.09)	(0.85)	(0.69)	(0.97)
Contagion_AS	-0.04	-0.02	0.02	0.11	0.00	-0.14
	(-0.79)	(-0.08)	(0.35)	(0.50)	(0.04)	(-0.67)
Contagion_EU	0.06	-0.02	0.02	-0.02	0.05	0.03
	(1.16)	(-0.28)	(0.40)	(-0.30)	(1.06)	(0.44)

Note: The table reports the test results of related hypotheses, as described in Section 6. Specifically, for daily standardized returns, we estimate the extended model $y_t = \alpha + \sum_{j=1}^{5} \beta_j y_{t-j} + \delta D_t + \varepsilon_t$, where $y_t = \{r_{us,t}/v_{us,t}, r_{jp,t}/v_{jp,t}, r_{as,t}/v_{as,t}, r_{eu,t}/v_{eu,t}\}'$, D_t is the Monday dummy that equals 1 on days following weekends and holidays and is 0 otherwise. Hence, the coefficient δ measures the Monday effect. For daily volatilities, we estimate the extended model $(1-L)^d y_t = \alpha + \sum_{j=1}^{5} \beta_j (1-L)^d y_{t-j} + \gamma(1-L)^d x_{t-1} + \delta D_t + \varepsilon_t$, where $y_t = \{lv_{us,t}, lv_{jp,t}, lv_{as,t}, lv_{eu,t}\}'$; $x_t = \{lv_{us,t} I(r_{us,t} < 0), lv_{jp,t} I(r_{jp,t} < 0), lv_{as,t} I(r_{as,t} < 0), lv_{eu,t} I(r_{eu,t} < 0)\}'$; $d = \{d_{us}, d_{jp}, d_{as}, d_{eu}\}'_{MLP}$; $I(\cdot)$ is the indicator function. Hence, the coefficient matrix γ measures the leverage effects. For daily correlations, we estimate the extended model $(1-L)^d y_t = \alpha + \sum_{j=1}^{5} \beta_j (1-L)^d y_{t-j} + \phi z_{t-1} + \delta D_t + \varepsilon_t$, where $y_t = \{cor_{us,jp,t}, cor_{us,as,t}, cor_{us,eu,t}\}'$; $z_t = \{I(r_{us,t} < r_{us,(q)}), I(r_{jp,t} < r_{jp,(q)}), I(r_{as,t} < r_{as,(q)}), I(r_{eu,t} < r_{eu,(q)}\}'$; $d = \{d_{us,jp}, d_{us,as}, d_{us,eu}\}'_{MLP}$, the subscript $_{(q)}$ denotes the qth quantile (we report the case $q = \frac{1}{100}$). Hence, the coefficient matrix ϕ measures the contagion effects. To save space, we report only the results for the Monday, leverage, and contagion effects.

Figure 1a. Five-Minute Trade Returns.

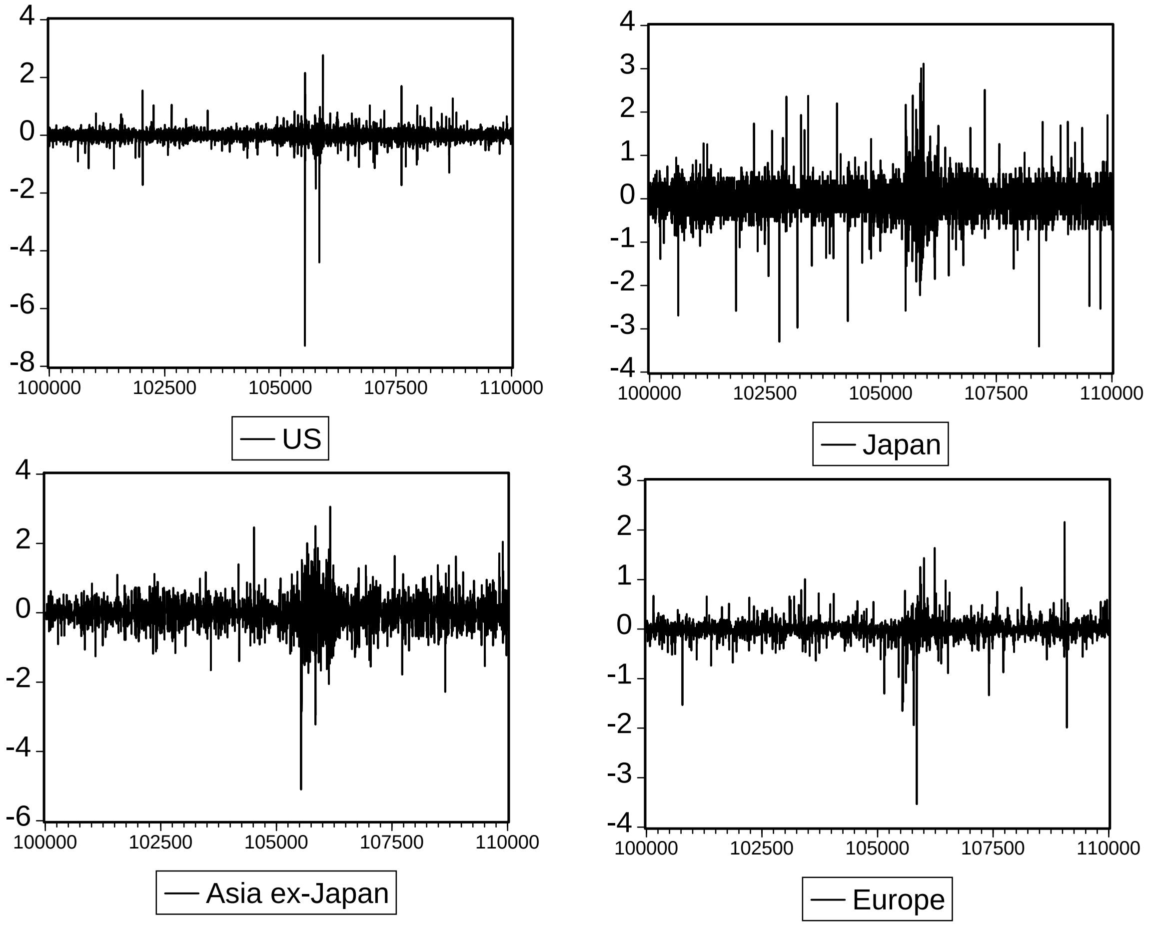


Figure 1b. Five-Minute Quote Returns.

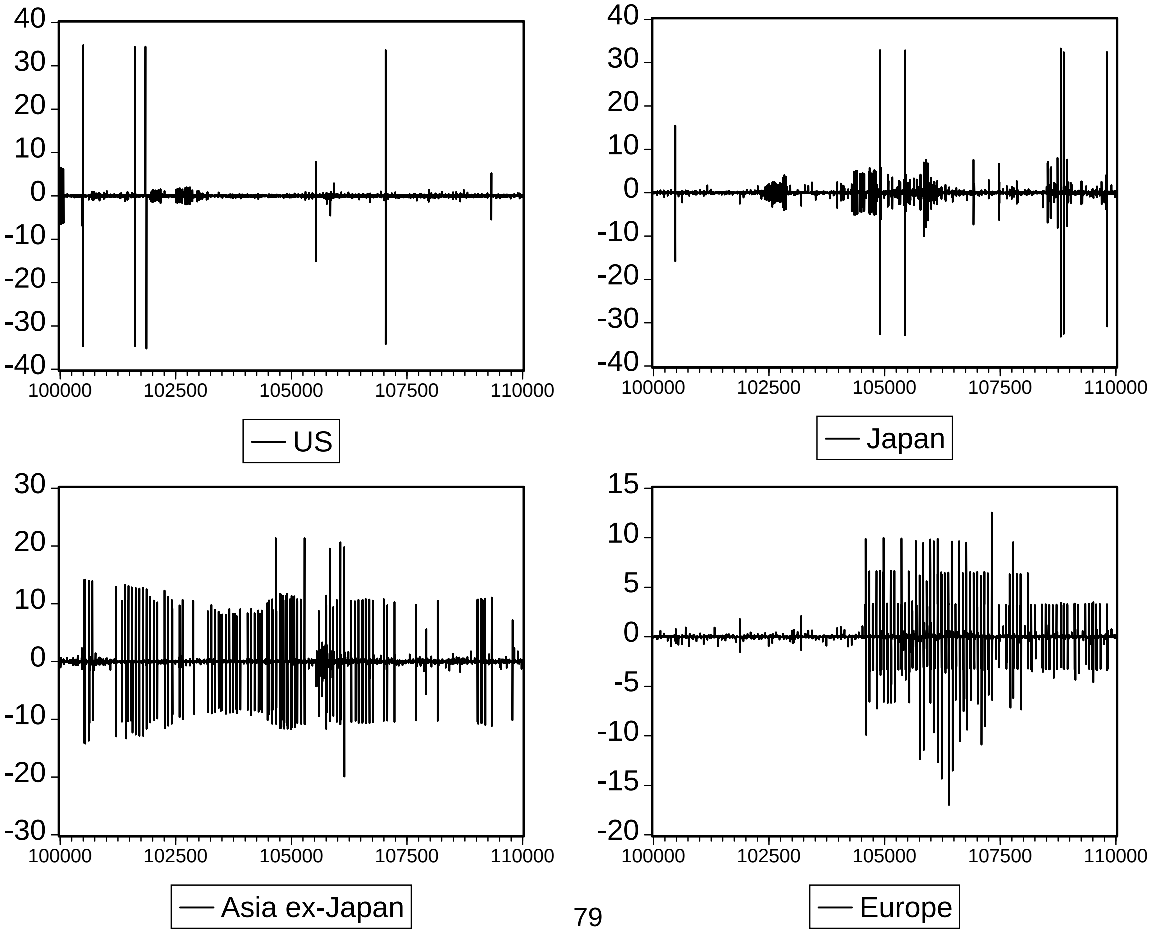
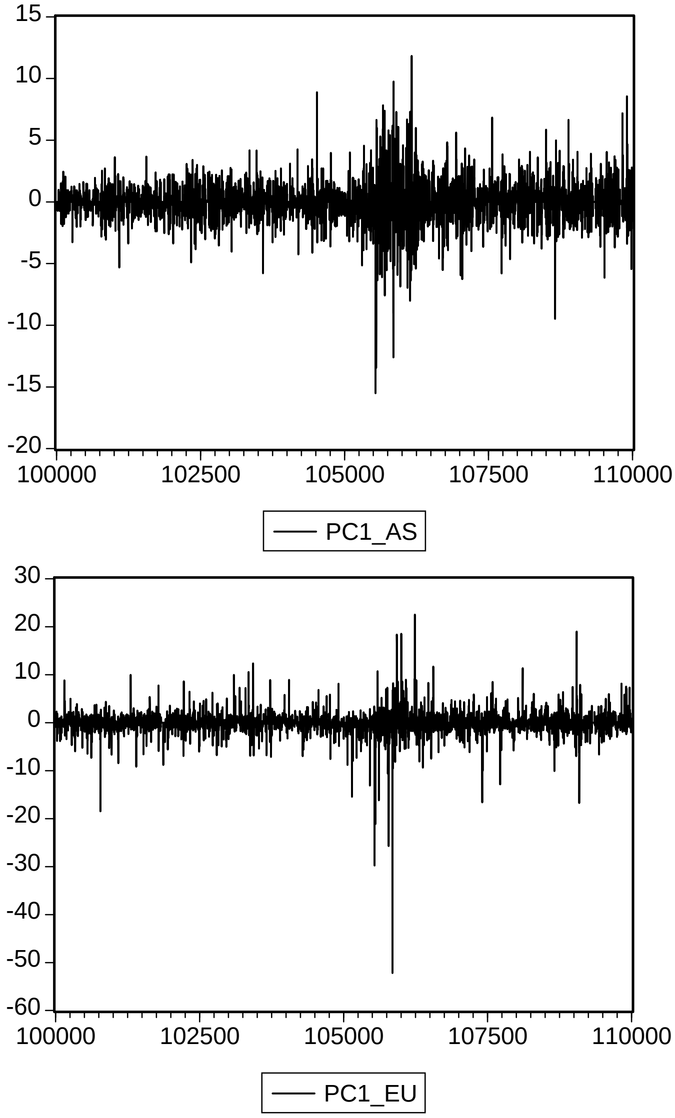
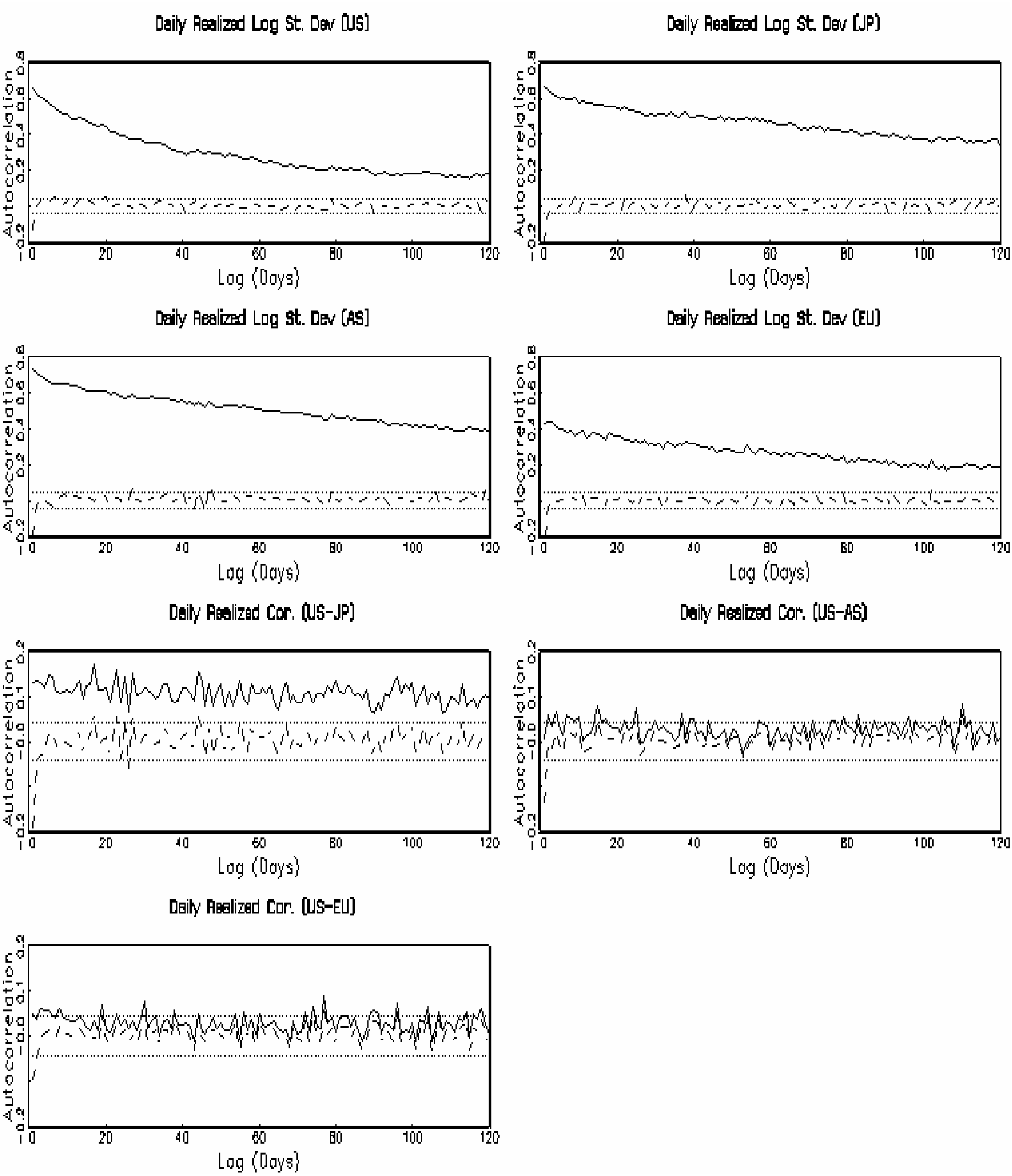


Figure 2. First Principal Components of Individual Returns.



Note: PC1_AS refers to the first principal components of individual Asian ex-Japan ETFs' five-minute trade returns. PC1_EU refers to the first principal components of individual European ETFs' five-minute trade returns. The figures are based on a randomly selected subsample (100000-110000) of the original sample (1-169884).

Figure 3. Realized Volatility and Correlation Autocorrelations.



Note: The graphs give the sample aucocorrelation functions (ACF) for daily realized volatilities and correlations. The sample covers May 1996 to December 2004. The solid lines in the figure represent the ACF of the original series, while the dashed lines are the ACF of the corresponding fractionally differenced series. The dotted lines give the 95% confidence bands of an i.i.d. Gaussian process.

CHAPTER III

HOW MUCH DO WE GAIN FROM HIGH-FREQUENCY DATA: INFORMATION TRANSMISSIONS, RISK HEDGING, AND PORTFOLIO OPTIMIZATION

Introduction

High-frequency finance (in particular, high-frequency volatility models) has become one of the two important frontiers of finance research (Engle, 2003).[54] In high-frequency finance, intra-day financial data, such as irregularly-spaced tick-by-tick data, or regularly-spaced minute-by-minute, hour-by-hour data, etc., are used for analysis. This approach was originally suggested by Merton (1980) according to which high-frequency sampling is essential for volatility estimation. Because of the difficulty of obtaining and manipulating high-frequency financial data, however, data at daily or lower frequencies are often used to estimate and forecast volatilities. And typically a parametric framework, such as a GARCH model, is assumed. With the development of computer technology and the accessibility of intra-day financial data, high-frequency finance literature has now been in a fast-developing phase.

One of the most popular high-frequency finance models is the realized volatility model. In this model, low-frequency (daily, weekly, or monthly, etc.) return volatilities are measured as the sum of high-frequency returns squared. The realized volatility model is an extension of the sample variance method, which has been used in the empirical finance literature (for example, Poterba and Summers, 1986; French, Schwert, and Stambaugh, 1987; Hsieh, 1991; Taylor and Xu, 1997). Despite its seemingly trivial extension from the sample variance method, the realized volatility method has desirable theoretical properties. Based on the theory of quadratic variation, the ex-post realized volatility is a consistent, approximately unbiased volatility estimator (Andersen, Bollerslev, and Diebold, ABD, 2005). In addition, relative to traditional models based on low-frequency data, the realized volatility is free from model misspecification, easy to implement in multi-variate systems, and provides a quick adapting estimate of current volatility. The advantages of the realized volatility method have been examined in Andersen, Bollerslev, Diebold, and Labys (ABDL, 2003), where they found that the realized volatility method produces superior out-of-sample volatility and Value-at-Risk forecasts relative to volatility models based on low-frequency data.

[54] The other frontier is high-dimension multivariate finance models.

This paper examines the gains of using high-frequency data from a more practical perspective. Volatility estimation and forecast have been applied in many academic researches and financial practices, including but not limited to detecting information (volatility) transmissions, risk hedging, and portfolio optimization. Intuitively, a superior volatility estimation method would lead to better detection of volatility transmissions, or enhanced efficiency of risk hedging and portfolio optimization. However, it remains uncertain how much the gains could be unless a rigorous statistical test is performed. Motivated by this empirical question, this paper does a comparative study where both traditional GARCH models based on low-frequency sampling and the realized volatility model based on high-frequency sampling are used to estimate (and forecast) daily volatilities. Implications in volatility transmissions, risk hedging, and portfolio optimization of these two methods are then compared.

The data used in this paper are high-frequency, intra-day transaction prices for two exchange-traded funds (ETFs): the S&P 500 index fund and the Ishares MSCI Japan index fund. These two funds are special instruments that aim to track the S&P 500 index and the MSCI Japan index. Unlike traditional index funds, these two funds are traded like common stocks in stock exchanges, such as the American Stock Exchange and the New York Stock Exchange. Therefore, their real-time transaction prices are available at intra-day frequencies. Besides the high-frequency feature of the data, these two funds are ideal candidates for the purpose of our study. First, because of their index-tracking feature, these two funds may be used to examine volatility transmissions between the US and Japan stock markets, two of the major international stock markets. This inter-market volatility transmission mechanism is important to understanding international financial integration, and has been frequently examined in the literature (see, Bauwens, Laurent, and Rombouts, 2003; Yang, 2006). Second, a global portfolio can be formed using these two index funds so as to hedge country-specific risk and achieve mean-variance efficiency. By focusing on these two major index funds as the portfolio components, we can capture a vast part of international stock markets while keeping our analysis within a manageable dimension.

The remainder of the paper is organized as follows. Section 2 describes the high-frequency financial data. Section 3 discusses the GARCH and the realized volatility models. Their implications in volatility transmissions, risk hedging and portfolio optimization are examined in Section 4 and 5, respectively. Some concluding remarks are made in Section 6.

High-frequency intra-day transaction data between 9:30 and 16:00 Eastern Standard Time (EST) are obtained from the Trade and Quote (TAQ) database for the S&P 500 index fund (US) and the Ishares MSCI Japan index fund (JP). The S&P 500 index fund was introduced on January 29, 1993, and the Japan index fund was introduced on March 12, 1996. To avoid thin trading during the introductory period of the Japan index fund, the period from May 1, 1996, to December 31, 2004, is examined.

Table 1 gives the ETF ticker symbols as well as a summary of their daily transactions. From Table 1, we can see that both ETFs are actively traded, with the minimum of 248 trades per day for JP (equivalently, one trade every 1.5 minutes). Based on the high-frequency tick-by-tick transaction data, artificial 5-minute trade and quote prices are constructed through the previous-tick interpolation method. The 5-minute frequency is chosen to achieve a reasonably large intra-day sample size while mitigating the market microstructure noise. A similar method has been used in ABDL (2001). The 5-minute trade and quote returns are then calculated as the logarithmic differences between adjacent prices.[55] As a result, there are 78 trade and quote returns per day.

Figure 1 plots a randomly selected sub-sample of the historical 5-minute trade and quote returns. Visual inspection shows that the quote data are very noisy, possibly due to discrete clustering and bid-ask bounce effects. Therefore, the analysis hereafter will be based on trade data.

Table 2 presents the summary statistics of 5-minute and daily trade returns and volatilities (measured by absolute returns). The 5-minute return series are approximately zero. In terms of the variability of 5-minute returns, JP is more widely dispersed with the standard deviation of arount 0.3. Both of the 5-minute return series are skewed and leptokurtic. The 5-minute absolute returns are skewed to the right and leptokurtic, with US having an extremely large kurtosis coefficient of 223.5. These are generally in line with the stylized facts of high-frequency returns and volatilities (e.g., Dacorogna, 2001). With regard to the daily data, the mean returns are approximately 78 times that of the 5-minute data. The variability of daily returns increases and the kurtosis decreases relative to those of the 5-minute returns. The skewness remains the same direction with small changes in magnitude. Relative to the 5-minute absolute returns,

[55] All the returns are expressed in percentage. And the first five-minute return of a day is equal to the logarithmic difference between the first five-minute price and the previous day's last five-minute price, thus is quivalently the overnight return.

the daily absolute returns have larger means and standard deviations, but smaller skewness and kurtosis.

Figure 2 plots the autocorrelation functions (ACFs) of the 5-minute and daily returns and absolute returns up to 10 days (2 weeks). From Figure 2a, the 5-minute returns are generally not serially correlated, except for the significantly negative first-order autocorrelations. Besides the first-order autocorrelations, the autocorrelations are generally not significant. The autocorrelations of the 5-minute absolute returns, however, are significantly positive at all lags. In addition, the 5-minute absolute returns possess strong daily seasonalities. From Figure 2b, the autocorrelations of daily returns are generally not significant. Exceptions are daily JP returns at the 1-day and 5-day lags, where significantly negative autocorrelations are observed. As regards daily absolute returns, significantly positive autocorrelations are observed at all lags, consistent with the volatility clustering effect documented in the literature.

Figure 3 presents the distribution of 5-minute returns and absolute returns over a day. For each 5-minute mark within the trading time of a day, the average 5-minute returns and absolute returns over the whole sample are plotted. This figure would reveal if the intra-day transactions are similar over different 5-minute intervals. For the first several 5-minute intervals, the returns are very volatile, possibly due to the reaction to overnight information. The magnitude of return variabilities is reduced after that. If the first several big jumps are removed, the returns are observed to follow a volatile-tranquile-volatile process, that is, the returns change frequently at the beginning and at the end, but are relatively quiet in the middle of a day. Correspondingly, the volatilities follow a "U" shape, which is documented in the high-frequency volatility literature.

Models

Both the bi-variate GARCH and the realized volatility models are used to estimate and forecast daily variance-covariance matrices, which are essential in various areas, such as detecting information transmissions, risk hedging, and portfolio optimization. Based on the daily data, the bi-variate GARCH method models the daily conditional variance-covariance matrix as a linear function of its own lags and the lagged cross products of return residuals. On the other hand, the realized volatility model makes use of intra-day high-frequency data and measures daily variance-covariance matrices as the sum of the cross products of intra-day returns. The conditional variance-covariance matrix is then considered as observed and can be directly modeled as an autoregressive moving average (ARMA) process for forecast purpose.

Bi-variate GARCH model

The mean return process is modeled as a VARMA(P,Q) system:

$$\mathbf{r}_t = \mathbf{a} + \sum_{p=1}^{P} \mathbf{b}_p \mathbf{r}_{t-p} + \sum_{q=1}^{Q} \mathbf{c}_q \mathbf{e}_{t-m} + \mathbf{e}_t,\ \mathbf{r}_t' = [r_{1t} \quad r_{2t}], \tag{20}$$

$$\mathbf{e}_t | \Omega_{t-1} \sim N(\mathbf{0}, H_t) \tag{21}$$

where $\mathbf{a}$ is a (2×1) vector, $\mathbf{b}$ and $\mathbf{c}$ are (2×2) matrices and $\mathbf{c}$ is restricted to be diagonal; r_1 and r_2 stand for the daily returns for US and JP, respectively.

The bi-variate GARCH method models the second moments (daily variance-covariance matrices) as follows:

$$H_t = C'C + \sum_{l=1}^{L} A_l' \mathbf{e}_{t-l} \mathbf{e}_{t-l}' A_l + \sum_{m=1}^{M} B_m' H_{t-m} B_m, \tag{22}$$

where C, A_l,and B_m are (2×2) matrices and C is restricted to be upper triangular. This model is proposed by Engle and Kroner (1995) and is generally referred to as the BEKK(L,M) model (the acronym comes from early work on multivariate GARCH models by Baba, Engle, Kraft, and Kroner). The BEKK model guarantees the positive definiteness of H_t. In addition, it incorporates cross-asset dependence while keeping the number of parameters relatively small. These properties are desirable for the purpose of our study. For propositions and proofs of the BEKK model, see Engle and Kroner (1995). For a survey on multi-variate GARCH models, see Kroner and Ng (1998), Bauwens, Laurent, and Rombouts (2006).

Realized volatility model

The realized volatility model measures daily variance-covariane matrices as:

$$H_t = \sum_{n=1}^{N} \mathbf{r}_{n,t}\mathbf{r}'_{n,t}, \ \ \mathbf{r}'_{n,t} = [r_{n,1t} \ \ r_{n,2t}], \tag{23}$$

where $r_{n,1t}$ and $r_{n,2t}$ are the nth 5-minute returns for US and Japan index funds on day t, respectively; $N = 78$ in our case. Assuming that the multivariate asset return process is a special semi-martingale, as well as some other mild conditions, the realized volatility method measures the conditional variance-covariance matrices approximately free of measurement error, without further assuming any specific return generating model. ABDL (2001, 2003) present formal derivations and proofs, while Barndorff-Nielsen and Shephard (2004) provide the asymptotic distribution theory by adding the assumption that the (logarithmic) asset price process is a continuous stochastic volatility semi-martingale. For a thorough review and comparison between the realized volatility and other volatility models, we refer to ABD (2005).

The measured daily realized variance-covariance matrices are then modeled as a long-memory VAR(K) system:

$$(1-L)^{\mathbf{d}} vech(H_t) = \boldsymbol{\alpha} + \sum_{k=1}^{K} \Psi_k (1-L)^{\mathbf{d}} vech(H_{t-k}) + \boldsymbol{\varepsilon}_t, \tag{24}$$

where $vech(H_t)$ denotes the lower triangular portion of H_t; $\mathbf{d}$ is the long memory parameter vector for $vech(H_t)$ estimated using the modified log periodogram regression (MLP) by Kim and Phillips (1999); and the product of $(1-L)^d$ and $vech(H_t)$ is calculated through element-by-element multiplication; $\boldsymbol{\alpha}$ is a (3×1) vector; Ψs are (3×3) matrices. The fractional difference filter $(1-L)^{\mathbf{d}}$ is used to capture the long memory, while the VAR captures the short-memory dynamics in H_t. The VAR model of realized volatilities (henceforth VAR-RV) has been used in ABDL 2003, and is shown to out-perform uni-variate daily GARCH and related models in terms of volatility forecast.

Information Transmissions

Information transmissions are referred to as the lead-lag relations in stock returns and volatilities across international markets. Important to understanding international financial integration, information transmissions have received much attention over the past two decades. Most of the research relies on GARCH models of different parametric specifications based on low-frequency data: Hamao, Masulis, and Ng (1990); Lin, Engle, and Ito (1994); Susmel and Engle (1994); Bae and Karolyi (1994); Karolyi (1995); Koutmos and Booth (1995); and Ng (2000), to name a few. Despite extensive studies in this literature, the issues of which GARCH model is the most appropriate, and in which direction the true causalities go remains unresolved. In a recent study, Yang (2006) utilizes high-frequency, intra-day ETFs data to examine information transmissions across the US, Japan, Asia ex-Japan, and Europe markets. In his study, the realized volatility model is used to avoid possible misspecification of volatility models. And uni-directional volatility spillovers from the US to other markets are observed. A direct comparison between the GARCH and the realized volatility models, however, has not been pursued in his study. In this section, both the bi-variate GARCH and the VAR-RV models are employed to examine volatility transmissions between the US and Japan ETFs, and their implications are compared.

With respect to the bi-variate GARCH model, some specifications need to be made. (1) Mean return equations. Following French, Schwert, and Stambaugh (1987), Hamao, Masulis, and Ng (1990), the mean return process is modeled as an MA(1) model (equivalently, VARMA(0,1) for the mean return vector) to accommodate possible first-order autocorrelations in the daily returns (as seen in the Japan ETF), which may be due to non-synchronous trading, bid-ask spreads, etc. Meanwhile, dummy variables for the day following a weekend (the Monday dummy) or a holiday (the Holiday dummy) are included in the return equations to take into account potential Monday and holiday effects, as documented by French (1980) and Gibbons and Hess (1981). (2) Variance-covariance equation. Based on the BIC criterion, the BEKK(2,1) specification is chosen. In addition, following Kroner and Ng (1998), the leverage terms are included to allow for the asymmetric effect (a negative daily return has a larger effect on the following day's return variance-covariance matrix). The Monday and Holiday dummies are also included in the BEKK specification. Therefore, the finalized BEKK model is as follows:

$$r_t = a + d_1 Hol_t + d_2 Mon_t + c_1 e_{t-1} + e_t,\ r_t' = [r_{1t} \ \ r_{2t}],\ e_t|\Omega_{t-1} \sim N(0, H_t), \tag{25}$$

$$H_t = C'C + \sum_{l=1}^{2} A_l' e_{t-l} e_{t-l}' A_l + B' H_{t-1} B + N' \eta_{t-1} \eta_{t-1}' N + D_1' D_1 Hol_t + D_2' D_2 Mon_t, \tag{26}$$

where Hol_t, Mon_t are Holiday and Monday dummies; $\mathbf{d}_1$ and $\mathbf{d}_2$ are (2×1) vectors; $\boldsymbol{\eta}_t$ $(= [\mathbf{e}_{1t} \cdot \mathbf{I}(\mathbf{e}_{1t} < 0), \mathbf{e}_{2t} \cdot \mathbf{I}(\mathbf{e}_{2t} < 0)]')$ is the leverage vector, $\mathbf{I}(\cdot)$ is the indicator function; N, D_1, D_2 are (2×2) matrices, and D_1, D_2 are restricted to be upper triangular. To test volatility transmissions from one market to the other, tests of nonlinear restrictions are involved. See Appendix A for discussion.

In the VAR-RV model, a lag order of 5 ($K = 5$) is chosen to take into account possible week seasonality. The leverage effects and Monday/holiday effects are taken into account as well, that is:

$$(1-L)^{\mathbf{d}} vech(H_t) = \boldsymbol{\alpha} + \sum_{k=1}^{5} \Psi_k (1-L)^{\mathbf{d}} vech(H_{t-k}) + \Phi(1-L)^{\mathbf{d}} x_{t-1} + \boldsymbol{\delta}_1 Hol_t + \boldsymbol{\delta}_2 Mon_t + \varepsilon_t \tag{27}$$

where x_t $(= [H_{11,t} \cdot \mathbf{I}(\mathbf{r}_{1t} < 0), H_{22,t} \cdot \mathbf{I}(\mathbf{r}_{2t} < 0)]')$ is the leverage vector; Φ is a (3×2) vectors; δ_1 and δ_2 are (2×1) vectors. All the inverse roots of the characteristic AR polynomial lie inside the unit circle, indicating covariance stationarity. To test overnight spillovers, the coefficients for the first lags and the associated t statistics are calculated. In addition, pairwise Granger causality tests are carried out to see whether there are significant return spillovers (up to 5 lags) from one endogenous variable to another.

The results are reported in Table 3. From diagnostic statistics, the daily return residuals standardized by the GARCH volatilities are fat tailed with the excess kurtosis of 1.2574 and 0.9248 for US and Japan, respectively; and they are very extremely non-normal distributed, with the Jarque-Bera statistics equal to 190.5356 and 78.7583, respectively. The daily return residuals standardized by the realized volatilities, on the contrary, are less fat tailed (the excess kurtosis equal to -0.2995 and -0.3652, respectively) and close to normally distributed (the Jarque-Bera statistic equal to 8.4412 and 12.3492, respectively). These findings are consistent with those reported in ABDL (2001). The Ljung-Box statistics indicate that the bi-variate GARCH specification is adequate in capturing the dynamics in the first and second moments of the standardized return residuals.

In both models, significant self volatility spillovers are observed, which are consistent with the volatility clustering effect. For example, $H_{11,t-1}$ in the $H_{11,t}$ equation and $H_{22,t-1}$ in the $H_{22,t}$ equation, are significant at a 1% level for both the BEKK and the VAR-RV models.[56] Results about cross-market volatility spillovers, on the other hand, are different between the two models. For the bi-variate GARCH model, generally no significant cross-market volatility spillovers are observed. An exception is the leverage effect term (η_{1t-1}^2) in the $H_{22,t}$ equation, which indicates that a negative US return is likely to increase the next day's Japan volatility. In contrast, significant volatility spillovers from US to JP are observed in the VAR-RV model ($H_{11,t-1}$ is 0.1462 at a 1% level in the $H_{22,t}$ equation). The difference in cross-market spillovers is corroborated by the Chi-squared tests of block exogeneity: no cross-market spillovers are found in the BEKK model, while significant volatility spillovers from US to JP are observed in the VAR-RV model at a 1% level.

In summary, while the bi-variate GARCH model based on daily data generally does not reveal significant volatility spillovers between the US and Japan ETFs, the VAR-RV model based on high-frequency, intra-day data detects a uni-directional volatility spillovers from US to Japan. The difference could be due to the following reasons. First, although daily data are adequate in terms of capturing daily returns, they may not incorporate intra-day return volatility information. In other words, both a trading day with large intra-day return volatility and one with small intra-day return volatility could yield the same daily return. This lack of intra-day information may lead to failure of the GARCH model in detecting cross-market volatility spillovers. Second, the GARCH measure of daily volatilities essentially relies on long and slowly decaying weighted moving averages of past daily squared returns (ABDL, 2003). This renders the GARCH volatility measure a relatively noisy and imprecise estimator, which in turn may obfuscate volatility transmissions. Obviously, these two reasons are intertwined with each other. The VAR-RV model, on the other hand, captures intra-day information and provides a quick-adapting estimate of daily volatilities. These features may provide the VAR-RV model better capability to detect volatility transmissions.

[56] Note that negative AR(1) coefficients in the VAR-RV model do not conflict with the volatility clustering effect, since they are applied to the fractional differenced instead of the original volatility series (see Andersen, Bollerslev, Diebold, and Ebens, 2001).

The variance-covariance matrix of asset returns plays an important role in portfolio construction. For example, the variance-covariance matrix can be used to minimize portfolio variance via the so-called "risk-minimizing hedge ratio" (Kroner and Ng, 1998). Also, the variance-covariance matrix is an important input in constructing the mean-variance frontier. According to ABD (2003, 2005), the VAR-RV model produces superior out-of-sample volatility forecast relative to GARCH(1,1) and other low-frequency volatility models. ABD (2003) also studied the application of VAR-RV model in return density forecast and associated Value-at-Risk estimation. To my best knowledge, however, the comparison between high-frequency and low-frequency models in a multivariate setup and its direct implication in portfolio efficiency have yet to be explored. Therefore, this section examines how the high-frequency method (VAR-RV) improves portfolio efficiency relative to the low-frequency method (GARCH).

For forecast purpose, daily returns are modeled as a VAR(1) system.[57] Daily variance-covariance matrices are modeled as a BEKK(1,1) model and a VAR(5)-RV model, respectively. Without loss of generosity, the Monday/holiday effects and the leverage effects are omitted in both return and variance-covariance equations. For evaluation of forecast performance, the whole sample (1:2178) is divided into two sub-samples with the ratio of 2:1. The first sub-sample (in-sample) is used for model estimation. Based on in-sample estimation, 1-day, 2-day, ..., up to 1-month (22-day) ahead returns and variance-covariance matrices are forecasted. Accordingly, the optimal portfolios (to be defined below) are constructed 1-day, 2-day, ..., up to 1-month ahead. One month later, the in-sample window is rolled over by including the latest one month data and excluding the earliest one month data. The in-sample estimation and out-of-sample optimization procedure are then repeated. This process continues until the end of our sample is reached. In total, the portfolio optimization procedure is repeated 33 times ($=(2178-1452)/22$).

Depending on the objective function, there are two kinds of optimal portfolios. (1) Minimum-variance portfolio. If our objective is to minimize portfolio risk (without considering portfolio returns), we can calculate the "risk-minimizing hedge ratio" h between the US and Japan ETFs, that is, how many dollars

[57] Both AIC and BIC choose the lag order of 1. Considering our large sample size, it might be desirable to choose a higher lag order. However, a high order VAR system for the daily returns will cause difficulty in estimating multi-variate GARCH models. Since our main goal is to compare different models of the second moments, I therefore set the lag order of 1 for returns.

to short in the US ETF in order to hedge the risk of longing one dollar in the Japan ETF. Specifically, at date t, $h_{t+k|t}$ is calculated so as to minimize portfolio variance $\sigma^2_{p,t+k|t}$:

$$Min\ \sigma^2_{p,t+k|t} \equiv Var_{t+k|t}(hr_1 + r_2) = h^2_{t+k|t}\sigma^2_{1,t+k|t} + \sigma^2_{2,t+k|t} + 2h_{t+k|t}\sigma_{12,t+k|t}.\ k = 1, 2, ..., 22, \qquad (28)$$

where ${}_{t+k|t}$ denotes the expected (forecast) value at $t + k$ based on information set at t. By FOC, we can easily get

$$h^*_{t+k|t} = -\frac{\sigma_{12,t+k|t}}{\sigma^2_{1,t+k|t}}. \qquad (29)$$

(2) Mean-variance efficient portfolio. Although the minimum variance portfolio shields the investors from price volatility risks, it does not take portfolio returns into account. Therefore, we may also choose appropriate portfolio weights $[w_{1,t+k|t}, w_{2,t+k|t}]$ to maximize the expected return-to-variance ratio $S_{t+k|t}$:

$$Max\ S_{t+k|t} \equiv \frac{E_{t+k|t}(w_1r_1 + w_2r_2)}{Var_{t+k|t}(w_1r_1 + w_2r_2)} = \frac{w_{1,t+k|t}r_{1,t+k|t} + w_{2,t+k|t}r_{2,t+k|t}}{w^2_{1,t+k|t}\sigma^2_{1,t+k|t} + w^2_{2,t+k|t}\sigma^2_{2,t+k|t} + 2w_{1,t+k|t}w_{2,t+k|t}\sigma_{12,t+k|t}}, \qquad (30)$$

subject to $w_{1,t+k|t} + w_{2,t+k|t} = 1$. It is easy to show that the optimal weights are (see Appendix B; for simplicity of notation, ${}_{t+k|t}$ is omitted)

$$w^*_1 = \frac{1}{r_1 - r_2}\left(\sqrt{\frac{r^2_2\sigma^2_1 - 2r_1r_2\sigma_{12} + r^2_1\sigma^2_2}{\sigma^2_1 + \sigma^2_2 - 2\sigma_{12}}} - r_2\right); \qquad (31)$$

$$w^*_2 = 1 - w^*_1. \qquad (32)$$

Based on Equation (29), (31) – (32) and out-of-sample forecasts of returns and variance-covariance matrices, the optimal portfolios are constructed. To evaluate portfolio performance, we need to use 'true' daily variance-covariance matrices (as well as daily returns), which are not directly observable. Following ABDL (2003), Ledoit, Santa-Clara, and Wolf (2003), I use the realized variance-covariance matrices estimated from intra-day 5-minute return data as a proxy.

To compare the two models in terms of minimizing portfolio risks, the realized portfolio variances at date $t+k$ are calculated based on the constructed hedge ratios for each model, that is,

$$\sigma^2_{p,t+k} \equiv var_{t+k}(h^* r_1 + r_2) = h^{*2}_{t+k|t}\sigma^2_{1,t+k} + \sigma^2_{2,t+k} + 2h^*_{t+k|t}\sigma_{12,t+h}, \tag{33}$$

where $h^*_{t+k|t}$ is calculated by Equation (29), and $\sigma^2_{1,t+k}, \sigma^2_{2,t+k}, \sigma_{12,t+k}$ are the realized variances and covariance at date $t+k$. For ease of comparison, the realized average daily portfolio variances for 1-day, 1-week, 2-week, and 1-month horizons are calculated as $\sigma^2_{p,t\to t+K} = \frac{1}{K}\sum_{k=1}^{K}\sigma^2_{p,t+k}$, where $K = 1, 5, 10, 22$. The average $\sigma^2_{p,t\to t+K}$ over the 33 periods ($\bar{\sigma}^2_{p,t\to t+K}$) are then calculated, and converted to annualized standard deviations ($\sqrt{\bar{\sigma}^2_{p,t\to t+K} \times 250}$).

Similarly, the realized portfolio return-to-variance ratio at date $t+k$ are calculated based on the constructed weighting vector for each model, that is,

$$S_{t+k} = \frac{w^*_{1,t+k|t} r_{1,t+k} + w^*_{2,t+k|t} r_{2,t+k}}{w^{*2}_{1,t+k}\sigma^2_{1,t+k} + w^{*2}_{2,t+k}\sigma^2_{2,t+k} + 2w^*_{1,t+k}w^*_{2,t+k}\sigma_{12,t+k}}, \tag{34}$$

where $w^*_{1,t+k|t}$ and $w^*_{2,t+k|t}$ are calculated by Equation (31) and (32), and $r_{1,t+k}, r_{2,t+k}, \sigma^2_{1,t+k}, \sigma^2_{2,t+k}, \sigma_{12,t+k}$ are the realized returns, variances and covariance at date $t+k$. I then calculate the realized daily portfolio return-to-variance ratio for 1-day, 1-week, 2-week, and 1-month horizons as $S_{p,t\to t+K} = \frac{1}{K}\sum_{k=1}^{K} S_{p,t+k}$, where $K = 1, 5, 10, 22$. The average $S_{p,t\to t+K}$ over the 33 periods ($\bar{S}_{p,t\to t+K}$) are then calculated, and converted to annualized return-to-variance ratio (actually it remains the same, since $\bar{S}_{p,t\to t+K} \times \frac{250}{250} = \bar{S}_{p,t\to t+K}$).

The results are report in Table 4. In the upper panel, the average annualized portfolio standard deviations for each model, as well as the associated matched pairs t statistic, are reported.[58] Over all the five horizons, the VAR-RV model consistently beats the GARCH model at a 1% significance level. Generally, the annualized portfolio standard deviations increase monotonically with the horizon, indicating that the forecast performance of both models worsens with the forecast horizon. There is no obvious evidence, however, showing that the difference of the forecast performance between the two models changes with the horizon.

[58] Based on descriptive statistics (not reported here but available upon request), the differences of average annualized portfolio standard deviations and return-variance ratios between GARCH and VAR-RV are approximately normal.

In the lower panel, the average annualized return-variance ratios for each model, as well as the associated matched pairs t statistics are reported. The VAR-RV return-variance ratios outperforms the GARCH return-variance ratios at a 1% significance level at the 1-day horizon, and a 5% significance level for all the other horizons. Both models have superior performance at the 1-day horizon relative to the other horizons. This short-run superiority could results from the short-term forecastability in both returns and variances. Again, it is not obvious that the difference between the VAR-RV return-variance ratio and the GARCH return-variance ratio changes with the horizon.

Conclusion

This paper examines the gains of using high-frequency data relative to low-frequency data in a multi-variate framework. Specifically, it compares a vector autoregressive model of the realized volatilities (VAR-RV) with a bi-variate GARCH model from the following aspects: detection of cross-market volatility transmissions, risk hedging, and portfolio optimization. In the analysis, the intra-day high-frequency returns for the US and Japan exchange-traded funds (ETFs) are used. Both ETFs are highly liquid assets and represent two of the largest international stock markets. The results show that while the bi-variate GARCH model generally does not detect any significant volatility transmissions between the two ETFs, the VAR-RV model reveals significant uni-directional volatility spillovers from US to Japan. In addition, the optimized portfolios based on the VAR-RV model outperform those based on the GARCH model in terms of minimizing portfolio risk (standard deviations) or maximizing portfolio return-to-variance ratios over various horizons.

There are at least two directions for future research. First, the comparison between high-frequency and low-frequency data analysis is based on two most popular multi-variate volatility models, the bi-variate GARCH and the VAR-RV model. Obviously, we need to check the robustness of the results to other model specifications. For example, among high-frequency volatility models, the range-based volatility method (which estimates the volatility over an interval as the difference between the highest and lowest intra-interval log prices) can also capture intra-day information. Moreover, it is less subject to market microstructure noise than the realized volatility model is. With its development in a multi-variate framework (e.g., Brandt and Diebold, 2006), the range-based method would be another ideal candidate for our analysis. Similarly, the bi-variate GARCH model specification is far from being ideal. One extension of the bi-variate GARCH model would be to incorporate the long memory characteristic.

Second, the gains from high-frequency data are examined based on their application in the estimation of low-frequency (daily) variance-covariance matrix. It will also be (or even more) interesting to see if exploring high-frequency information would enable us to gain from high-frequency trading. Intuitively, high-frequency data provide us the opportunity to examine the predictability of asset return, volatility, and trading time over a very short horizon. The predictability, if any, could be used to generate high-frequency trading opportunities. Indeed, high-frequency trading strategies have been pursued by financial practitioners. For example, Renaissance Technologies, a hedge fund company founded by Dr. Jim Simons in 1982, has been developing sophisticated high-frequency trading models and generating a compounded annual return in excess of 30 percent for the past 20 years. Undoubtedly, it remains a challenging topic to combine short-term predictability and high transaction costs to examine the gains from high-frequency data. I leave these for future research.

Appendix A

Tests of Information Spillovers

This appendix documents the specifics of testing information (volatility) spillovers using the bi-variate GARCH model. The model is:

$$r_t = a + d_1 Hol_t + d_2 Mon_t + c_1 e_{t-1} + e_t,\ r_t' = [r_{1t} \quad r_{2t}],\ e_t|\Omega_{t-1} \sim N(0, H_t), \tag{35}$$

$$H_t = C'C + \sum_{l=1}^{2} A_l' e_{t-l} e_{t-l}' A_l + B' H_{t-1} B + N' \eta_{t-1} \eta_{t-1}' N + D_1' D_1 Hol_t + D_2' D_2 Mon_t, \tag{36}$$

The variables of interest are the diagonal elements of H_t ($H_{11,t}$ and $H_{22,t}$), that is, the daily variance of US and Japan ETFs. Based on (36), we can write

$$\begin{aligned} H_{11,t} =& C_{11}^2 + \sum_{l=1}^{2} [(A_{11}^{(l)} e_{1t-l})^2 + 2A_{11}^{(l)} A_{21}^{(l)} e_{1t-l} e_{2t-l} + (A_{21}^{(l)} e_{2t-l})^2] + (B_{11}^2 H_{11,t-1} + 2B_{11} B_{21} H_{12,t-1} + B_{21}^2 H_{22,t-1}) \\ &+ [(N_{11} \eta_{1t-1})^2 + 2N_{11} N_{21} \eta_{1t-1} \eta_{2t-1} + (N_{21} \eta_{2t-1})^2] + D_{1,11}^2 Hol_t + D_{2,11}^2 Mon_t, \\ H_{22,t} =& (C_{12}^2 + C_{22}^2) + \sum_{l=1}^{2} [(A_{12}^{(l)} e_{1t-l})^2 + 2A_{12}^{(l)} A_{22}^{(l)} e_{1t-l} e_{2t-l} + (A_{22}^{(l)} e_{2t-l})^2] + (B_{12}^2 H_{11,t-1} + 2B_{12} B_{22} H_{21,t-1} + B_{22}^2 H_{22,t-1}) \\ &+ [(N_{12} \eta_{1t-1})^2 + 2N_{12} N_{22} \eta_{1t-1} \eta_{2t-1} + (N_{22} \eta_{2t-1})^2] + (D_{1,12}^2 + D_{1,22}^2) Hol_t + (D_{2,12}^2 + D_{2,22}^2) Mon_t. \end{aligned}$$

And the hypotheses tested can be summarized in the following table:

From	Spillovers to $H_{11,t}$ Null Hypothesis	Spillovers to $H_{22,t}$ Null Hypothesis
Self overnight spillovers		
$H_{11,t-1}$	$B_{11}^2 = 0$	$B_{12}^2 = 0$
e_{1t-1}^2	$(A_{11}^{(1)})^2 = 0$	$(A_{12}^{(1)})^2 = 0$
η_{1t-1}^2	$N_{11}^2 = 0$	$N_{12}^2 = 0$
Cross-market overnight spillovers		
$H_{22,t-1}$	$B_{21}^2 = 0$	$B_{22}^2 = 0$
e_{2t-1}^2	$(A_{21}^{(1)})^2 = 0$	$(A_{22}^{(1)})^2 = 0$
η_{2t-1}^2	$N_{21}^2 = 0$	$N_{22}^2 = 0$
Cross-market spillovers up to one week		
	$B_{21}^2 = (A_{21}^{(1)})^2 = (A_{21}^{(2)})^2 = N_{21}^2 = 0$	$B_{12}^2 = (A_{12}^{(1)})^2 = (A_{12}^{(2)})^2 = N_{12}^2 = 0$

Therefore, nonlinear restrictions are involved. For overnight spillovers, the values of these nonlinear functions are calculated based on estimated coefficients (for example, $\hat{B}_{11}^2$), and the associated t statistics are calculated based on the first order Taylor expansion method (Greene, 2003, (6-24)-(6-26)). Similarly, to test the cross-market spillovers up to one week, the chi-squared statistics are calculated (Greene, 2003, (6-27)-(6-29)).

Appendix B

Deriving the Optimal Portfolio Weights

Suppose we know $E_{t+k|t}[r_{us}] = r_1, Var_{t+k|t}[r_{us}] = \sigma_1^2, E_{t+k|t}[r_{jp}] = r_2, Var_{t+k|t}[r_{jp}] = \sigma_2^2, Cov_{t+k|t}(r_{us}, r_{jp}) = \sigma_{12}$; We want to choose $[w_1, w_2]$ to:

$$Max\ S_{t+k|t} \equiv \frac{w_1 r_1 + w_2 r_2}{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}},\ \text{subject to } w_1 + w_2 = 1. \tag{37}$$

Substituting $1 - w_1$ for w_2, and by FOC, we have

$$(r_1-r_2)[w_1^2\sigma_1^2+(1-w_1)^2\sigma_2^2+2w_1(1-w_1)\sigma_{12}] = [(r_1-r_2)w_1+r_2][2w_1\sigma_1^2-2(1-w_1)\sigma_2^2+2\sigma_{12}-4w_1\sigma_{12}] \tag{38}$$

$$(r_1-r_2)[(\sigma_1^2+\sigma_2^2-2\sigma_{12})w_1^2+2(\sigma_{12}-\sigma_2^2)w_1+\sigma_2^2] = 2[(r_1-r_2)w_1+r_2][(\sigma_1^2+\sigma_2^2-2\sigma_{12})w_1+\sigma_{12}-\sigma_2^2] \tag{39}$$

$$(r_1-r_2)(\sigma_1^2+\sigma_2^2-2\sigma_{12})w_1^2+2r_2(\sigma_1^2+\sigma_2^2-2\sigma_{12})w_1+2r_2\sigma_{12}-(r_1+r_2)\sigma_2^2 = 0 \tag{40}$$

$$w_1^* = -\frac{r_2}{r_1-r_2} \pm \frac{1}{r_1-r_2}\sqrt{\frac{r_2^2\sigma_1^2-2r_1r_2\sigma_{12}+r_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2-2\sigma_{12}}} \tag{41}$$

Without loss of generality, I will use $w_1^* = -\frac{r_2}{r_1-r_2}+\frac{1}{r_1-r_2}\sqrt{\frac{r_2^2\sigma_1^2-2r_1r_2\sigma_{12}+r_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2-2\sigma_{12}}}$,and $w_2^* = 1-w_1^*$.

REFERENCES

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association,* 96, 42-55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics,* 61, 43-76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica,* 71, 579-625.

Andersen, T. G., T. Bollerslev, and F. X. Diebold, 2005, "Parametric and Nonparametric Measurement of Volatility," in Yacine Aït-Sahalia and Lars Peter Hansen (eds.), *Handbook of Financial Econometrics*, North Holland, Amsterdam.

Bae, K., and G. A. Karolyi, 1994, "Good News, Bad News and International Spillovers of Stock Return Volatility between Japan and the U.S.," *Pacific-Basin Finance Journal*, 2, 405-438.

Barndorff-Nielsen, O. E., and N. Shephard, 2004, "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica,* 72*,* 885-925.

Bauwens, L., S. Laurent, and J.V.K. Rombouts, 2006, "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79-109.

Brandt, M.W., and F.X. Diebold, 2006, "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations," Journal of Business, 79, 61-74.

Dacorogna, M. M., R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet, 2001, *An Introduction to High-Frequency Finance*, Academic Press, San Diego.

Engle, R., and F. Kroner, 1995, "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

Engle, R. F., 2003, "Risk and Volatility: Econometric Models and Financial Practice," Nobel Lecture, December 8, 2003.

French, K., 1980, "Stock Returns and the Weekend Effect," *Journal of Financial Economics,* 8, 55-69.

French, K.R., G.W. Schwert, and R.F. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.

Gibbons, M. R., and P. Hess, 1981, "Day of the Week Effects and Asset Returns," *Journal of Business,* 54, 579-596.

Greene, W. H., 2003, *Econometric Analysis*, Prentice-Hall, New Jersey.

Hamao, Y., R. W. Masulis, and V. Ng, 1990, "Correlations in Price Changes and Volatilities across International Stock Markets," *Review of Financial Studies,* 3, 281-307.

Hsieh, D.A., 1991, "Chaos and Nonlinear Dynamics: Application to Financial Markets," *Journal of Finance*, 46, 1839-1877.

Karolyi, G. A., 1995, "A Multivariate GARCH Model of International Transmission of Stock Returns and Volatility: The Case of the United States and Canada," *Journal of Business & Economic Statistics,* 13, 11-25.

Kim, C. S., and P. C. B. Phillips, 1999, "Modified Log Periodogram Regression," mimeo, Yale University.

Koutmos, G., and G. G. Booth, 1995, "Asymmetric Volatility Transmission in International Stock Markets," *Journal of International Money and Finance,* 14, 747-762.

Kroner, F., and V. Ng, 1998, "Modelling Asymmetric Comovements of Asset Returns," *Review of Financial Studies*, 11, 817–844.

Ledoit, O., P. Santa-Clara, and M. Wolf, 2003, "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets," *Review of Economics and Statistics*, 85, 735-747.

Lin, W., R. F. Engle, and T. Ito, 1994, "Do Bulls and Bears Move across Borders? International Transmission of Stock Returns and Volatilities," *Review of Financial Studies,* 7, 507-538.

Merton R. C., 1980, "On Estimating the Expected Return on the Market: An Exploraty Investigation," *Journal of Financial Economics*, 8, 323-361.

Ng, A., 2000, "Volatility Spillover Effects from Japan and the US to the Pacific–Basin," *Journal of International Money and Finance,* 19, 207-233.

Poterba, J., and L. Summers, 1986, "The Persistence of Volatility and Stock Market Fluctuations," American Economic Review, 76, 1124-1141.

Susmel, R., and R. F. Engle, 1994, "Hourly Volatility Spillovers between International Equity Markets," *Journal of International Money and Finance,* 13, 3-25.

Taylor, S.J., and X. Xu, 1997, "The Incremental Volatility Information in One Million Foreign Exchange Quotations," *Journal of Empirical Finance*, 4, 317-340.

Yang, K., 2006, "Inter-Market Information Transmissions: Evidence from High-Frequency Index Funds Data," working in progress, Vanderbilt University.

.Table 1. Daily Transaction Summary (1996-2004)

ETFs	Ticker	Trades per day	Quotes per day
US	SPY	3877	39605
		(6 sec/trade)	(0.5 sec/quote)
Japan	EWJ	248	516
		(1.5 min/trade)	(0.7 min/trade)

Note: The trades and quotes per day refer to the average number of trades and quotes from 9:30 EST until 16:00 EST. The calculation is based on the transaction records of the American Stock Exchange (AMEX) and the Nasdaq National Market System (NMS) over the period May 1996 to December 2004.

Table 2. Summary Statistics of Returns and Volatilities

ETFs	Mean	Std.Dev.	Skewness	Kurtosis	Obs.
Five-Minute Returns					
US	0.0004	0.1592	-0.2421	90.9676	170274
Japan	-0.0002	0.2999	0.3810	50.5501	170274
Five-Minute Absolute Returns					
US	0.1010	0.1230	8.2028	223.5364	170274
Japan	0.1320	0.2692	5.0051	66.4719	170274
Daily Returns					
US	0.0281	1.2329	-0.0845	5.9290	2183
Japan	-0.0189	1.7005	0.3910	6.2427	2183
Daily Absolute Returns					
US	0.9079	0.8345	2.0806	10.6712	2183
Japan	1.2645	1.1368	2.1711	12.5501	2183

Note: The sample covers the period May 1996 to December 2004. The number of working days is 2183. With 78 five-minute intervals per day, we thus have a total of 170274 (= 2183×78) observations for each series. The returns are expressed in percentage.

Table 3. Volatility Transmissions Based on a Bi-variate GARCH Model and a VAR-RV Model

	GARCH model				VAR_RV model			
	Equation 1		Equation 2		Equation 1		Equation 2	
	$H_{11,t}$		$H_{22,t}$		$H_{11,t}$		$H_{22,t}$	
	Coef.	t value	Coef.	t value	Coef.	t value	Coef.	t value
$H_{11,t-1}$	0.896	$[39.63]^a$	0.0000	[0.24]	-0.122	$[-7.63]^a$	0.146	$[3.31]^a$
$\mathbf{e}^2_{1t-1}$	0.004	[1.21]	0.014	[1.64]	–	–	–	–
$\boldsymbol{\eta}^2_{1t-1}$ (x_{1t-1})	0.153	$[4.18]^a$	0.056	$[2.68]^a$	0.975	$[46.69]^a$	0.084	[1.46]
$H_{22,t-1}$	0.000	[0.13]	0.903	$[42.09]^a$	0.012	[1.82]	-0.105	$[-5.57]^a$
$\mathbf{e}^2_{2t-1}$	0.000	[0.48]	0.055	$[3.85]^a$	–	–	–	–
$\boldsymbol{\eta}^2_{2t-1}$ (x_{2t-1})	0.002	[0.61]	0.006	[0.78]	0.017	[1.61]	0.965	$[32.94]^a$
Chi-squared tests of block exogeneity								
from US	–	–	8.9193	(0.11)	–	–	3.672^a	$(0.00)^a$
from Japan	8.505	(0.13)	–	–	1.808	(0.11)	–	–
Standardized return residuals diagonostic statistics								
Mean	0.003		0.005		0.027		0.011	
Std. dev.	1.001		0.997		0.907		0.663	
Skew.	-0.360^a		0.056		0.029		0.026	
Excess Kurt.	1.257^a		0.925^a		-0.300^a		-0.365^a	
JB	190.54^a		78.758^a		8.441^b		12.349^a	
$LB(12)$	13.079		9.261		22.254^b		19.108	
$LB^2(12)$	13.143		4.819		49.664^a		358.30^a	

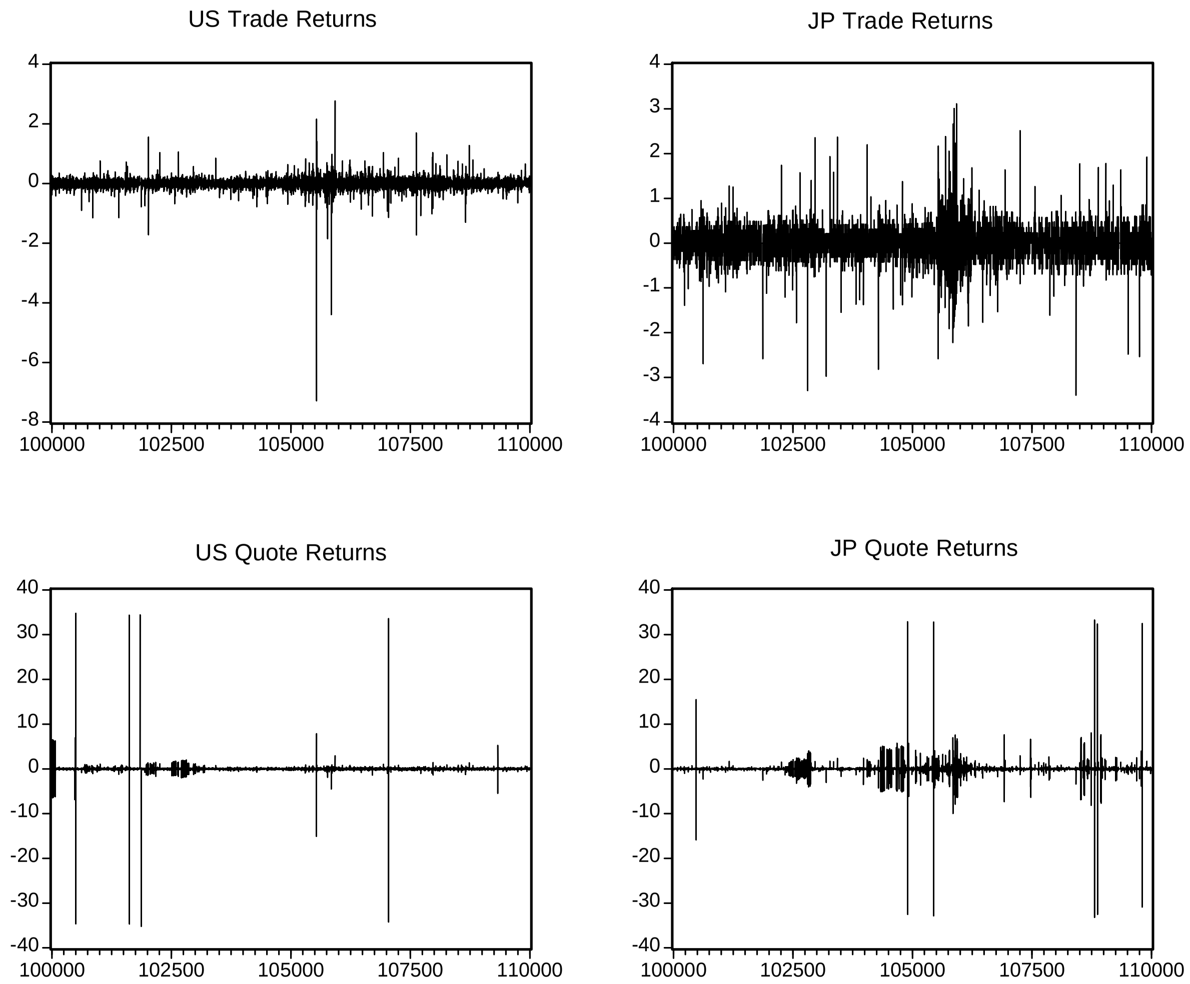
Note: The coefficients and associated t values in the bi-variate GARCH model are calculated based on the procedure described in Appendix A. $LB(12)$ and $LB^2(12)$ are the Ljung-Box statistics for serial correlation in the standardized return residuals and standardized return residuals squared at lag 12. The symbol ^{a} and ^{b} indicate significance at the 1% and 5% level, respectively.

Table 4. Performance of Optimized Portfolio based on a bi-variate GARCH and a VAR-RV model

Annualized Portfolio Std. Dev.				
Horizon	GARCH	VAR-RV	Diff.	t value
1 day	25.627	24.186	1.441	3.872^a
1 week	27.565	26.139	1.426	5.340^a
2 week	28.391	27.337	1.054	5.037^a
3 week	28.676	27.493	1.183	4.340^a
1 month	28.601	27.614	0.987	4.265^a
Annualized Return-Variance Ratio				
Horizon	GARCH	VAR-RV	Diff.	t value
1 day	0.168	0.388	-0.220	-2.816^a
1 week	0.036	0.099	-0.063	-2.188^b
2 week	0.015	0.051	-0.036	-1.970^b
3 week	0.012	0.052	-0.040	-2.002^b
1 month	0.022	0.058	-0.036	-2.293^b

Note: t value is one-sided, matched pairs t statistic, with VAR-RV as the benchmark. A positive t value in annualized portfolio std. dev. indicates a better performance of VAR-RV than that of GARCH; while a negative t value in annualized return-variance (and sharpe ratio) indicates a better performance of VAR-RV than that of GARCH.

Figure 1. Historical Five-Minute Trade and Quote Returns.



Note: The figures are based on a subsample (100000-110000) of the original sample (1-170274).

Figure 2a. Autocorrelations of Five-Minute Returns and Volatilities

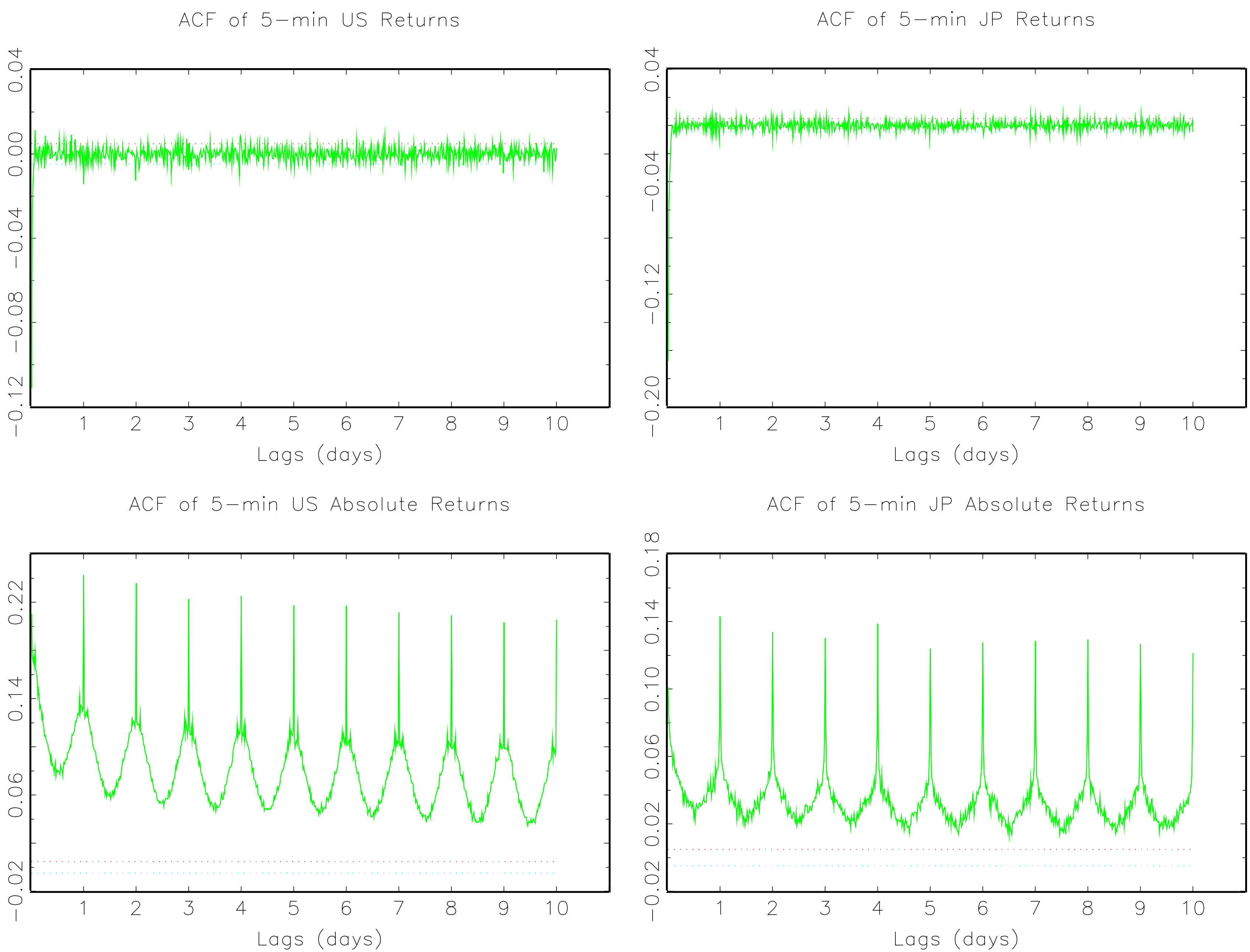


Figure 2b. Autocorrelations of Daily Returns and Volatilities

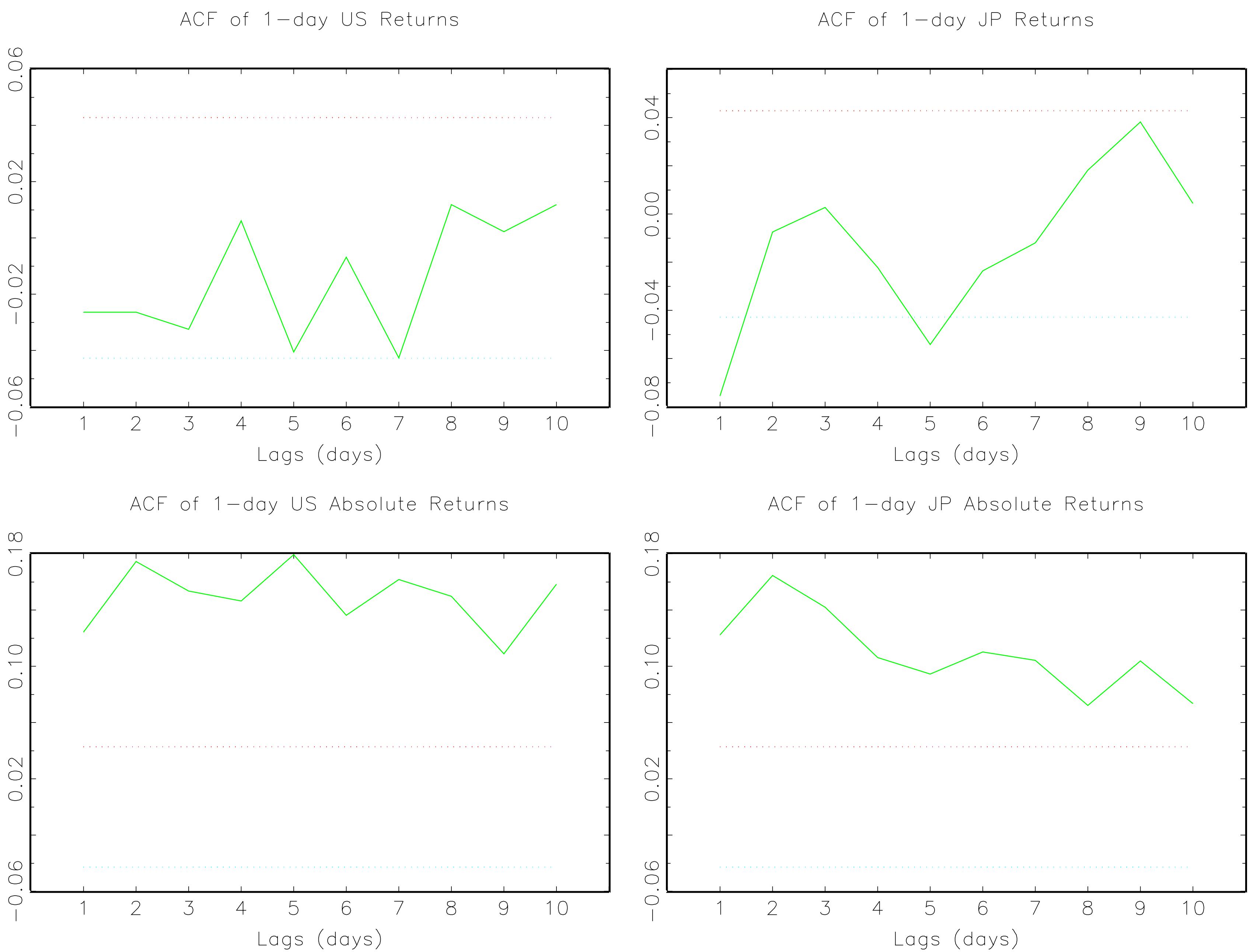


Figure 3. Intra-Day Distribution of Five-Minute Returns and Volatilities