Journal of Financial Forecasting (1–22) Volume 1/Number 2, Fall 2007

Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence

Peter F. Christoffersen
Desautels Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, Quebec H3A1G5, Canada; email: peter.christoffersen@mcgill.ca
and
CREATES, School of Economics and Management, University of Aarhus, Building 1322, DK-8000 AarhusC, Denmark

Francis X. Diebold
Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA1914-6297, USA; email: fdiebold@wharton.upenn.edu
and
NBER, 1050 Massachusetts Avenue, Cambridge, MA 02138-5398, USA

Roberto S. Mariano
School of Economics and Social Sciences, Singapore Management University, 90 Stanford Road, Singapore, 178903; email: rsmariano@smu.edu.sg

Anthony S. Tay
School of Economics and Social Sciences, Singapore Management University, 90 Stanford Road, Singapore, 178903; email: anthonytay@smu.edu.sg

Yiu Kuen Tse
School of Economics and Social Sciences, Singapore Management University, 90 Stanford Road, Singapore, 178903; email: yktse@smu.edu.sg

Recent theoretical work has revealed a direct connection between asset return volatility *forecastability and asset return* sign *forecastability. This suggests that the pervasive volatility forecastability in equity returns could, through induced sign forecastability, be used to produce direction-of-change forecasts useful for market timing. We attempt to do so in an*

This work was supported by FQRSC, IFM2, SSHRC (Canada), the National Science Foundation, the Guggenheim Foundation, the Wharton Financial Institutions Center (US) and Wharton – Singapore Management University Research Centre. We thank participants of the 9th World Congress of the Econometric Society and numerous seminars for comments, but we emphasize that any errors remain ours alone.

international sample of developed equity markets, with some success, as assessed by formal probability forecast scoring rules such as the Brier score. An important ingredient is our conditioning not only on conditional mean and variance information, but also on conditional skewness and kurtosis information, when forming direction-of-change forecasts.

1 INTRODUCTION

Recent work by Christoffersen and Diebold (2006) has revealed a direct connection between asset return *volatility* dependence and asset return sign dependence (and hence sign forecastability). This suggests that the pervasive volatility dependence in equity returns could, through induced sign dependence, be used to produce direction-of-change forecasts useful for market timing.

To see this, let R_t be a series of returns and Ω_t be the information set available at time t. $\Pr[R_t > 0]$ is the probability of a positive return at time t. The conditional mean and variance are denoted, respectively, as $\mu_{t+1|t} = \mathrm{E}[R_{t+1}|\Omega_t]$ and $\sigma^2_{t+1|t} = \mathrm{Var}[R_{t+1}|\Omega_t]$. The return series is said to display conditional mean predictability if $\mu_{t+1|t}$ varies with Ω_t; conditional variance predictability is defined similarly. If $\Pr[R_t > 0]$ exhibits conditional dependence, ie, $\Pr[R_{t+1} > 0|\Omega_t]$ varies with Ω_t, then we say the return series is sign predictable (or the price series is direction-of-change predictable).

For clearer exposition, and to emphasize the role of volatility in return sign predictability, suppose that there is no conditional mean predictability in returns, so $\mu_{t+1|t} = \mu$ for all t. In contrast, suppose that $\sigma^2_{t+1|t}$ varies with t in a predictable manner, in keeping with the huge literature on volatility predictability reviewed in Andersen *et al* (2006). Denoting $D(\mu,\sigma^2)$ as a generic distribution dependent only on its mean μ and variance σ^2, assume

$$R_{t+1}|\Omega_t \sim D(\mu,\sigma^2_{t+1|t})$$

Then the conditional probability of positive return is

$$\begin{aligned}\Pr(R_{t+1} > 0 \mid \Omega_t) &= 1 - \Pr(R_{t+1} \le 0 \mid \Omega_t)\\ &= 1 - \Pr\left(\frac{R_{t+1} - \mu}{\sigma_{t+1|t}} \le \frac{-\mu}{\sigma_{t+1|t}}\right)\\ &= 1 - F\left(\frac{-\mu}{\sigma_{t+1|t}}\right) \qquad (1)\end{aligned}$$

where F is the distribution function of the "standardized" return $(R_{t+1|t} - \mu)/\sigma_{t+1|t})$. If the conditional volatility is predictable, then the sign of the return is predictable even if the conditional mean is unpredictable, provided $\mu \neq 0$. Note also that if the distribution is asymmetric, then the sign can be predictable even if the mean is zero: time-varying skewness can be driving sign prediction in this case.

In practice, interaction between mean and volatility can weaken or strengthen the link between conditional volatility predictability and return sign predictability. For instance, time variation in conditional means of the sort documented in recent work by Brandt and Kang (2004) and Lettau and Ludvigson (2005) would strengthen our results. Interaction between volatility and higher-ordered conditional moments can similarly affect the potency of conditional volatility as a predictor of return signs.

In this paper, we use

$$\Pr(R_{t+1} > 0 \mid \Omega_t) = 1 - F\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right) \tag{2}$$

to explore the sign predictability of one-, two- and three-month returns in three stock markets,[1] in which we examine out-of-sample predictive performance. We also use an extended version of Equation (2) that explicitly considers the interaction between volatility and higher-ordered conditional moments. We estimate the parameters of the models recursively and we evaluate the performance of sign probability forecasts. We proceed as follows: in Section 2, we discuss our data and its use for the construction of volatility forecasts; in Section 3, we discuss our direction-of-change forecasting models and evaluation methods; in Section 4, we present our empirical results; and in Section 5, we conclude.

2 DATA AND VOLATILITY FORECASTS

Estimates and forecasts of realized volatility are central to our analysis; for background, see Andersen *et al* (2003, 2006). Daily values for the period 1980:01 to 2004:06 of the MSCI index for Hong Kong, UK and US were collected from Datastream. From these, we constructed one-, two- and three-month returns and realized volatility. The latter is computed as the sum of squared daily returns within each one-, two- and three-month period. We use data from 1980:01 to 1993:12 as the starting estimation sample, which will be recursively expanded as more data becomes available. We reserve the period 1994:01 to 2004:06 for our forecasting application.

Tables 1 and 2 summarize some descriptive statistics of the return and the log of the square root of realized volatility (hereafter "log realized volatility") for the three markets. All markets have low positive mean returns for the period (see

[1] The same analysis extended to 20 stock markets is available from the authors upon request. The results for the three markets shown are representative for the 20 markets.

Table 1). Returns have negative skewness and are leptokurtic at all three frequencies. The *p*-values of the Jarque–Bera statistics indicate non-normality of all returns series. Log realized volatility is positively skewed and slightly leptokurtic (see Table 2). As with the returns series, the *p*-values of the Jarque–Bera statistics indicate non-normality of all volatility series.

Figure 1 presents the plots of the log realized volatilities. There appears to be some clustering of return volatility. Predictability of volatility is indicated by the

TABLE 1 Summary statistics of the full sample of returns, 1980:01–2004:06.

	Mean	Standard deviation	Skewness	Kurtosis	Jarque–Bera *p*-value
			Hong Kong		
1 month	0.008	0.091	–1.029	8.902	0.000
2 months	0.016	0.125	–0.422	5.046	0.000
3 months	0.025	0.164	–0.684	3.712	0.011
			UK		
1 month	0.008	0.049	–1.228	8.236	0.000
2 months	0.015	0.064	–0.611	4.186	0.000
3 months	0.023	0.085	–1.047	5.254	0.000
			US		
1 month	0.008	0.045	–0.841	6.124	0.000
2 months	0.016	0.058	–0.884	5.875	0.000
3 months	0.024	0.082	–0.799	4.237	0.000

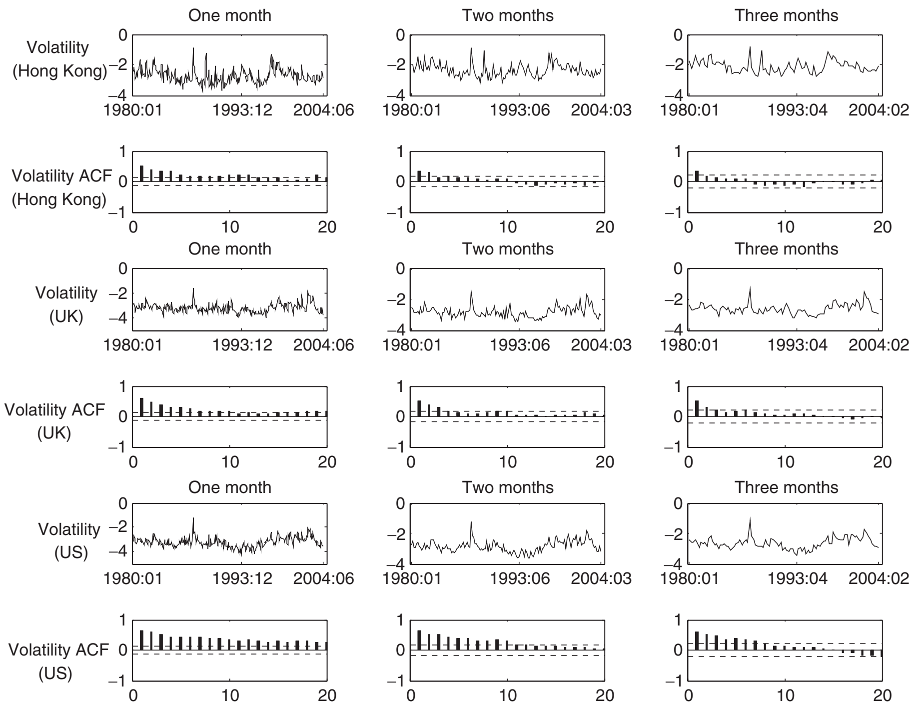
Returns are in percent per time interval (one month, two months or one quarter, not annualized).

TABLE 2 Summary statistics of the full sample of realized volatility, 1980:01–2004:06.

	Mean	Standard deviation	Skewness	Kurtosis	Jarque–Bera *p*-value
			Hong Kong		
1 month	–2.737	0.451	0.682	3.835	0.000
2 months	–2.357	0.427	0.719	3.659	0.001
3 months	–2.126	0.405	0.727	3.437	0.011
			UK		
1 month	–3.213	0.362	0.821	4.677	0.000
2 months	–2.843	0.341	0.909	4.496	0.000
3 months	–2.626	0.323	0.936	4.799	0.000
			US		
1 month	–3.226	0.400	0.566	4.619	0.000
2 months	–2.851	0.377	0.658	4.476	0.000
3 months	–2.637	0.367	0.679	4.523	0.000

"Volatility" refers to log of the square root of realized volatility computed from daily returns.

FIGURE 1 Realized volatility.



"Volatility" refers to the log of the square root of realized volatility constructed from daily returns.

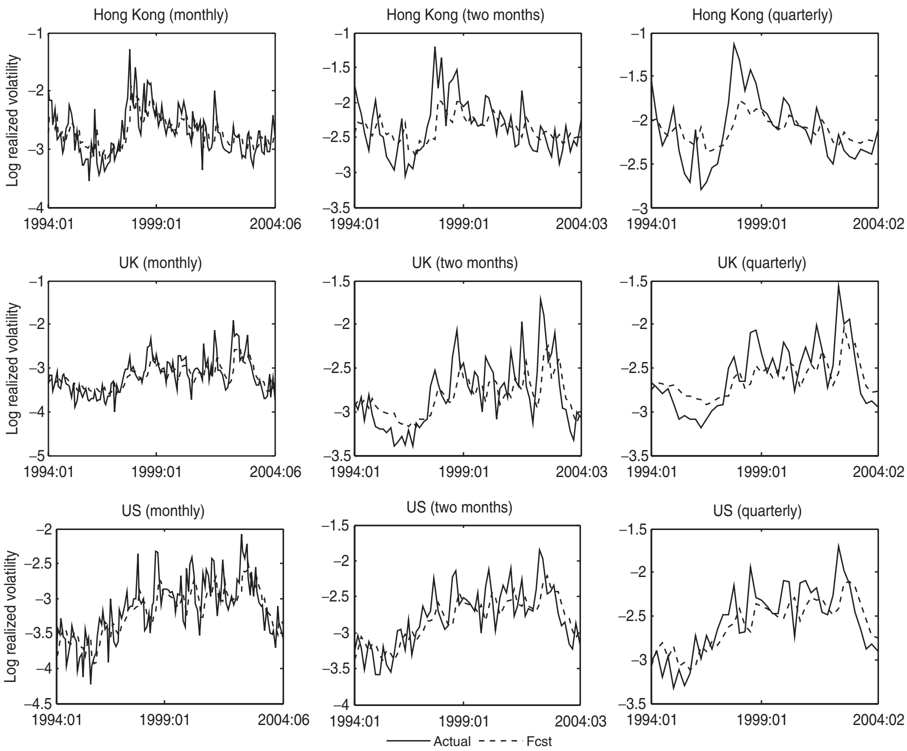
corresponding correlograms. As we move from the monthly frequency to the quarterly frequency, the autocorrelations diminish but still indicate predictability. The correlograms (none of which are reported here) of the return series of the indexes show that they are all serially uncorrelated.

Our method for forecasting the probability of positive returns will require forecasts of volatility, which we discuss here. We use the data from 1980:01 to 1993:12 as the base estimation sample. Out-of-sample one-step-ahead forecasts are generated for the period 1994:01 to 2004:06, with recursive updating of parameter estimates, ie, a volatility forecast for period $t + 1$ made at time t will use a model estimated over the period 1980:1 to t. In addition, we also choose our models recursively: at each period, we select ARMA models for log-volatility by minimizing the Akaike information criterion (AIC).[2]

[2] We repeated the analysis using the SIC criterion, but because the subsequent probability forecasts generated by the Schwarz information criterion (SIC), and the corresponding evaluation results, are very similar to the AIC results, we report results only for the models selected by the AIC. The SIC results are available from the authors upon request.

Broadly speaking, the AIC favors ARMA(1,1) models, particular at the monthly frequency. In Figure 2, we display the volatility forecasts (with actual log realized volatilities included for comparison) for the three markets. The forecasts generated by the AIC track actual log realized volatility fairly reasonably. The ratios of the mean square prediction errors (MSPEs) to the sample variance of log realized volatility are given in Table 3. The forecasts capture a substantial amount of the variation in actual log realized volatility.

FIGURE 2 Realized volatility and recursive realized volatility forecasts.



"Volatility" refers to log of the square root of realized volatility constructed from daily returns. "Fcst" is the one-step-ahead forecasts of volatility generated from recursively estimated ARMA models chosen recursively using the AIC criterion.

TABLE 3 Ratio of mean square prediction error (MSPE) of forecasts to sample variance, realized volatility.

	Hong Kong	UK	US
1 month	0.479	0.648	0.537
2 months	0.632	0.756	0.592
3 months	0.575	0.855	0.563

A comment on our notation: throughout this paper, we use $\hat{\sigma}_t$ to represent the square root of realized volatility. The symbol $\hat{\sigma}_{t+1|t}$ will represent the period t forecast of the square root of period $t + 1$ realized volatility. Note also that our volatility forecasting models use (and forecast) the log of these objects, so that $\hat{\sigma}_{t+1|t}$ actually represents the exponent of the forecasts of (log) realized volatility. Finally, for simplicity of notation, we will also write $\Pr[R_{t+1|t} > 0]$ for $\Pr[R_{t+1} > 0|\Omega_t]$.

3 FORECASTING MODELS AND EVALUATION METHODS

3.1 Forecasting models

We will evaluate the forecasting performance of two sets of forecasts and compare their performance against forecasts from a baseline model. Our baseline forecasts are generated using the empirical cumulative distribution function (cdf) of the R_t using data from the beginning of our sample period right up to the time the forecast is made, ie, at period k, we compute

$$\widehat{\Pr}\,(R_{k+1|k} > 0) = \frac{1}{k}\sum_{t=1}^{k} I(R_t > 0) \tag{3}$$

where $I(\cdot)$ is the indicator function.

Our first forecasting model makes direct use of Equation (2). Using all available data at time k, we first regress R_t on a constant, $\log(\hat{\sigma}_t)$, and $[\log(\hat{\sigma}_t)]^2$, and compute

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1\log(\hat{\sigma}_t) + \hat{\beta}_2[\log(\hat{\sigma}_t)]^2, \quad t = 1,\ldots,k \tag{4}$$

where $\hat{\sigma}_t$ is the square root of (actual, not forecasted) realized volatility. The period $k+1$ forecast is then generated by

$$\begin{aligned}\widehat{\Pr}(R_{k+1|k} > 0) &= 1 - \hat{F}\left(-\frac{\hat{\mu}_{k+1|k}}{\hat{\sigma}_{k+1|k}}\right)\\ &= 1 - \frac{1}{k}\sum_{t=1}^{k} I\left(\frac{R_t - \hat{\mu}_t}{\hat{\sigma}_t} \le \frac{\hat{\mu}_{k+1|k}}{\hat{\sigma}_{k+1|k}}\right)\end{aligned} \tag{5}$$

ie, $\hat{F}$ is the empirical cdf of $(R_t - \hat{\mu}_t)/\hat{\sigma}_t$. The one-step-ahead volatility forecast $\hat{\sigma}_{t+1|t}$ is generated from a recursively estimated model selected, at each period, by minimizing the AIC, as described in the previous section. The one-step-ahead mean forecast $\hat{\mu}_{t+1|t}$ is computed as

$$\hat{\mu}_{t+1|t} = \hat{\beta}_0 + \hat{\beta}_1\log(\hat{\sigma}_{t+1|t}) + \hat{\beta}_2[\log(\hat{\sigma}_{t+1|t})]^2 \tag{6}$$

The coefficients $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are recursively estimated using Equation (4).

A linear relationship between the return mean and the time-varying log return volatility was first found in the seminal ARCH-in-Mean work of Engle *et al* (1987). A quadratic return mean specification is used here as the quadratic term is significant for almost all series in the starting estimation sample. Although the coefficients are recursively estimated, at each recursion no attempt is made to refine the model. We refer to forecasts from Equation (5) as non-parametric forecasts (even though the realized volatility forecasts are generated using fully parametric models) to differentiate it from forecasts from our next model.

The second model is an extension of Equation (1) and explicitly considers the interaction between volatility, skewness and kurtosis. This is done by using the Gram–Charlier expansion:

$$1 - F\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right) \approx 1 - \Phi\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right) + \Phi\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right)\left[\frac{\gamma_{3,t+1|t}}{3!}\left(\frac{\mu^2_{t+1|t}}{\sigma^2_{t+1|t}} - 1\right) + \frac{\gamma_{4,t+1|t}}{4!}\left(\frac{-\mu^3_{t+1|t}}{\sigma^3_{t+1|t}} + \frac{3\mu_{t+1|t}}{\sigma_{t+1|t}}\right)\right]$$

where $\Phi(\cdot)$ is the distribution function of a standard normal, and γ_3 and γ_4 are, respectively, the skewness and excess kurtosis, with the usual notation for conditioning on Ω_t. This equation can be rewritten as

$$1 - F(-\mu_{t+1|t}\, x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t}\, x_{t+1}) (\beta_{0t} + \beta_{1t} x_{t+1} + \beta_{2t} x^2_{t+1} + \beta_{3t} x^3_{t+1})$$

with $\beta_{0t} = 1 + \gamma_{3,t-1|t}/6$, $\beta_{1t} = -\gamma_{4,t+1|t}\,\mu_{t+1|t}/8$, $\beta_{2t} = -\gamma_{3,t+1|t}\,\mu^2_{t+1|t}/6$ and $\beta_{3t} = \gamma_{4,t+1|t}\,\mu^3_{t+1|t}/24$, where for notational convenience, we denote $x_{t+1} = 1/\sigma_{t+1|t}$.

Several points should be noted. The sign of returns is predictable for non-zero $\mu_{t+1|t}$ even when there is no volatility clustering, as long as the skewness and kurtosis are time varying. On the other hand, even if $\mu_{t+1|t}$ is zero, sign predictability arises as long as conditional skewness dynamics is present, regardless of whether volatility dynamics is present. If there is no conditional skewness and excess kurtosis, the above equation is reduced to

$$1 - F(-\mu_{t+1|t}\, x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t}\, x_{t+1})$$

so that normal approximation applies. If returns are conditionally symmetric but leptokurtic (ie, $\gamma_{3,t+1|t} = 0$ and $\gamma_{4,t+1|t} > 0$), then $\beta_{0t} = 1$ and $\beta_{2t} = 0$, and we have

$$1 - F(-\mu_{t+1|t}\, x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t}\, x_{t+1})(1 + \beta_{1t} x_{t+1} + \beta_{3t} x^3_{t+1})$$

Furthermore, if $\mu_{t+1|t} > 0$, we have $\beta_{1t} < 0$ and $\beta_{1t} > 0$, and the converse is true for $\mu_{t+1|t} < 0$. Finally, if $\mu_{t+1|t}$ is small, as in the case of short investment horizons, then β_{2t} and β_{3t} can safely be ignored, resulting in

$$1 - F(-\mu_{t+1|t}\, x_{t+1}) \approx 1 - \Phi(-\mu_{t+1|t}\, x_{t+1})(\beta_{0t} + \beta_{1t} x_{t+1})$$

Thus, conditional skewness affects sign predictability through β_{0t}, and conditional kurtosis affects sign predictability through β_{1t}. When there is no conditional dynamics in skewness and kurtosis, the above equation is reduced to

$$1 - F(-\mu_{t+1|t}\, x_{t+1}) \simeq 1 - \Phi(-\mu_{t+1|t}\, x_{t+1})(\beta_0 + \beta_1 x_{t+1}) \quad (7)$$

for some time-invariant quantities β_0 and β_1.

We use Equation (7) as our second model for sign prediction, ie, we generate forecasts of the probability of positive returns as

$$\widehat{\Pr}(R_{t+1|t} > 0) = 1 - \Phi(-\hat{\mu}_{t+1|t}\, \hat{x}_{t+1})(\hat{\beta}_0 + \hat{\beta}_1\, \hat{x}_{t+1}) \quad (8)$$

where $\hat{x}_{t+1|t} = 1/\hat{\sigma}_{t+1|t}$, and where $\hat{\mu}_{t+1|t}$ and $\hat{\sigma}_{t=t+1|t}$ are as defined earlier. We refer to these as forecasts from the "extended" model. The parameters β_0 and β_1 are estimated by regressing $1 - I(R_t > 0)$ on $\Phi(-\hat{\mu}_t \hat{x}_t)$ and $\Phi(-\hat{\mu}_t \hat{x}_t)\hat{x}_t$ for $t = 1,\ldots,$k. Although we have not explicitly placed any constraints on this model to require $\Phi(-\hat{\mu}_t \hat{x}_t)(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_t)$ to lie between 0 and 1, this was inconsequential as all our predicted probabilities turn out to lie between 0 and 1.

3.2 Forecast evaluation

We perform post-sample comparison of the forecast performance of Equations (5) and (8) for the sign of return. Both are compared against baseline forecasts [Equation (3)]. This is done for one-, two- and three-month returns. We assess the performance of the forecasting models using Brier scores; for background, see Diebold and Lopez (1996).

Two Brier scores are used:

$$\text{Brier(Abs)} = \frac{1}{T-k} \sum_{t=k}^{T} |\widehat{\Pr}(R_{t+1|t} > 0) - z_{t+1}|$$

$$\text{Brier(Sq)} = \frac{1}{T-k} \sum_{t=k}^{T} 2(\widehat{\Pr}(R_{t+1|t} > 0) - z_{t+1})^2$$

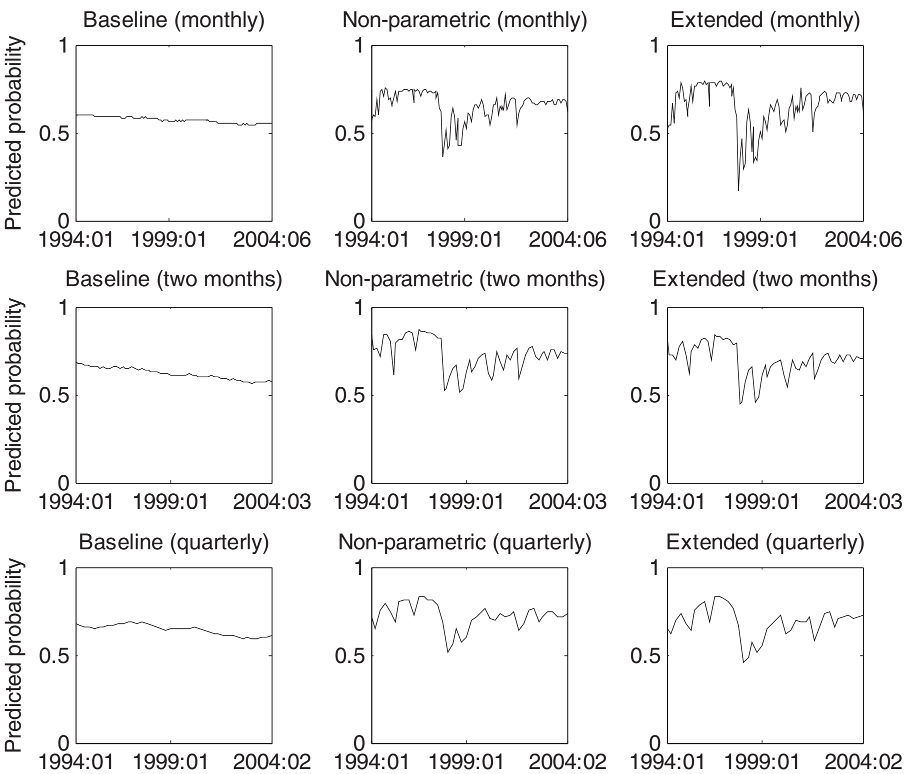
where $z_{t+1} = I(R_{t+1} > 0)$. The latter is the traditional Brier score for evaluating the performance of probability forecasts and is analogous to the usual MSPE. A score of 0 for Brier(Sq) occurs when perfect forecasts are made: where at each period, correct probability forecasts of 0 or 1 are made. The worst score is 2 and occurs if at each period probability forecasts of 0 or 1 are made but turn out to be wrong each time. Note that if we follow the usual convention where a correct probability forecast of $I(R_{t+1} > 0)$ is 1 that is greater than 0.5, then correct forecasts will have an individual Brier(Sq) score between 0 and 0.5, whereas incorrect forecasts have individual scores between 0.5 and 2. A few incorrect forecasts can therefore dominate a majority of correct forecasts.

For this reason, we also consider a modified version of the Brier score, which we call Brier(Abs). Like Brier(Sq), the best possible score for Brier(Abs) is 0. The worst score is 1. Here correct forecasts have individual scores between 0 and 0.5, whereas incorrect forecasts carry scores between 0.5 and 1.

4 EMPIRICAL RESULTS

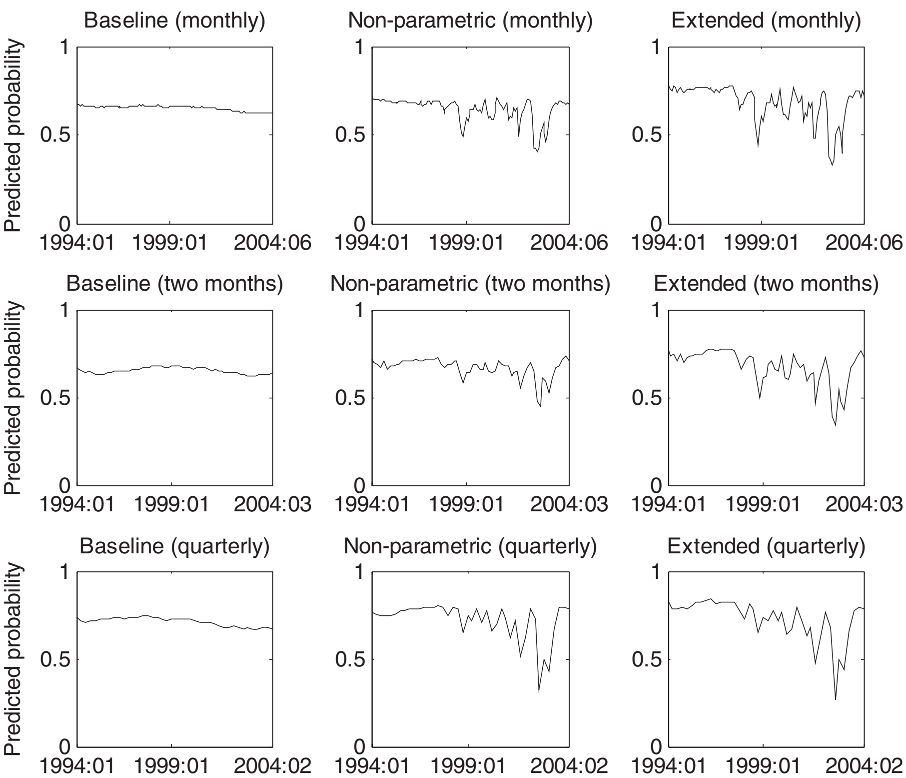
Figures 3–5 show, for the Hong Kong, UK and US markets, respectively, the predicted probabilities generated by the baseline model, the non-parametric model

FIGURE 3 Predicted probabilities (Hong Kong).



"Nonparametric" forecasts (second column) refer to forecasts generated using $\widehat{\Pr}(R_{t+1} > 0) = 1 - \hat{F}(-\hat{\mu}_{t+1|t}/\hat{\sigma}_{t+1|t})$, where $\hat{F}$ is the empirical cdf of $(R_t - \hat{\mu}_t)/\hat{\sigma}_t$. "Extended" (third column) refers to forecasts generated from the extended model $\widehat{\Pr}(R_{t+1} > 0) = 1 - \Phi(-\hat{\mu}_t\, \hat{x}_{t+1})\, (\hat{\beta}_0 + \hat{\beta}_1\, \hat{x}_{t+1})$.

FIGURE 4 Predicted probabilities (UK).

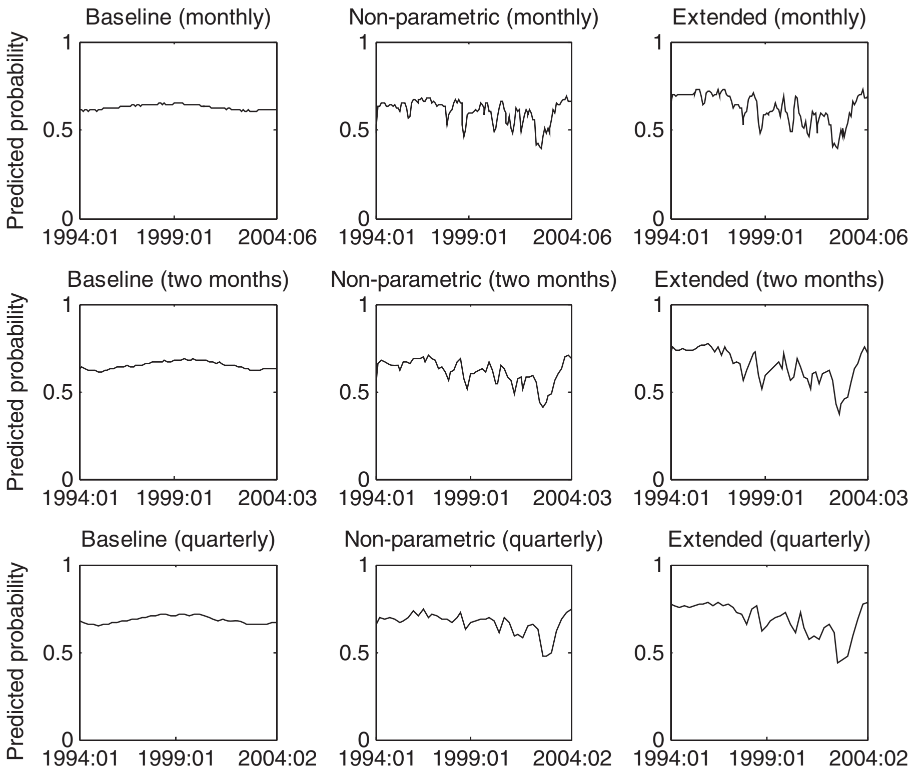


See Figure 3 footnote.

and the extended model (columns 1, 2 and 3, respectively) for the monthly, two-month and quarterly frequencies (rows 1, 2 and 3, respectively). For the non-parametric and extended models, forecasts based on the AIC volatility forecasts are plotted. For all three markets, the baseline forecasts are very flat, at values slightly above 0.5. The non-parametric and extended forecasts show more variability, especially in the later periods.

Before reporting our main results, we highlight some interesting regularities in the Brier scores. In Table 4, we report the mean and standard deviation of the Brier(Abs) scores from the AIC-based probability forecasts for the Hong Kong, UK and US markets. Results are reported for three "subperiods". The 'low' volatility subperiod comprises all dates for which realized volatility falls in the 1st to 33rd percentile range. The 'medium' and 'high' volatility subperiods comprise all dates for which realized volatility falls in the 34–66th and 67–100th percentile ranges, respectively. In all three markets, the Brier score for the low volatility subperiod is lower than the corresponding Brier score for the high volatility subperiod. In contrast, the standard deviations of the Brier scores for the non-parametric

FIGURE 5 Predicted probabilities (US).



See Figure 3 footnote.

and extended models are higher in the low volatility subperiods than in the high volatility subperiods. For instance, the mean Brier score for the extended model in the US market at the monthly frequency is 0.378 in the low volatility subperiod and 0.532 in the high volatility subperiod. The standard deviation of the same Brier scores falls from 0.143 in the low volatility subperiod to 0.088 in the high volatility subperiod. These findings seem perfectly reasonable: we should expect our models to have more to say in subperiods of low volatility and little to say in subperiods of high volatility. In high volatility subperiods, the models tend to generate probability forecasts that are close to 0.5. The corresponding Brier scores in turn tend to be close to 0.5, resulting in the lower standard deviation of Brier scores in high volatility subperiods.

Our main results are reported in Tables 5–8. Table 5 contains our results for the full forecast period. Tables 6–8 contain the results for the low, medium and high volatility subperiods, respectively. In all four tables, both Brier(Abs) and Brier(Sq) are given for the baseline model. The Brier scores for the non-parametric and extended models are expressed relative to the baseline Brier

TABLE 4 Forecast performance, Brier(Abs), three markets.

		Baseline		Nonparametric		Extended	
		Mean	**Standard deviation**	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**
				Hong Kong			
1 month	Low volatility	0.507	0.075	0.510	0.208	0.504	0.239
	Medium volatility	0.491	0.070	0.475	0.168	0.475	0.182
	High volatility	0.520	0.076	0.537	0.137	0.535	0.151
2 months	Low volatility	0.512	0.126	0.518	0.298	0.512	0.265
	Medium volatility	0.487	0.121	0.487	0.230	0.488	0.203
	High volatility	0.579	0.107	0.640	0.173	0.622	0.145
3 months	Low volatility	0.435	0.132	0.390	0.241	0.402	0.223
	Medium volatility	0.515	0.144	0.526	0.256	0.524	0.222
	High volatility	0.551	0.154	0.548	0.202	0.534	0.159
				UK			
1 month	Low volatility	0.430	0.137	0.418	0.164	0.388	0.224
	Medium volatility	0.507	0.157	0.516	0.165	0.519	0.215
	High volatility	0.500	0.151	0.522	0.123	0.531	0.153
2 months	Low volatility	0.398	0.106	0.362	0.147	0.333	0.181
	Medium volatility	0.511	0.160	0.526	0.178	0.546	0.193
	High volatility	0.517	0.160	0.514	0.154	0.518	0.158
3 months	Low volatility	0.343	0.156	0.308	0.204	0.288	0.226
	Medium volatility	0.470	0.220	0.484	0.247	0.491	0.254
	High volatility	0.566	0.201	0.638	0.173	0.638	0.162

TABLE 4 Continued.

		Baseline		Nonparametric		Extended	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
				US			
1 month	Low volatility	0.419	0.093	0.404	0.104	0.378	0.143
	Medium volatility	0.468	0.129	0.488	0.127	0.489	0.144
	High volatility	0.525	0.130	0.536	0.088	0.532	0.088
2 months	Low volatility	0.417	0.109	0.412	0.134	0.375	0.196
	Medium volatility	0.451	0.149	0.466	0.125	0.461	0.157
	High volatility	0.518	0.165	0.499	0.107	0.502	0.131
3 months	Low volatility	0.433	0.162	0.405	0.172	0.379	0.231
	Medium volatility	0.366	0.143	0.380	0.139	0.364	0.158
	High volatility	0.524	0.199	0.529	0.158	0.528	0.165

$\text{Brier(Abs)} = 1/T \sum_{t=k}^{T} |p_{t+1|t} - z_{t+1}|$, where k is the start of the estimation sample, $p_{t+1|t}$ is the one-step-ahead forecast of $\Pr(R_{t+1} > 0)$ made at time t and $z_{t+1} = I(R_{t+1} > 0)$. At each time t, data from the first period up to time t is used to estimate the forecasting model. "Baseline" refers to forecasts generated from the unconditional empirical distribution of R_t. "Non-parametric" refers to forecasts generated using $\widehat{\Pr}(R_{t+1|t} > 0) = 1 - \hat{F}(-\hat{\mu}_{t+1|t}/\hat{\sigma}_{t+1|t})$, where $\hat{F}$ is the empirical cdf of $(R_t - \hat{\mu}_t)/\hat{\sigma}_t$. "Extended" refers to forecasts generated from $\widehat{\Pr}(R_{t+1|t} > 0) = 1 - \Phi(-\hat{\mu}_{t+1|t}\hat{x}_{t+1})(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{t+1})$.

TABLE 5 Relative forecast performance (full sample).

	Number of forecasts	Brier(Abs)			Brier(Sq)		
		Baseline	**Non-parametric**	**Extended**	**Baseline**	**Non-parametric**	**Extended**
Hong Kong							
1 month	126	0.506	1.005	1.001	0.524	1.102	1.123
2 months	63	0.526	1.043	1.028	0.584	1.231	1.158
3 months	42	0.500	**0.975**	**0.973**	0.544	1.081	1.027
UK							
1 month	126	0.482	1.013	1.003	0.510	1.031	1.084
2 months	63	0.477	**0.984**	**0.981**	0.500	**0.999**	1.031
3 months	42	0.460	1.036	1.027	0.510	1.124	1.129
US							
1 month	126	0.469	1.011	**0.991**	0.471	1.015	1.003
2 months	63	0.462	**0.994**	**0.967**	0.470	**0.964**	**0.969**
3 months	42	0.441	**0.993**	**0.961**	0.450	**0.972**	**0.966**

$\text{Brier(Abs)} = 1/T \sum_{t=k}^{T} |p_{t+1|t} - z_{t+1}|$ and $\text{Brier(Sq)} = 1/T \sum_{t=k}^{T} 2(p_{t+1|t} - z_{t+1})^2$, where k is the start of the estimation sample, $p_{t+1|t}$ is the one-step-ahead forecast of $\Pr(R_{t+1} > 0)$ made at time t and $z_{t+1} = I(R_{t+1} > 0)$. At each time t, data from the first period up to time t is used to estimate the forecasting model. "Baseline" refer to forecasts generated from the unconditional empirical distribution of R_t. "Non-parametric" refers to forecasts generated using $\widehat{\Pr}(R_{t+1|t} > 0) = 1 - \hat{F}(-\hat{\mu}_{t+1|t}/\hat{\sigma}_{t+1|t})$, where $\hat{F}$ is the empirical cdf of $(R_t - \hat{\mu}_t)/\hat{\sigma}_t$. "Extended" refers to forecasts generated from $\widehat{\Pr}(R_{t+1|t} > 0) = 1 - \Phi\left(-\hat{\mu}_{t+1|t}\hat{x}_{t+1}\right)\left(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{t+1}\right)$. Actual Brier scores are reported for the baseline forecasts. All other scores are Brier scores for the given model divided by the Brier score for the baseline forecast. Ratios below 1 are in bold print.

TABLE 6 Forecast performance, low volatility periods (1st to 33rd percentile).

	Number of forecasts	Brier(Abs)			Brier(Sq)		
		Baseline	**Non-parametric**	**Extended**	**Baseline**	**Non-parametric**	**Extended**
Hong Kong							
1 month	42	0.507	1.004	**0.993**	0.526	1.148	1.178
2 months	21	0.512	1.011	1.000	0.555	1.269	1.186
3 months	14	0.435	**0.896**	**0.923**	0.411	1.002	1.010
UK							
1 month	42	0.430	**0.972**	**0.904**	0.406	**0.990**	**0.985**
2 months	21	0.398	**0.909**	**0.836**	0.338	**0.896**	**0.839**
3 months	14	0.343	**0.897**	**0.839**	0.281	**0.950**	**0.929**
US							
1 month	42	0.419	**0.964**	**0.903**	0.368	**0.944**	**0.887**
2 months	21	0.417	**0.989**	**0.899**	0.370	1.011	**0.956**
3 months	14	0.433	**0.934**	**0.875**	0.423	**0.902**	**0.911**

See Table 5 footnote.

TABLE 7 Forecast performance, medium volatility periods (34–66th percentile).

	Number of forecasts	Brier(Abs)			Brier(Sq)		
		Baseline	Non-parametric	Extended	Baseline	Non-parametric	Extended
Hong Kong							
1 month	42	0.491	**0.967**	**0.966**	0.492	1.029	1.046
2 months	21	0.487	1.000	1.002	0.502	1.144	1.104
3 months	14	0.515	1.022	1.019	0.568	1.187	1.129
UK							
1 month	42	0.507	1.017	1.023	0.563	1.040	1.116
2 months	21	0.511	1.028	1.068	0.572	1.072	1.168
3 months	14	0.470	1.029	1.045	0.532	1.093	1.133
US							
1 month	42	0.468	1.043	1.047	0.470	1.080	1.106
2 months	21	0.451	1.032	1.023	0.449	1.031	1.052
3 months	14	0.366	1.040	**0.996**	0.306	1.065	1.021

See Table 5 footnote.

TABLE 8 Forecast performance, high volatility periods (66th–100th percentile).

	Number of forecasts	Brier(Abs)			Brier(Sq)		
		Baseline	Non-parametric	Extended	Baseline	Non-parametric	Extended
Hong Kong							
1 month	42	0.520	1.032	1.029	0.552	1.110	1.118
2 months	21	0.579	1.105	1.075	0.692	1.265	1.176
3 months	14	0.551	**0.995**	**0.970**	0.652	1.040	**0.949**
UK							
1 month	42	0.500	1.044	1.063	0.544	1.053	1.121
2 months	21	0.517	**0.995**	1.002	0.583	**0.984**	1.002
3 months	14	0.566	1.127	1.126	0.717	1.214	1.204
US							
1 month	42	0.525	1.021	1.015	0.583	1.009	**0.997**
2 months	21	0.518	**0.964**	**0.969**	0.588	**0.885**	**0.912**
3 months	14	0.524	1.009	1.008	0.622	**0.974**	**0.977**

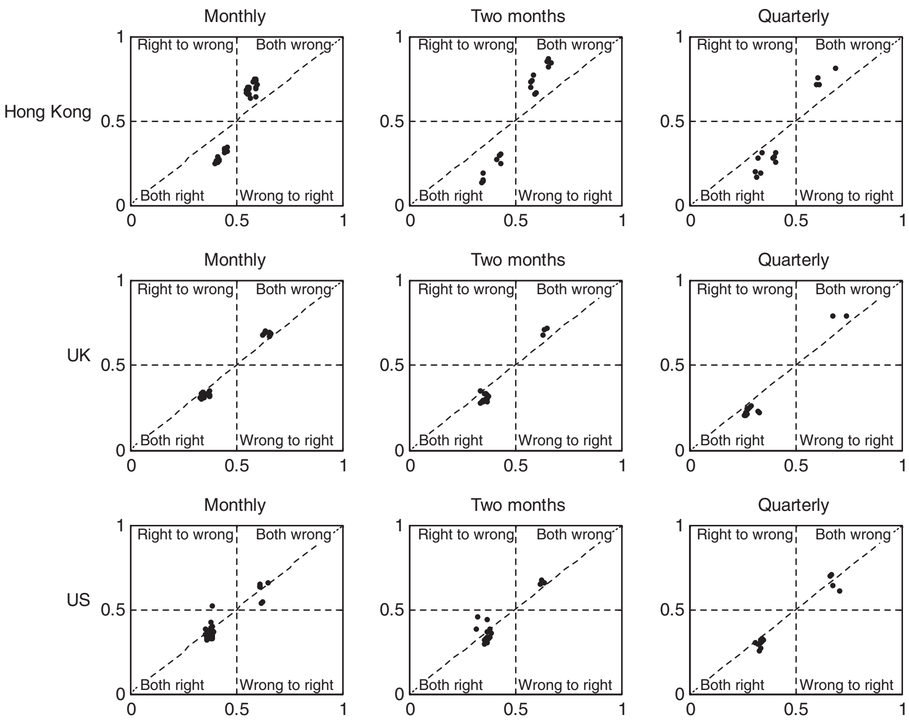
See Table 5 footnote.

scores. A relative measure of less than 1 therefore implies improvement in forecast performance.

Table 5 reports improvement in the performance of the non-parametric or extended models over the baseline forecasts in half of the cases considered, when using Brier(Abs) as a measure of performance. All of the improvements, however, are very small. The situation is usually worse when the forecasts are evaluated using Brier(Sq) instead.

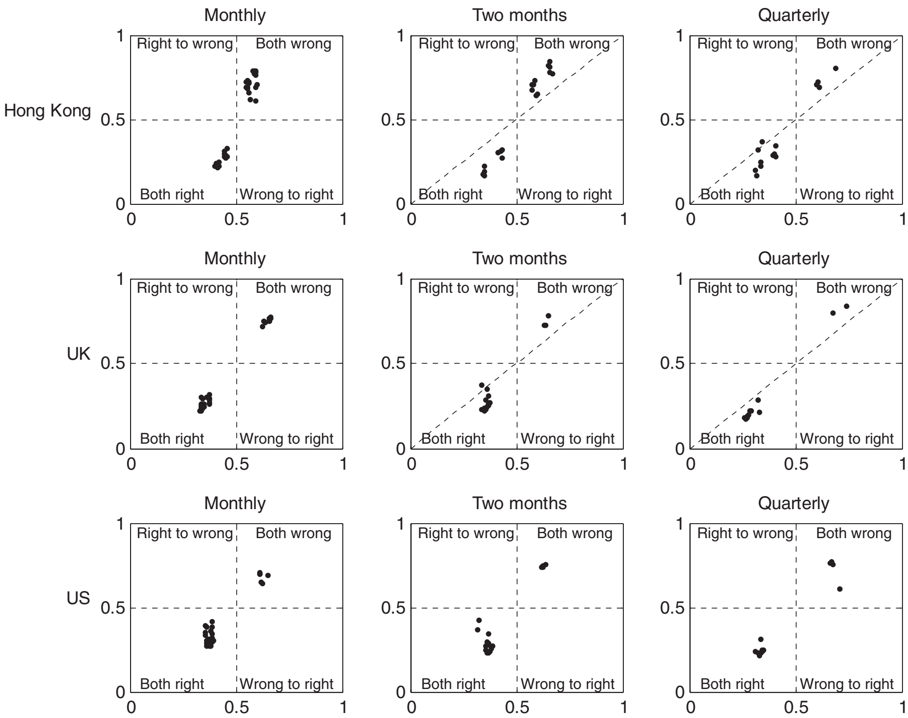
The fact that the non-parametric and extended models perform better during low volatility subperiods than during high volatility subperiods suggests that their performance relative to the baseline model might be better during low volatility subperiods than during high volatility subperiods. This is verified by the relative performances reported in Tables 6–8. In Table 6, the improvements are widespread and sometimes large. In a number of cases, the ratio of the Brier(Abs) scores for the extended/parametric models to the baseline model is less than 0.9.

FIGURE 6 Comparative Brier(Abs) scores, low volatility (nonparametric versus baseline).



The horizontal axis measures individual Brier(Abs) scores for baseline forecasts. The vertical axis measures corresponding Brier(Abs) scores for the nonparametric forecasts. A score below 0.5 indicates a correct forecast. Only observations with volatility in the 1st to 33rd percentile range are included.

FIGURE 7 Comparative Brier(Abs) scores, low volatility (extended model versus baseline).



The horizontal axis measures individual Brier(Abs) scores for baseline forecasts. The vertical axis measures corresponding Brier(Abs) scores for the extended forecasts. A score below 0.5 indicates a correct forecast. Only observations with volatility in the 1st to 33rd percentile range are included.

When Brier(Sq) is used to measure forecast performance, there are fewer instances where the non-parametric and extended models perform better than the baseline. The notable differences between the Brier(Sq) scores and Brier(Abs) scores occur for Hong Kong, where Brier(Sq) shows no improvements from the non-parametric and extended models.

The ratios under Brier(Abs) also show that the extended model performs much better than the non-parametric model. We note also that for both Brier(Abs) and Brier(Sq), the performance of the non-parametric and extended models, relative to the baseline, is generally better at the quarterly frequency than at the monthly frequency. This is to be expected, as the theory indicates that volatility-aided prediction depends on a sizable mean return, and the mean return increases in all markets as we go from monthly to quarterly frequencies.

In the medium and high volatility subperiods in Tables 7 and 8, respectively, much less improvement in the performance of the non-parametric and extended models is found. It appears that volatility in these subperiods is simply too large relative to the mean to be useful in guiding direction-of-change forecasts.

Figures 6 and 7 show a clear picture of the forecast performance of the non-parametric and extended models compared to the baseline forecasts. At each frequency, we show a scatterplot of the Brier(Abs) scores of individual forecasts. We include only observations when volatility is low, as previously defined. In both figures, the horizontal axis measures the Brier(Abs) scores for individual baseline forecasts. In Figure 6, the vertical axis measures the Brier(Abs) scores for individual non-parametric forecasts. In Figure 7, the vertical axis measures the Brier(Abs) scores for individual forecasts from the extended model. In addition to the scatterplots, we include a horizontal and vertical gridline at 0.5 and a 45° line. As a Brier(Abs) score below 0.5 indicates a correct forecast, points in the lower left quadrant indicate that both competing forecasts are correct, whereas a point in the lower right quadrant indicates that the baseline forecast for this observation is incorrect, with the competing forecast correct. Points below the 45° line indicate improvements in the Brier(Abs) scores over the baseline.

In all three markets, the non-parametric and extended models clearly provide better signals than the baseline model when both the baseline and the competing forecasts are correct. However, for Hong Kong and UK, the performance of the non-parametric and extended model is worse than the baseline model when the baseline and the competing forecasts are wrong. Note that the upper left and lower right quadrants of Figures 6 and 7 are mostly empty, which implies that the models by and large make predictions that are similar to the baseline forecasts. Nonetheless, there is evidence that when volatility is low, forecasts of volatility can improve the quality of the signal, in the sense of providing probability forecasts with improved Brier scores.

5 SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

Methodologically, we have extended the Christoffersen and Diebold (2006) direction-of-change forecasting framework to include the potentially important effects of higher-ordered conditional moments. Empirically, in an application to a sample of three equity markets, we have verified the importance of allowing for higher-ordered conditional moments and taken a step toward evaluating the real-time predictive performance. In future work, we look forward to using our direction-of-change forecasts to formulate and evaluate actual trading strategies and to exploring their relationships to the "volatility timing" strategies recently studied by Fleming *et al* (2003), in which portfolio shares are dynamically adjusted based on forecasts of the variance–covariance matrix of the underlying assets.

REFERENCES

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71**, 579–626.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of Economic Forecasting*, Elliott, G., Granger, C. W. J. and Timmermann, A. (eds). North-Holland, Amsterdam 777–878.

Brandt, M. W., and Kang, Q. (2004). On the relationship between the conditional mean and volatility of stock returns: a latent VAR approach. *Journal of Financial Economics* **72**, 217–257.

Christoffersen, P. F., and Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* **52**, 1273–1287.

Diebold, F. X. and Lopez, J. (1996). Forecast evaluation and combination. *Handbook of Statistics*, Maddala, G. S. and Rao, C. R. (eds). North-Holland, Amsterdam, pp. 241–268.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the Arch-M model. *Econometrica* **55**, 391–407.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics* **67**, 473–509.

Lettau, M., and Ludvigson, S. (2005). Measuring and modeling variation in the risk-return tradeoff. *Handbook of Financial Econometrics*, Ait-Shalia, Y. and Hansen, L. P. (eds). North-Holland, Amsterdam, forthcoming.