Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2007/04

Volatility Forecasting Using Explanatory Variables and Focused Selection Criteria

Christian T. Brownlees,
Giampiero M. Gallo

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Volatility Forecasting Using Explanatory Variables and Focused Selection Criteria

Christian T. Brownlees* Giampiero M. Gallo*

This version: May 2007

Abstract

This paper assesses the performance of volatility forecasting using focused selection and combination strategies to include relevant explanatory variables in the forecasting model for realized volatility. The focused selection/combination strategies consist of picking up the model that minimizes the estimated risk (e.g. MSE) of a given smooth function of the parameters of interest to the forecaster. The proposed focused methods are compared with other strategies, including the well established AIC and BIC. The methodology is applied to a daily recursive 1–step ahead value–at–risk (VaR) forecasting exercise of 4 widely traded New York Stock Exchange stocks. Results show that VaR forecasts can significantly be improved upon using focused forecast strategies for the selection of relevant predetermined variables. The set of explanatory variables that helps improving prediction is stock dependent: however leverage effects, daily range and overnight volatility proves to be relevant. In line with recent theoretical findings, the predictive performance of the BIC (which tends to overly exclude useful explanatory variables) appears to be modest.

Keywords: Forecasting, Volatility, Value–at–Risk, Realized Volatility, Model Selection, FIC, AIC, BIC
JEL: C22, C52, C53

*Dipartimento di Statistica "G. Parenti", Viale G.B. Morgagni 59, I-50134 Firenze, Italy, e-mail: ctb@ds.unifi.it gallog@ds.unifi.it.

1 Introduction

The financial econometric literature has devoted much attention to the modelling of the dynamics of volatility, exploiting the features of clustering exhibited by returns (Engle (1982), Bollerslev (1986)). In spite of some notable exceptions in the early '90s (e.g. Lamoureux & Lastrapes (1990)), the role of explanatory variables on driving the evolution of conditional variance (other than the returns' own past) in forecasting has received less attention. This is possibly due to the difficulties in jointly modeling the variables involved (as in, e.g., Lamoureux & Lastrapes (1994) for returns and volumes) in a multi–step forecasting perspective. The potential relevance in exploiting available information is stressed by Poon & Granger (2005) (cf. the references therein), who remark that while some studies have established the explanatory relevance of some predetermined variables for volatility, they have not been conclusive about their actual contribution in improving forecasting.

The scenario assumed in this work is that we are interested in modeling and projecting the conditional expectation of a volatility proxy on the basis of an information set containing a group of potentially useful explanatory variables. We will make use of some model selection/combination scheme is used to select/weigh predictions constructed from volatility models using different subsets of explanatory variables.

Specifically, we propose the use *focused* selection and combination strategies for volatility prediction using explanatory variables. The term "focused" refers to a wide family of model selection devices called Focused Information Criteria (FIC) introduced by Claeskens & Hjort (2003). The focused selection/combination strategies consist of picking up the model that minimizes the estimated risk (e.g., but not limited to, MSE) of a given smooth function of the parameters of interest to the forecaster: for example, precision in the estimation of a (nonlinear) function of the parameters (e.g. the persistence, the unconditional variance or the half-life of a shock in a GARCH model) may be more important than that of single parameters.

We are interested in fitting and predicting the realized volatility with a nested set of Multiplicative Error Models (Engle (2002), Engle & Gallo (2006)) with an expanded information set including explanatory variables other than the realized volatility own past.

The aim of this present paper is therefore to assess whether the predictions of a parsimonious benchmark volatility model can be beaten using focused strategies. The performance of focused methods is also compared with the well established AIC and BIC. The ultimate prediction problem of interest is a daily recursive 1–step ahead VaR forecasting application, where at each period the set of volatility models is estimated and model selection/combination strategies are used to select/combine VaR forecasts. The evaluation framework of the predictions explicitly accommodates a risk-management viewpoint.

The empirical applications involves four stocks widely traded at the New York Stock Exchange (NYSE): AXP (American Express), GE (General Electric), MCD (McDonald's) and WMT (Walmart). The sample period spans from January 2, 2001[1]to December 30, 2005 (1256 trading days). Daily frequency series are constructed from the TAQ high frequency database using the procedures described in Brownlees & Gallo (2006). Note that the pragmatic view adopted in the paper is that it is of interest to assess which are *useful forecasting methods*. This is somehow different from a large segment of the literature which seems to be more concerned in searching for the *best forecasting model*.

The empirical results of the analysis show that it is possible to beat the benchmark using focused strategies. In most cases, focused selection strategies for VaR forecasting produce predictions that are more accurate than the benchmark using a risk management loss metric. Such improvement appears to be statistically significant using standard predictive ability tests. These results are also interesting given that benchmark parsimonious models for volatility like a $GARCH(1,1)$ do not appear easy to beat (cf. Hansen & Lunde (2005)). The set of explanatory variables that help improving forecasts appears to be stock dependent. Another empirical result that emerges from this application is that traditional model selection criteria such as the AIC and BIC are not too successful in selecting better forecasts. In particular the BIC produces forecasts that appear to be very close to the ones produced by the benchmark base model.

There is a number of different contributions in the literature that relate to this work. The literature on the FIC family of selection devices was introduced by Hjort & Claeskens (2003) and Claeskens & Hjort (2003). Further extensions are proposed in Claeskens, Croux & Van Kerckhoven (2006) and Brownlees & Gallo (2007*b*). Several contributions have addressed the issue of modelling and forecasting realized volatility series using different types of time series models, some of which also include exogenous information. The list of contributions includes Andersen, Bollerslev, Diebold & Labys (2003), Pong, Shackleton, Taylor & Xu (2004), Koopman, Jungbacker & Hol (2005), Andersen, Bollerslev & Diebold (2006). The problem of model selection with explanatory variables for volatility modelling in a MEM framework was addressed in Engle & Gallo (2006) and Gallo & Velucchi (2007). VaR forecasting applications using realized volatility are presented in Andersen et al. (2003), Giot & Laurent (2003), Brownlees & Gallo (2007*a*).

The paper is structured as follows. Section 2 describes the series used for this empirical application. Section 3 presents the set of volatility models with explanatory variables used for prediction. Section 4 illustrates the various selection and combination techniques used to select/combine forecasts. Section 5 explains how VaR forecasts are constructed and evaluated. Concluding remarks follow in Section 6.

2 Returns, Volatility and Volatility Predictors

As customary, the daily return series $\{r_t\}$ is defined as the difference of the logarithmic closing prices (adjusted for dividends and splits) multiplied by 100, that is

$$r_t = 100[\, p_{c,t} - p_{c,t-1} \,].$$

The object of interest throughout is the realized volatility $\{rv_t\}$, defined as the square root of realized variance[2], that is the square root of the sum of intra-daily squared returns

$$rv_t = \sqrt{RV_t} = \sqrt{\sum_{i=0}^{1/\tau} r^2_{t-1+i\tau}},$$

[1]In January 2001 the NYSE moved the minimum price variation allowed on its stocks to USD 0.01. This microstructural change is likely to have caused changes in the characteristics of the high–frequency series. In the forecasting exercise we thus only use the data starting from February 2001.

where τ is the sampling frequency, $r_{t-1+i\tau} = 100[p_{t-1+i\tau} - p_{t-1+(i-1)\tau}]$, assuming that $1/\tau$ is an integer for notational convenience. Some authors, like Andersen et al. (2003), emphasize the use of realized volatility measures for volatility modelling as such measures provide a much more precise estimate of volatility and under much more general assumptions than traditional ARCH–like models. In fact, several authors (including Andersen & Bollerslev (1998), Andersen, Bollerslev, Diebold & Labys (2001), Barndorff-Nielsen & Shephard (2002*a*), Barndorff-Nielsen & Shephard (2002*b*)) have shown, using the theory of quadratic variation, that realized variance is a consistent ex–post estimator of the integrated variance of the underlying continuous time diffusion process that drives the logarithm of prices in the absence of jumps. The sampling frequency τ has been set to 5 minutes as is customary in many studies (e.g. Andersen et al. (2006)).[3]

In modeling the dynamics of realized volatility we consider a set of five variables observable at time $t-1$ which may be of interest, as they capture different aspects of market activity. With formal definitions to follow, we consider:

- *leverage effects*, represented by an indicator representing negative returns multiplied by realized volatility (reminiscent of the asymmetric GARCH);
- *overnight volatility*, defined as the absolute difference between opening price and previous closing price, as the realized volatility is calculated on the intra–daily returns and some impact of the former on the intra-daily volatility has been documented in (e.g. Gallo (2001));
- *trading intensity*, i.e. the average volume per trade, reflecting the link between volume and volatility;
- *daily range* as an alternative measure of volatility;
- *jumps* as a relatively rare event affecting realized volatility (cf. Barndorff-Nielsen & Shephard (2004)).

Leverage Effects Past negative stock return shocks have a greater impact on volatility than positive shocks (Black (1976), Andersen, Bollerslev, Diebold & Ebens (2001)). In order to capture this asymmetry in the behavior of volatility, we consider the series of realized volatility multiplied by a negative return indicator

$$rv_t^- = rv_t \mathbf{1}_{\{r_t<0\}}.$$

Leverage effects are known to be relevant for the improvement of volatility predictions as pointed out by many contributions (e.g. Hansen & Lunde (2005)).

[3]Following, for example, Barndorff-Nielsen & Shephard (2006) we use the term realized variance for the sum of square returns and realized volatility for its square root.

[3]We are aware that the measurement of volatility can be improved in the presence of jumps (Barndorff-Nielsen & Shephard (2004)) or microstructure noise (e.g. Bandi & Russell (2006), Aït-Sahalia, Mykland & Zhang (2005)). By choosing 5-minutes we make our measure less prone to the latter, while we include the former in the information set.

Overnight Volatility Since realized volatility is a measure of intra–daily volatility, it omits the informative content of overnight returns which have peculiar characteristics (many zeros and high density in the tails). As shown in Gallo (2001), a GARCH model for intra-daily conditional variance is affected by an *overnight surprise bias* in that squared overnight returns have significant explanatory power. Thus, we define the overnight volatility series as

$$rv_t^o = 100|p_{o,t} - p_{c,t-1}|.$$

where $p_{o,t}$ is the series of logarithmic adjusted opening prices. The importance of this series can also be justified as a result of the fact that news accumulates during market closure, and is reflected in the stock price at opening time.

Trading Intensity Several authors have established that there is a significant link between volatility and volume (e.g. Tauchen & Pitts (1983)). We transform volume into the log of average volume per trade,

$$ti_t = \log\left(\frac{V_t}{N_t}\right)$$

where V_t is the sum of daily transaction volumes and N_t is the number of daily transactions.

Daily Range Engle & Gallo (2006) show that the high–low range is a measure of volatility which has relevant explanatory power for other measures of volatility. We define the daily range as

$$hl_t = 100[p_{\max,t} - p_{\min,t}]$$

where $p_{\max,t}$ and $p_{\min,t}$ are respectively the logarithm of the daily maximum and minimum price.

Jumps As mentioned, realized variance is not a consistent estimator of the underlying integrated variance in the presence of jumps, in which case one may use bipower variation (e.g. Barndorff-Nielsen & Shephard (2004))

$$BRV_t = \frac{\pi}{2}\sum_{i=0}^{1/\tau} |r_{t+i\tau}||r_{t-1+i\tau}|,$$

As shown by Barndorff-Nielsen & Shephard (2004), the jump component can be estimated as the difference between realized and bipower variation when such difference is positive and zero otherwise. In modelling realized volatility we consider the square root of such quantity

$$j_t = \max(RV_t - BRV_t, 0)^{1/2}.$$

This type of measures of jumps has been considered in some studies such as Andersen et al. (2006) and Gallo & Velucchi (2007).

Figure 1 displays the series of returns and realized volatilities for the 4 stocks analyzed in this paper. Table 1 shows selected values of the sample ACF and PACF of the realized

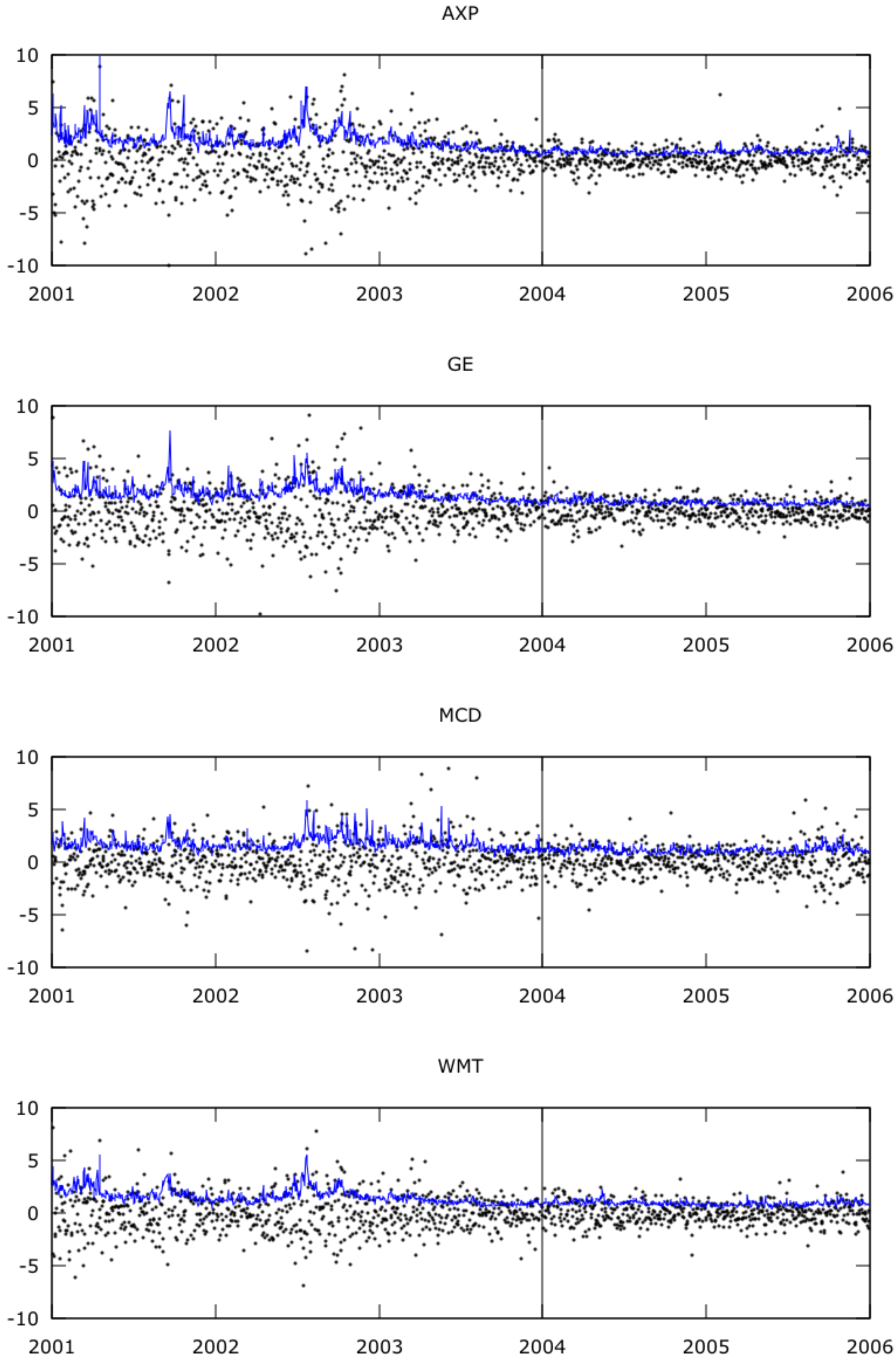

Figure 1: Returns (dots) and realized volatility (continuous line). The vertical line denotes the beginning of the prediction period.

	ACF_1	ACF_5	ACF_{20}	$PACF_5$	$PACF_{20}$
AXP	0.823	0.715	0.584	0.099	0.063
GE	0.779	0.663	0.557	0.062	0.003
MCD	0.608	0.468	0.381	0.111	0.033
WMT	0.801	0.691	0.584	0.007	0.041

Table 1: Realized volatility: Descriptive Statistics.

	rv^-_{t-1}	rv^o_{t-1}	ti_{t-1}	hl_{t-1}	j_{t-1}
AXP	0.08	0.15	0.05	0.05	0.01
GE	0.18	0.01	0.02	0.10	0.00
MCD	0.01	0.10	0.05	0.04	0.05
WMT	0.14	0.02	0.05	0.02	-0.04

Table 2: Partial correlation of $\{rv_t\}$ with the lagged explanatory variables conditional on $\{rv_{t-1}\}$.

volatility series for the 4 stocks. As is well documented in the literature and confirmed by the sample ACF statistic, realized volatility exhibits a very slow decaying memory.

Table 2 gives a rough picture of the links between the realized volatility series and the set of explanatory variables. The table reports the partial correlations of $\{rv_t\}$ with the explanatory variables conditional on $\{rv_{t-1}\}$. Most partial correlations are nonnegative and each stock seems to give different importance to the explanatory variables. The question of actual relevance for forecasting is thus still open.

3 A Multiplicative Error Model for Realized Volatility

The Multiplicative Error Model is suitable to model the realized volatility, also in view of the empirical regularities shown. MEMs are a generalization of the ARCH family of models proposed in Engle (2002). Discussions about the properties of the model can be found in Engle (2002) and Engle & Gallo (2006).

Alternative approaches for the modeling of realized volatility have been proposed in the literature. Several authors adopt ARFIMA–type of models on the logarithm of realized variance. This is motivated by the fact that the logarithm of realized variance exhibits long memory and is approximately Gaussian (e.g. Andersen et al. (2003)). Some authors however have reported that the long memory assumption does not seem to be crucial for forecasting (Pong et al. (2004)). Another strand of literature has proposed mixing information at different frequencies in a linear regression framework (e.g. the so called Heterogeneous AR (HAR) model of Corsi (2003), extended by Andersen et al. (2006) for the inclusion of explanatory variables; and Forsberg & Ghysels (2007) in a MIDAS frameworkGhysels, Santa-Clara & Valkanov (2006)).

Let $\{rv_t\}$ be a realized volatility series and let $\mathcal{F}_{t-1}$ be the information set at $t-1$.

The generic MEM specification we adopt is

$$rv_t = \sigma_t \, \epsilon_t \quad \epsilon_t|\mathcal{F}_{t-1} \sim Gamma(\phi, 1/\phi), \tag{1}$$

where conditionally on $\mathcal{F}_{t-1}$, σ_t is the conditionally deterministic component of the process and ϵ_t is an i.i.d. innovation process term with unit expected value.

The are a number of reasons why we argue that MEMs are a suitable specification for modeling volatility measures. Firstly, the MEM is a nonnegative time series model and hence it always produces nonnegative predictions. Secondly, the Gamma innovation term assumption is a rather flexible distributional assumption that is able to capture the different shapes exhibited by different volatility proxies from absolute returns to realized volatility measures. Lastly, if $\sigma_t = E(rv_t|\mathcal{F}_{t-1})$, the expected value of the score of θ evaluated at the true parameters is zero irrespective of the Gamma assumption on $\epsilon_t|\mathcal{F}_{t-1}$, the ML estimator is also a QML estimator.

The challenge lies in choosing an appropriate specification for the volatility σ_t. The minimal specification for σ_t able to capture the series dynamics is

$$\sigma_{\text{narr},t} = \omega + \alpha rv_{t-1} + \beta \sigma_{\text{narr},t-1}, \tag{2}$$

which we will refer to as the *narrow* model reminiscent of the GARCH(1,1). The narrow model is also the benchmark specification that we will attempt to beat in the empirical application. On the other extreme a rather rich specification for σ_t is

$$\begin{aligned} \sigma_{\text{wide},t} &= \omega + \alpha rv_{t-1} + \beta \sigma_{\text{wide},t-1} \\ &\quad + \gamma_1 rv^-_{t-1} + \gamma_2 rv^o_{t-1} + \gamma_3 ti_{t-1} + \gamma_4 hl_{t-1} + \gamma_4 j_{t-1} \\ &= \omega + \alpha rv_{t-1} + \beta \sigma_{\text{wide},t-1} + \gamma' x_{t-1} \end{aligned} \tag{3}$$

with $\gamma \in \mathbb{R}^5$ and $x_t = (rv^-_{t-1}, rv^o_{t-1}, ti_{t-1}, hl_{t-1}, j_{t-1})'$, which we will refer to as the *wide* model.

Table 3 reports for each stock the maximum likelihood estimation results of the wide model on the full sample together with the p–values of the Ljung–Box test for serial correlation of the standardized residuals $rv_t/\widehat{rv}_{t|t-1}$ and squared standardized residuals $(rv_t/\widehat{rv}_{t|t-1})^2$. Apart from the leverage effect and, sometimes, the daily range, most of the coefficients associated with the explanatory variables do not appear to be statistically significant, The p–values of the Ljung–Box test indicate that the wide specification captures the dynamics of the series adequately for the AXP and GE stock and slightly less convincingly for the MCD and WMT stock. The full sample estimation results suggest that the wide model provides a reasonable description of the volatility dynamics, and that probably a more parsimonious specification might offer some gains in terms of estimation and prediction accuracy.

Table 4 reports the RMSPE and MAPE of the 1–step ahead realized volatility predictions of the narrow and wide model using different rolling window sizes R. The prediction period used for the forecasting exercise ranges from January 2, 2004 to December 30, 2005, corresponding to 503 trading days. The forecasting results show that there is no clear cut winner between the wide and the narrow model and suggest to attempt to search for a compromise between the wide and narrow specifications. Interestingly, smaller rolling window sizes seem to produce better forecasts. This is probably due to the fact that in the prediction period volatility is relatively low while between 2001 and 2004 volatility is much higher, also because of 9/11.

	AXP	GE	MCD	WMT
ω	0.124 (0.023)	0.138 (0.014)	0.081 (0.009)	0.093 (0.028)
α	0.194 (0.033)	0.168 (0.030)	0.184 (0.031)	0.165 (0.028)
β	0.688 (0.020)	0.696 (0.020)	0.769 (0.019)	0.749 (0.021)
γ_1	0.059 (0.009)	0.052 (0.008)	0.014 (0.008)	0.045 (0.009)
γ_2	0.014 (0.009)	0.025 (0.011)	-0.028 (0.011)	-0.016 (0.010)
γ_3	-0.006 (0.031)	-0.006 (0.026)	0.090 (0.029)	0.036 (0.028)
γ_4	0.098 (0.024)	0.136 (0.029)	0.006 (0.032)	0.058 (0.022)
γ_5	0.046 (0.053)	0.031 (0.043)	0.017 (0.044)	0.052 (0.037)
ϕ	16.948 (0.511)	16.987 (0.565)	13.759 (0.439)	19.891 (0.684)
LB_{10}	0.201	0.112	0.046	0.019
LB_{10}^2	0.507	0.220	0.097	0.056

Table 3: Full sample estimation results of the wide model. Standard errors are reported in parenthesis. LB reports the Ljung–Box test p-values on the first 10 lags of the residuals and squared residuals, respectively.

		RMSPE			MAPE		
		R=500	R=600	R=700	R=500	R=600	R=700
AIX	Wide	3.324	3.338	3.373	2.351	2.379	2.408
	Narrow	3.346	3.353	3.369	2.374	2.383	2.404
GE	Wide	3.113	3.120	3.124	2.460	2.455	2.447
	Narrow	3.111	3.117	3.124	2.472	2.477	2.483
MCD	Wide	4.799	4.774	4.789	3.596	3.582	3.624
	Narrow	4.767	4.767	4.774	3.613	3.627	3.666
WMT	Wide	3.384	3.378	3.412	2.527	2.528	2.570
	Narrow	3.429	3.437	3.453	2.571	2.579	2.596

Table 4: Volatility forecasts performance using different rolling window sample sizes.

4 Model Selection Strategies for Prediction

Let S be a subset of $\{1, 2, ..., 5\}$, and denote by v_S the subvector of $v \in \mathbb{R}^5$ of components v_j with $j \in S$. We will denote by $\sigma_{S,t}$ the generic specification obtained by only including those explanatory variables x_i whose index is contained in the set S. Such generic specification can be compactly represented as

$$\sigma_{S,t} \quad = \quad \omega + \alpha rv_{t-1} + \beta\sigma_{S,t-1} + \gamma_S' x_{S,t-1} \quad S \subseteq \{1, ..., 5\}, \tag{4}$$

and it will referred to as submodel S. The number of specifications that can be used for prediction is made up of a collection $\mathcal{M}$ of $32 = 2^5$ submodels ranging from the narrow model to the wide model. As it is assumed that it is not known which subset of explanatory variables produces the best forecasts, some selection or combination mechanism is needed to select or combine predictions.

4.1 Focused Information Criterion

Focused selection criteria attempt at selecting models on the basis of a specific statistical problem of interest. The FIC is derived in a particular asymptotic framework called *local misspecification framework* developed in Hjort & Claeskens (2003). The presentation of the theory in this paper is heuristic and follows the lines of Claeskens et al. (2006). More rigorous treatment of the asymptotic framework can be found in Hjort & Claeskens (2003).

In order to derive the criterion, a *focus parameter* has to be chosen. The focus parameter is a function of the parameters $(\omega, \alpha, \beta, \phi, \gamma_1, ..., \gamma_5)'$. It represents a quantity the researcher assumes to be relevant in the context of the application that is wished to be estimated as precisely as possible. As our objective is to evaluate which explanatory variables are more relevant in explaining volatility, the choice of the focus function adopted in this work is unconditional volatility keeping the values of the explanatory variables fixed at a certain x:

$$\sigma(x) \equiv \frac{\omega + \gamma' x}{1 - \alpha - \beta}.$$

Note that for each submodel S there corresponds a submodel estimator $\hat{\sigma}_S(x)$ for $\sigma(x)$, and the task hence consists of selecting the model that provides the most precise estimate of $\sigma(x)$. The FIC is in fact an estimator of the limiting risk of the focus parameter under the loss function of interest and the FIC model selection strategy consists of selecting the submodel with the smallest estimated risk for the focus parameter.

Some further notation has to be introduced. Let the parameter vector be partitioned in $\theta' \equiv (\omega, \alpha, \beta, \phi)'$ and $\gamma \equiv (\gamma_1, ..., \gamma_5)'$, the latter containing the parameters to be constrained. The local misspecification framework assumption consists of assuming that the true parameter value is $(\theta_0', \delta_0'/\sqrt{n})'$ where n denotes the sample size. This assumption is called local misspecification assumption in that it implies that the null model $(\theta_0', \mathbf{0}')'$ is locally misspecified. Let $B_{0,n}$ denote the information matrix of the wide model computed in $(\theta_0', \mathbf{0}')'$ with the expectation taken with respect to the density in $(\theta_0', \mathbf{0}')'$, that is

$$\begin{aligned} B_{0,n} &\equiv -n^{-1} \sum_{t=1}^{n} \mathrm{E}_0 \begin{pmatrix} \frac{\partial^2 \log f_t(rv_t, x_{t-1}, \theta_0, \mathbf{0})}{\partial\theta\partial\theta'} & \frac{\partial^2 \log f_t(rv_t, x_{t-1}, \theta_0, \mathbf{0})}{\partial\theta\partial\gamma'} \\ \frac{\partial^2 \log f_t(rv_t, x_{t-1}, \theta_0, \mathbf{0})}{\partial\gamma\partial\theta'} & \frac{\partial^2 \log f_t(rv_t, x_{t-1}, \theta_0, \mathbf{0})}{\partial\gamma\partial\gamma'} \end{pmatrix} \\ &= \begin{pmatrix} B_{0,n,11} & B_{0,n,12} \\ B_{0,n,21} & B_{0,n,22} \end{pmatrix} \end{aligned}$$

and let $K_n \equiv (B_{0,n,22} - B_{0,n,21} B_{0,n,11}^{-1} B_{0,n,12})^{-1}$. Then, under regularity assumptions (cf. Hjort & Claeskens (2003)), it can be shown that

$$D_n = \sqrt{n}(\hat{\gamma}_{\text{wide}}) \xrightarrow{d} N(\delta_0, K_n)$$

where $\hat{\gamma}_{\text{wide}}$ is the unrestricted maximum likelihood estimator obtained from the wide model.

After some work, this framework allows one to obtain the asymptotic distribution of the focus parameter for the generic submodel S that is

$$\begin{aligned} \sqrt{n}(\hat{\sigma}_S(z) - \sigma_{\text{true}}(z)) &\xrightarrow{d} \Lambda_S \\ &\overset{a}{\sim} \frac{\partial\sigma}{\partial\theta}' B_{0,n,11}^{-1} M_n + \nu_n' (\delta_0 - K_n^{1/2} H_{n,S} K_n^{-1/2} D_n) \\ &\overset{a}{\sim} N(\ b_{n,S}\ ,\ \tau_{n,S}^2\ ) \end{aligned} \tag{5}$$

where $M_n \sim N(\mathbf{0}, B_{0,n,11})$, $H_{n,S} \equiv K_n^{-1/2} \pi_S' K_{n,S} \pi_S K_n^{-1/2}$, $\pi_S \in \mathbb{R}^{|S| \times 5}$ is the projection matrix mapping v to the subvector v_S, $K_{n,S} \equiv (\pi_S (B_{0,n,22} - B_{0,n,20} B_{0,n,11}^{-1} B_{0,n,12}) \pi_S')^{-1}$, $\nu_n \equiv B_{0,n,21} B_{0,n,11}^{-1} \frac{\partial\sigma}{\partial\theta} - \frac{\partial\sigma}{\partial\gamma}$, $\frac{\partial\sigma}{\partial\theta}$ and $\frac{\partial\sigma}{\partial\gamma}$ are the partial derivatives of $\sigma(\cdot)$ with respect to θ and γ in $(\theta_0', \gamma_0')'$.

The result in Equation 5 states that under local alternatives the distribution of the focus parameter of any submodel estimator S is Gaussian with bias b_S and variance τ_S^2. It is then possible to obtain expression of the risk of the submodel estimator using the loss function of interest for the problem at hand and their estimators, i.e. the FIC.

A number of FICs have been derived using different loss functions. Hjort & Claeskens (2003) initially proposed an estimator of the MSE of the focus parameter, Claeskens et al. (2006) obtain more general results for the generic L_p–error of the estimator as well as other losses. Brownlees & Gallo (2007*b*) derives some asymmetric FICs using the linex and linlin losses. In this paper we restrict the analysis to the MSE and MAE FIC. The Mean Square Error FIC of submodel S is

$$\text{FIC}_{\text{MSE}}(S) = (\hat{b}_{n,S})^2 + 2\nu_n' K_n^{1/2} H_{n,S} K_n^{1/2} \nu_n$$

and the Mean Absolute Error counterpart is

$$\text{FIC}_{\text{MAE}}(S) = 2\hat{b}_{n,S} \left(\Phi\left(\frac{\hat{b}_{n,S}}{\hat{\tau}_{n,S}} \right) - \frac{1}{2} \right) + 2\hat{\tau}_S \phi\left(\frac{\hat{b}_S}{\hat{\tau}_S} \right).$$

where $\hat{b}_{n,m} = \nu_n' (I - K_n^{1/2} H_{n,S} K_n^{-1/2}) D_n$, $\hat{\tau}_{n,s}^2 = \frac{\partial\sigma}{\partial\theta}' B_{0,n,11}^{-1} \frac{\partial\sigma}{\partial\theta} + \nu_n' K_n^{1/2} H_{n,S} K_n^{1/2} \nu_n$ $\Phi(\cdot)$ is the distribution function of a standard normal r.v. and $\phi(\cdot)$ is the probability density function of standard normal r.v..

The focused information criteria just defined depends on the some value of the explanatory variables x. This is an appealing feature of the FIC from a forecasting perspective in that it allows one to select a model conditionally on the values of the explanatory variables that will be used to produce the 1–step ahead forecast for the next period.

4.2 Other Competing Criteria

In order to have more insights on the forecasting ability of the proposed methods, the focused selection strategies will be compared with other well known selection methods.

One of the most famous model selection criteria is the AIC (An Information Criterion) introduced by Akaike in a series of papers in the first half of the 1970s (e.g. Akaike (1973), Akaike (1974))

$$\mathrm{AIC}(S) = -2\log L_S(\hat{\theta}_S, \hat{\gamma}_S) + 2(3 + |S|)$$

where $L_S(\cdot)$ is the likelihood function of submodel S and $\hat{\theta}_S$, $\hat{\gamma}_S$ are the maximum likelihood estimators of its parameters. The criterion is an estimate of the expected Kullback-Leibler distance from the unknown data generating process. Another famous selection criteria is the BIC (Bayesian Information Criterion, also known as the Schwartz Criterion) proposed by Schwartz (1978)

$$\mathrm{BIC}(S) = -2\log L_S(\hat{\theta}_S, \hat{\gamma}_S) + \log(n)(3 + |S|).$$

The criterion is an approximation of the posterior probability of submodel S. The AIC and BIC model selection strategies consist of selecting the submodel with the smallest value of the information criterion. Many other model selection criteria are often asymptotically equivalent to either the AIC or BIC, hence these two criteria are representative.

The AIC and BIC are simple and elegant model selection devices that are commonly employed in statistical analysis even if it is not clear what are their statistical properties, as, for example, Bollerslev, Engle & Nelson (1994) point out in the context of ARCH models. Furthermore, some of the asymptotic properties of selection that have been found in the literature can sometimes be very misleading. As pointed out by Leeb & Pötscher (2005) for instance, consistent model selection strategies (that is, a selection strategy that selects the "true" model asymptotically) like the BIC imply using a post–model selection estimator whose asymptotic maximum mean square error is infinite, a rather unpleasant result. On the other hand, it is unclear how to assess the optimality of conservative model selection strategies (that is, a selection strategy that asymptotically selects an over-parametrised model) like the AIC. For instance, some authors have proved that AIC selects asymptotically some "optimal" model (e.g. Shibata (1984)), but as the convergence of the post–model selection estimator to its limit is not uniform, it is unclear whether these results are informative as to the actual performance of the estimation method.

4.3 Combining Forecasts

As well known in the literature since the seminal contribution of Bates & Granger (1969), if two series of (unbiased) forecasts from different sources are available, choosing one of them and discarding the other is suboptimal. In fact, there exists a combination of forecasts that has smaller MSE than both of them[4]. In recent times some of these themes have become quite popular out of the forecasting literature under the more general heading of model averaging (Hoeting, Madigan, Raftery & Volinsky (1999), Burnham & Anderson (2002)).

Rather than forecasting using the model chosen from some information criteria, some authors such as Buckland, Burnham & Augusting (1997), Hjort & Claeskens (2003) have

[4]See Timmermann (2006) for recent review and theoretical results on combination of forecasts

proposed using a weighted combinations of forecasts, with weights proportional to the (negative) value of a reference information criterion. The weight w_S assigned to the prediction coming from submodel S is

$$w_{\text{axIC},S} = \frac{\exp\left(-\frac{1}{2}\text{xIC}(S)\right)}{\sum_{S \in \mathcal{M}} \exp\left(-\frac{1}{2}\text{xIC}(S)\right)},$$

where xIC is the reference information criterion. This leads to the aFIC$_{\text{MSE}}$, aFIC$_{\text{MAE}}$, aAIC and aBIC forecast combination schemes, where the "a" prefix stands for averaged. However, it is unclear what are the benefits of using such weights for forecasting, though, Stock & Watson (2006) have interestingly pointed out (under specific conditions) that some model averaging estimators from a set of nested models can be represented as a shrinkage estimator.

4.4 Out–of–Sample Realized Volatility Analysis

For each period in the validation sample the forecasting procedure consists of estimating the set of submodels $\mathcal{M}$ and in selecting or combining the submodels' 1–step ahead predictions using some selection/combination strategy. The parameter estimates and selection criteria are computed each period using a rolling window scheme. The set of selection/combination strategies is made up of FIC$_{\text{MSE}}$, FIC$_{\text{MAE}}$, AIC, BIC selection strategies together with their "averaged" counterparts aFIC$_{\text{MSE}}$, aFIC$_{\text{MAE}}$, aAIC and aBIC. The performance of such strategies is compared with the narrow and wide model. The aim of the exercise is to assess if any strategy beats the narrow model benchmark. As the focused criteria attempt at selecting the submodel with the highest estimated precision in estimating volatility, it is of interest to assess if such strategies actually do predict realized volatility better.

Table 5 reports the RMSPE and MAPE of the various strategies, together with the RMSPE and MAPE of the wide and narrow model. Table 5 also reports the γ–parameter average norm reduction and the average persistence of the various selection/combination strategies. Focused methods perform well. In all cases the forecasting method which achieves the best performance is always a FIC based strategy. The FIC$_{\text{MSE}}$ and FIC$_{\text{MAE}}$ almost always beat the narrow and the wide model from both a RMSPE and MAPE point of view. The BIC has the worst performance among the forecasting strategies from both and RMSPE and MAPE perspective in 3 out 4 cases. Note that combining predictions from different models seems to pay off well as the RMSPE and MAPE of the various averaged strategies almost always beat their selection counterparts.

This forecasting exercise also allows one to gain some insights as to which explanatory variables seem to be more relevant in modelling volatility. Table 6 reports the average frequencies of inclusions of the explanatory variables for each strategy. In the case of the combination schemes, the average frequency of inclusion for the explanatory variable x_i is computed as the average sum of weights of the model containing the explanatory variable x_i in each period. Interestingly, across stocks the various selection strategies appear to give different importance to each of the explanatory variables. Table 6 shows that FIC strategies give much more uniform weights to the explanatory variables in comparison to the AIC and BIC strategies and also tend to select larger models. The weights given to the explanatory variables by the AIC and BIC are much more diversified. Leverage effects

AXP

Strategy	RMSPE	MAPE	Norm. Red.	Pers.
AIC	3.335	2.374	13.8%	0.88
BIC	3.337	2.373	32.0%	0.89
FIC_s	3.340	2.366	18.9%	0.90
FIC_a	3.327	2.357	30.2%	0.91
aAIC	3.331	2.371	14.3%	0.88
aBIC	3.323	2.363	30.2%	0.89
aFIC_s	3.330	2.369	26.8%	0.90
aFIC_a	3.321	2.363	31.4%	0.91
Wide	3.338	2.379		0.88
Narrow	3.353	2.383		0.97

GE

Strategy	RMSPE	MAPE	Norm. Red.	Pers.
AIC	3.124	2.459	11.0%	0.84
BIC	3.128	2.475	23.6%	0.87
FIC_s	3.115	2.451	22.4%	0.86
FIC_a	3.110	2.448	35.3%	0.88
aAIC	3.119	2.459	10.9%	0.84
aBIC	3.121	2.465	23.4%	0.87
aFIC_s	3.104	2.445	28.0%	0.87
aFIC_a	3.102	2.450	31.8%	0.87
Wide	3.120	2.455		0.82
Narrow	3.117	2.477		0.97

MCD

Strategy	RMSPE	MAPE	Norm. Red.	Pers.
AIC	4.791	3.613	21.3%	0.94
BIC	4.815	3.685	68.9%	0.97
FIC_s	4.767	3.580	9.3%	0.93
FIC_a	4.774	3.580	12.5%	0.93
aAIC	4.781	3.607	26.4%	0.95
aBIC	4.789	3.647	64.0%	0.97
aFIC_s	4.756	3.568	18.6%	0.94
aFIC_a	4.757	3.582	36.2%	0.94
Wide	4.774	3.582		0.93
Narrow	4.767	3.627		0.98

WMT

Strategy	RMSPE	MAPE	Norm. Red.	Pers.
AIC	3.399	2.546	11.6%	0.87
BIC	3.403	2.543	41.4%	0.90
FIC_s	3.381	2.525	16.8%	0.88
FIC_a	3.397	2.531	23.4%	0.89
aAIC	3.388	2.534	14.5%	0.88
aBIC	3.403	2.547	42.2%	0.90
aFIC_s	3.397	2.538	22.9%	0.89
aFIC_a	3.392	2.538	27.7%	0.89
Wide	3.378	2.528		0.86
Narrow	3.437	2.579		0.95

Table 5: RMSPE, MAPE, average γ–parameter norm reduction, average persistence.

AXP

Strategy	rv^-	rv^o	ti	hl	j
		Frequency of Inclusion			
AIC	100.0%	22.4%	35.4%	100.0%	6.6%
BIC	85.8%	0.4%	1.2%	81.0%	2.2%
FIC$_s$	79.0%	81.6%	80.0%	82.4%	82.0%
FIC$_a$	72.4%	64.6%	59.2%	70.0%	63.6%
		Average Weights			
aAIC	95.8%	42.4%	45.4%	95.9%	33.7%
aBIC	83.1%	9.1%	11.4%	78.2%	7.2%
aFIC$_s$	75.6%	62.8%	60.2%	73.5%	63.8%
aFIC$_a$	69.2%	51.9%	51.4%	68.2%	55.3%

GE

Strategy	rv^-	rv^o	ti	hl	j
		Frequency of Inclusion			
AIC	92.2%	98.6%	0.0%	100.0%	27.8%
BIC	40.6%	56.0%	0.0%	100.0%	4.0%
FIC$_s$	60.6%	70.8%	66.2%	75.2%	75.8%
FIC$_a$	50.6%	55.8%	41.0%	60.8%	58.0%
		Average Weights			
aAIC	81.8%	88.8%	29.1%	99.6%	44.8%
aBIC	48.2%	54.9%	4.5%	97.1%	11.8%
aFIC$_s$	60.9%	60.7%	52.3%	71.6%	61.9%
aFIC$_a$	56.8%	55.7%	49.6%	67.7%	58.2%

MCD

Strategy	rv^-	rv^o	ti	hl	j
		Frequency of Inclusion			
AIC	78.8%	73.6%	74.2%	46.6%	6.0%
BIC	2.0%	44.0%	17.6%	10.6%	0.2%
FIC$_s$	90.2%	91.8%	91.8%	91.8%	91.4%
FIC$_a$	87.2%	89.2%	91.0%	90.2%	90.6%
		Average Weights			
aAIC	66.2%	73.0%	68.0%	54.4%	32.2%
aBIC	20.8%	45.1%	24.9%	20.7%	5.4%
aFIC$_s$	80.6%	81.9%	83.9%	83.4%	83.4%
aFIC$_a$	63.6%	60.2%	61.4%	63.6%	60.8%

WMT

Strategy	rv^-	rv^o	ti	hl	j
		Frequency of Inclusion			
AIC	90.6%	18.8%	8.0%	97.8%	35.4%
BIC	52.8%	0.0%	0.0%	52.4%	12.8%
FIC$_s$	71.0%	74.4%	74.8%	79.0%	83.6%
FIC$_a$	57.4%	59.2%	59.6%	69.2%	72.6%
		Average Weights			
aAIC	81.8%	40.6%	37.3%	83.2%	51.8%
aBIC	59.0%	7.2%	7.8%	53.9%	18.2%
aFIC$_s$	62.5%	58.7%	59.3%	68.4%	68.4%
aFIC$_a$	60.1%	50.5%	51.3%	61.7%	59.6%

Table 6: Frequencies of inclusion and average weights of the explanatory variables.

and the daily range are often judged as relevant by such criteria for AXP, GE and WMT stocks. Overnight volatility appears quite frequently in the GE and MCD stock, while not so often for the AXP and WMT stocks. Trading intensity appears frequently in the MCD stock only. The jump component appears to be the least favorite explanatory variable for all stocks.

5 VaR Forecasting

5.1 Construction the VaR Forecasts

Volatility forecasts are one of the most important ingredients for Value–at–Risk (VaR) prediction. VaR became very popular in the 1990s as a regulatory tool, to help keeping capital exposure under control. The $(1-\alpha)100\%$ VaR is defined as the maximum loss that is expected with a $(1-\alpha)100\%$ confidence over a prespecified time horizon: here we will consider 99% and 95% VaRs over a 1 day time horizon.

An important stylized fact pointed out by a number of contributions (Andersen, Bollerslev, Diebold & Labys (2001)) and confirmed by the data at hand is that daily returns standardized by contemporary realized volatility $z_t = r_t/rv_t$ are approximately distributed as standard normal. Table 7 reports some descriptive statistics of the standardized returns series r_t/rv_t together with a Jarque–Bera test p–value. The standardized series look reasonably close to a Gaussian distribution and exhibit weak persistence in the square levels, as reported in the findings of Andersen, Bollerslev, Diebold & Ebens (2001) on stock data. Under the normality assumption of standardized returns and correct specification of the volatility model, the distribution of returns standardized by predicted realized volatility *conditional* on the information available at $t-1$ is

$$\frac{r_t}{rv_{t|t-1}} = \frac{r_t}{rv_t}\,\frac{rv_t}{rv_{t|t-1}} = z_t\;\epsilon_t,$$

which is the product of an independent standard normal r.v. with a gamma r.v., and is clearly not normal. However, as in Andersen et al. (2003), for simplicity, it is assumed that the returns standardized by realized volatility conditional on $t-1$ are normal $r_t/\kappa rv_{t|t-1} \sim N(0,1)$ where κ is a scale factor, which has to be estimated, ensuring that the standardized returns have a unit variance (see also Giot & Laurent (2003)).

The 1 day horizon conditional VaR at period $t-1$ at a $(1-\alpha)100\%$ level is thus given by

$$\mathrm{VaR}^{\alpha}_{t|t-1} = -z_\alpha\,\kappa\,rv_{t|t-1},$$

where z_α is the α quantile of the normal distribution.[5]

As a consequence, the VaR forecasts constructed using model selection strategy have the form

$$\widehat{\mathrm{Var}}^{\alpha}_{\mathrm{xIC},t|t-1} = z_\alpha\,\hat{\kappa}_{S'}\,\widehat{rv}_{S',t|t-1} \qquad S' = \arg\min_{S\in\mathcal{M}} \mathrm{xIC}(S)$$

[5] It is not totally clear whether the normality assumption is satisfactory. This assumption seems to work well for Andersen et al. (2003). On the other hand, Giot & Laurent (2003) report that that standardized skewed t distribution performs much better.

	Var.	Skew.	Kurt.	$\#\{z_t < z_{0.01}\}$	$\#\{z_t < z_{0.05}\}$	$\text{ACF}_1(z_t^2)$	JB
AXP	1.3	0.15	2.96	0.80	3.60	-0.062	0.083
GE	1.4	0.24	3.13	0.56	3.11	-0.070	0.001
MCD	1.2	0.14	3.15	0.45	3.99	0.021	0.069
WMT	1.3	0.14	3.33	0.96	4.47	-0.022	0.008

Table 7: Standardized Returns $z_t = r_t/rv_t$: Descriptive Statistics.

while the averaged/combined forecasts have the form

$$\widehat{\text{VaR}}^{\alpha}_{\text{axIC},t|t-1} = \sum_{S\in\mathcal{M}} w_{\text{axIC},S}\; z_\alpha\, \hat{\kappa}_S\, \widehat{rv}_{S,t|t-1},$$

where $\widehat{rv}_{S,t|t-1}$ denotes the one step ahead forecast of rv_t in model S given the information available until time $t-1$ and $\hat{\kappa}_S$ is the estimated scale factor for submodel S.

5.2 Forecast Evaluation

Despite the recent significant improvements in the evaluation of volatility predictions (e.g. Andersen & Bollerslev (1998), Andersen, Bollerslev & Meddahi (2005)), a volatility evaluation metric might fail to asses the usefulness of volatility forecasts form a risk-management point of view, as noted among others by Brooks & Persand (2003). The literature on VaR forecasting has developed evaluation tools which explicitly undertake a risk-management viewpoint.

The assessment of the performance of the VaR forecast is carried out according to a two step procedure. The first step assesses the *adequacy* of the VaR forecasting methods using a battery of tests and the second step assesses the *accuracy* of those forecasting methods found significantly adequate using a risk management loss function. Such an approach follows the lines of methodology proposed by Sarma, Thomas & Shah (2003), which is built, among others, on the contributions of Christoffersen (1998) and Lopez (1999). The following discussion follows Sarma et al. (2003).

Adequate VaR forecasts should have a *"correct conditional coverage"*, that is, the failure rate of the forecasts should be close to the nominal level α conditionally at every point in time. Conditional coverage is a particular important property of VaR forecasts in that as volatility tends to cluster, the clustering of VaR failures could be particularly harmful from a risk management perspective. Let $\{\widehat{\text{VaR}}^{\alpha}_{t|t-1}\}$ be a sequence of 1–step ahead $(1-\alpha)100\%$ VaR forecasts and define the failure process $\{I_t\}$ as

$$I_t = \begin{cases} 1 & \text{if } r_t < \widehat{\text{VaR}}^{\alpha}_{t|t-1} \\ 0 & \text{else} \end{cases};$$

Christoffersen (1998) develops three LR tests to assess the correct conditional coverage properties based the failure process $\{I_t\}$.

Unconditional coverage test. Assuming that $\{I_t\}$ is an independently distributed failure process, the null hypothesis of the unconditional coverage test is that the failure probability is equal to α, and it is tested against the alternative of a failure rate different from

α. The test statistic is

$$LR_{\text{uc}} = -2\log\frac{\alpha^{n_1}(1-\alpha)^{n_0}}{\hat{\pi}^{n_1}(1-\hat{\pi})^{n_0}} \sim \chi^2_{(1)},$$

where n_0 and n_1 are, respectively, the number of 0's and 1's in the series and $\hat{\pi} = n_1/(n_0 + n_1)$.

Independence test. The null hypothesis of the independence test is that the failure process $\{I_t\}$ is independently distributed, and it is tested against the alternative of a first order Markov process. The test statistic is

$$LR_{\text{ind}} = -2\log\frac{(1-\hat{\pi}_2)^{(n_{00}+n_{10})}\hat{\pi}_2^{(n_{01}+n_{11})}}{(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}} \sim \chi^2_{(1)},$$

where n_{ij} is the number of i values followed by a j in the I_t series, $\hat{\pi}_{01} = n_{01}/(n_{00}+n_{01})$, $\hat{\pi}_{11} = n_{11}/(n_{10}+n_{11})$ and $\hat{\pi}_2 = (n_{01}+n_{11})/(n_{00}+n_{01}+n_{10}+n_{11})$.

Conditional coverage test. The null hypothesis of the conditional coverage is that the failure process $\{I_t\}$ is an independent failure process with failure probability α, and it is tested against the alternative of a first–order Markov failure process with a different transition probability matrix. The test statistic is

$$LR_{\text{cc}} = -2\log\frac{\alpha^{n_1}(1-\alpha)^{n_0}}{(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}} \sim \chi^2_{(2)}.$$

Note that conditionally on the first observation $LR_{\text{cc}} = LR_{\text{uc}} + LR_{\text{ind}}$.

The literature has also proposed some more refined tests for the assessment of the adequacy of VaR forecast, like Christoffersen & Diebold (2000). However these adequacy tests suffice for the purposes of the current analysis.

Statistical adequacy is a necessary requirement that VaR forecasts must satisfy, but it does not provide information as to the accuracy of such predictions and it does not always help to discriminate among different VaR forecasting methods. Lopez (1999) hence proposed the use of loss functions that reflect how accurate the VaR forecasts are from a risk-management perspective, and suggested using asymmetric loss functions of the form

$$L_t(r_t, \widehat{\text{VaR}}^{\alpha}_{t|t-1}) = \begin{cases} f(r_t, \widehat{\text{VaR}}^{\alpha}_{t|t-1}) & \text{if } r_t < \widehat{\text{VaR}}^{\alpha}_{t|t-1} \\ g(r_t, \widehat{\text{VaR}}^{\alpha}_{t|t-1}) & \text{else} \end{cases};$$

where $f(\cdot)$ is meant to penalize more severely than $g(\cdot)$. Sarma et al. (2003) proposed using a loss function named "Firm's Loss Function",

$$FLF_t(r_t, \widehat{\text{VaR}}^{\alpha}_{t|t-1}) = \begin{cases} (r_t - \widehat{\text{VaR}}^{\alpha}_{t|t-1})^2 & \text{if } r_t < \widehat{\text{VaR}}^{\alpha}_{t|t-1} \\ i\widehat{\text{VaR}}^{\alpha}_{t|t-1} & \text{else} \end{cases};$$

which penalizes quadratically VaR violations, while in case the VaR is not violated the loss is equal to the cost of capital (i is the opportunity cost of capital). In order to assess the superior predictive ability between two competing forecasting methods, it is suggested one should resort to a predictive ability test (e.g. Diebold & Mariano (1995)).

AXP

	99% VaR				95% VaR				
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	JB
$\widehat{rv}_{t\|t-1}$	1.0	0.996	0.751	0.951	4.8	0.828	0.879	0.966	0.000
rv_t	1.0	0.996	0.751	0.951	3.8	0.196	0.747	0.411	0.823

GE

	99% VaR				95% VaR				
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	JB
$\widehat{rv}_{t\|t-1}$	0.6	0.329	0.849	0.610	3.4	0.081	0.600	0.189	0.000
rv_t	0.6	0.329	0.849	0.610	5.0	0.992	0.155	0.363	0.056

MCD

	99% VaR				95% VaR				
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	JB
$\widehat{rv}_{t\|t-1}$	1.2	0.666	0.703	0.847	4.0	0.284	0.233	0.277	0.000
rv_t	0.6	0.329	0.849	0.610	5.0	0.992	0.105	0.268	0.258

WMT

	99% VaR				95% VaR				
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	JB
$\widehat{rv}_{t\|t-1}$	0.4	0.124	0.899	0.304	4.0	0.284	0.233	0.277	0.000
rv_t	0.4	0.124	0.899	0.304	3.2	0.048	0.530	0.115	0.281

Table 8: In–sample 99% and 95% VaR prediction results. The table reports the failure rates, the p–values of the adequacy tests and the p–value of the Jarque–Bera test of the standardised series.

5.3 In–Sample VaR Prediction Analysis

Table 8 reports the results of the adequacy tests for the VaR forecasts in the prediction period using realized volatility forecasts based on the wide model estimates on the full sample as well as realized volatility itself. This information is of interest as a reference for the following out–of–sample recursive prediction exercise. Table 8 reports the percentage of VaR violations, the p–values of the VaR adequacy tests previously defined and the Jarque–Bera test under the null of normality of the standardized returns. Generally speaking, the wide model provides adequate ex-post VaR forecasts. Realized volatility forecasts provide a satisfactory VaR coverage in most of the cases despite the fact that the normality assumption of the standardized returns seems to be hard to justify on the grounds of the Jarque–Bera test. Realized volatility of course provides very good VaR forecasts and, interestingly, in the prediction period there is much less evidence against the hypothesis of normality in comparison to the full sample judging from the p–values of the Jarque–Bera test. The number of VaR violations is almost below the nominal level in almost all the cases. This can be explained by the fact that the forecasting period is a period of relatively low volatility while parameter estimates are obtained on a sample which also contains extreme volatility episodes.

5.4 Out–of–Sample VaR Prediction Analysis

Table 9 reports for each stock the number of VaR violations of the various forecasting strategies together with the battery of tests for VaR adequacy. The overall adequacy of the VaR forecast does not seem to change dramatically across the various forecasting methods and the narrow model. Moreover, the results of the adequacy tests are coherent with the in–sample forecasts of the wide model. However, it is interesting to note that the inclusion of the explanatory variables does not increase the number of VaR violations in almost all the cases. Therefore it can be claimed that explanatory variables can capture some aspect of the trading activity that helps to improve VaR forecast qualitatively. As said for the in–sample VaR forecasts, the number of VaR violations is smaller than the nominal failure rate in almost all cases, which might be due to the fact the forecast period is a relatively low volatility period.

Table 10 reports the FLF losses associated with each strategy together with the p–value of Diebold–Mariano Predictive Ability test under the null of equal predictive ability with the benchmark. As the validation sample has a moderate size and the FLF heavily penalizes VaR violations, forecasting methods with a higher number of VaR violation are always beaten by those methods with a lower number of VaR violations. This makes the comparisons of different methods much harder. Luckily enough, the number of VaR violations is sufficiently uniform and in 3 cases out of 8 it is equal for all strategies, making comparisons still possible. Despite these complications, results show that the FIC model selection criteria always have a better performance than the benchmark whenever the number of failures of the FIC is lower than *or* equal to the benchmark. In one case the number of failures of the FIC is greater than in the benchmark (MCD at 95%). Furthermore, the Diebold Mariano predictive ability tests show little evidence in favor of the hypothesis of equal predictive ability with the benchmark. Another interesting result is that the FIC selection methods always achieve the lowest loss among all other methods with an equal number of violations. The AIC strategies produces VaR forecast with significantly better performance than the benchmark in 3 out of 8 cases. The BIC based strategies are the ones with the less convincing performance. Despite the case of the AXP stock, BIC strategies always produce VaR forecasts that do not seem to have a significantly different performance in comparison with the benchmark. Note that combination schemes do not seem to uniformly improve the quality of the basic selection strategy predictions, as in the case for predicting realized volatility.

6 Concluding Remarks

In this paper we engaged in a forecasting competition between the predictions of a benchmark parsimonious volatility model and the predictions constructed using model selection and combination strategies using also explanatory variables. The success in improving upon the benchmark model is ultimately assessed on the grounds of a VaR forecasting application.

The focus model selection and combination strategies proposed in this paper are interesting in that they both have a clearer interpretation in the context of the forecasting exercise of interest and also because they are able to produce a significantly better forecast than the benchmark in most cases. This result is also of interest in that it implies

AXP

	99% VaR				95% VaR			
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}
AIC	0.8	0.641	0.799	0.868	3.2	0.047	0.531	0.118
BIC	0.8	0.641	0.799	0.868	3.6	0.131	0.674	0.293
FIC_s	0.8	0.641	0.799	0.868	3.6	0.131	0.674	0.293
FIC_a	0.8	0.641	0.799	0.868	3.4	0.082	0.601	0.193
aAIC	0.8	0.641	0.799	0.868	3.4	0.082	0.601	0.193
aBIC	0.8	0.641	0.799	0.868	3.6	0.131	0.674	0.293
$aFIC_s$	0.8	0.641	0.799	0.868	3.4	0.082	0.602	0.193
$aFIC_a$	0.8	0.641	0.799	0.868	3.4	0.082	0.602	0.193
Wide	0.8	0.641	0.799	0.868	3.2	0.049	0.531	0.118
Narrow	1.0	0.999	0.750	0.950	3.8	0.199	0.748	0.417

GE

	99% VaR				95% VaR			
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}
AIC	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
BIC	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
FIC_s	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
FIC_a	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
aAIC	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
aBIC	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
$aFIC_s$	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
$aFIC_a$	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
Wide	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066
Narrow	0.6	0.332	0.849	0.613	3.0	0.027	0.463	0.066

MCD

	99% VaR				95% VaR			
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}
AIC	1.2	0.663	0.702	0.845	4.0	0.289	0.2340	0.28
BIC	1.2	0.663	0.702	0.845	3.6	0.131	0.1554	0.17
FIC_s	1.2	0.663	0.702	0.845	4.0	0.289	0.2340	0.28
FIC_a	1.2	0.663	0.702	0.845	4.0	0.289	0.2340	0.28
aAIC	1.2	0.663	0.702	0.845	3.8	0.199	0.1922	0.19
aBIC	1.2	0.663	0.702	0.845	3.6	0.131	0.1554	0.12
$aFIC_s$	1.2	0.663	0.702	0.845	4.0	0.289	0.2340	0.28
$aFIC_a$	1.2	0.663	0.702	0.845	4.0	0.289	0.2340	0.28
Wide	1.2	0.663	0.702	0.845	4.0	0.289	0.2340	0.28
Narrow	1.2	0.663	0.702	0.845	3.8	0.199	0.1922	0.19

WMT

	99% VaR				95% VaR			
	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}	Fail. Rate	LR_{uc}	LR_{in}	LR_{cc}
AIC	0.4	0.125	0.899	0.306	3.6	0.131	0.155	0.117
BIC	0.4	0.125	0.899	0.306	3.6	0.131	0.155	0.117
FIC_s	0.4	0.125	0.899	0.306	3.6	0.131	0.155	0.117
FIC_a	0.4	0.125	0.899	0.306	3.8	0.199	0.192	0.188
aAIC	0.4	0.125	0.899	0.306	4.0	0.289	0.234	0.280
aBIC	0.4	0.125	0.899	0.306	3.6	0.131	0.155	0.117
$aFIC_s$	0.4	0.125	0.899	0.306	4.0	0.289	0.234	0.280
$aFIC_a$	0.4	0.125	0.899	0.306	4.0	0.289	0.234	0.280
Wide	0.4	0.125	0.899	0.306	3.8	0.199	0.192	0.188
Narrow	0.6	0.331	0.849	0.613	4.0	0.289	0.234	0.280

Table 9: VaR adequacy tests.

AXP

Strategy	99% VaR FLF	99% VaR PA	95% VaR FLF	95% VaR PA
AIC	0.2488	0.000	0.4510	0.001
BIC	0.2472	0.000	0.4791	0.000
FIC_s	0.2468	0.001	0.4872	0.001
FIC_a	0.2467	0.001	0.4649	0.000
aAIC	0.2486	0.000	0.4685	0.000
aBIC	0.2480	0.000	0.4801	0.000
$aFIC_s$	0.2485	0.002	0.4674	0.003
$aFIC_a$	0.2485	0.000	0.4669	0.000
Wide	0.2498		0.4543	
Narrow	0.2949		0.5106	

GE

Strategy	99% VaR FLF	99% VaR PA	95% VaR FLF	95% VaR PA
AIC	0.2042	0.964	0.4682	0.964
BIC	0.2032	0.893	0.4636	0.893
FIC_s	0.2007	0.002	0.4609	0.003
FIC_a	0.2003	0.000	0.4613	0.000
aAIC	0.2046	0.420	0.4676	0.420
aBIC	0.2056	0.396	0.4654	0.396
$aFIC_s$	0.2037	0.001	0.4649	0.014
$aFIC_a$	0.2061	0.004	0.4665	0.036
Wide	0.2048		0.4678	
Narrow	0.2106		0.4684	

MCD

Strategy	99% VaR FLF	99% VaR PA	95% VaR FLF	95% VaR PA
AIC	0.5601	0.000	1.0224	0.000
BIC	0.5660	0.218	0.9634	0.263
FIC_s	0.5561	0.000	1.0156	0.000
FIC_a	0.5561	0.000	1.0149	0.000
aAIC	0.5590	0.002	0.9937	0.002
aBIC	0.5638	0.687	0.9643	0.754
$aFIC_s$	0.5563	0.000	1.0165	0.000
$aFIC_a$	0.5602	0.000	1.0193	0.000
Wide	0.5561		1.0149	
Narrow	0.5646		0.9984	

WMT

Strategy	99% VaR FLF	99% VaR PA	95% VaR FLF	95% VaR PA
AIC	0.1895	0.263	0.5861	0.304
BIC	0.1900	0.496	0.5904	0.555
FIC_s	0.1880	0.017	0.5856	0.011
FIC_a	0.1879	0.041	0.6091	0.040
aAIC	0.1890	0.720	0.6360	0.654
aBIC	0.1887	0.304	0.5876	0.347
$aFIC_s$	0.1877	0.017	0.6355	0.210
$aFIC_a$	0.1877	0.002	0.6349	0.035
Wide	0.1885		0.6118	
Narrow	0.2273		0.6533	

Table 10: VaR accuracy comparison.

that it is possible to significantly improve volatility forecasts using a wider information set and appropriate devices for the selection of the inclusion of the relevant information. The explanatory variables that appear to be more important for prediction are leverage effects, overnight volatility and the daily range. However, different series seem to react differently to the information contained in the explanatory variables. Classic information criteria such as the AIC and the BIC do not appear to be able to select better forecasts.

This paper also uses combinations/model averaging strategies for volatility predictions. Such schemes seem to produce better volatility forecasts but their performance in the VaR exercise is not outstanding. An important implication of this latter result is that different strategies might perform better in different types of forecasting problems. Finally, the paper embraces the view that it is interesting to assess which are useful forecasting methods rather than searching for the best forecasting model and it has shown how such an approach can successfully be applied in empirical applications.

References

Aït-Sahalia, Y., Mykland, P. & Zhang, L. (2005), 'How often to sample a continuous–time process in the presence of market microstructure noise', *The Review of Financial Studies* **28**, 351–416.

Akaike, H. (1973), *Information theory as an extension of the maximum likelihood principle*, Second International Symposium on Information Theory, Akademiai Kiado, Budapest, pp. 267–281.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**, 716–723.

Andersen, T. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T., Bollerslev, T. & Diebold, F. (2006), 'Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility', *Review of Economics and Statistics* . forthcoming.

Andersen, T., Bollerslev, T., Diebold, F. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–45.

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 529–626.

Andersen, T., Bollerslev, T. & Meddahi, N. (2005), 'Correcting the errors: Volatility forecast evaluation using high–frequency data and realized volatilities', *Econometrica* **73**, 279–296.

Bandi, F. M. & Russell, J. R. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics* **79**, 655–692.

Barndorff-Nielsen, O. & Shephard, N. (2002*a*), ‘Econometric analysis of realized volatility and its use in estimating stochastic volatility models’, *Journal of the Royal Statistical Society* **63**, 253–280.

Barndorff-Nielsen, O. & Shephard, N. (2002*b*), ‘Estimating quadratic variation using realized variance’, *Journal of Applied Econometrics* **17**, 457–458.

Barndorff-Nielsen, O. & Shephard, N. (2004), ‘Power and bipower variation with stochastic volatility and jumps (with discussion)’, *Journal of Financial Econometrics* **2**, 1–48.

Barndorff-Nielsen, O. & Shephard, N. (2006), ‘Econometrics of testing for jumps in financial economics using bipower variation’, *Journal of Financial Econometrics* **4**, 1–30.

Bates, J. M. & Granger, C. W. J. (1969), ‘The combination of forecasts’, *Operations Research* **20**, 451–468.

Black, F. (1976), Studies of stock market volatility changes, *in* ‘Proceedings of the American Statistics Association, Business and Economic Statitics Section’, pp. 177–181.

Bollerslev, T. (1986), ‘Generalized autoregressive conditional autoconditional heteroscedasticity’, *Journal of Econometrics* **31**, 304–327.

Bollerslev, T., Engle, R. & Nelson, D. (1994), *ARCH Models, Handbook of Econometrics*, Vol. IV, Elsevier Science B.V, chapter 49, pp. 2959–3038.

Brooks, C. & Persand, G. (2003), ‘Volatility forecasting for risk management’, *Journal of Forecasting* **22**, 1–22.

Brownlees, C. & Gallo, G. (2007*a*), Comparison of volatility measures: a risk management perspective, Technical report.

Brownlees, C. & Gallo, G. (2007*b*), Flexible time series forecasting using shrinkage techniques and focused selection criteria, Technical report, Dipartimento di Statistica ‘G.Parenti’, Università di Firenze. WP 2007-2.

Brownlees, C. & Gallo, G. M. (2006), ‘Financial econometric analysis of ultra–high frequency: data handling concerns’, *Computational Statistics and Data Analysis* **51**, 2232–2245.

Buckland, S. T., Burnham, K. P. & Augusting, N. (1997), ‘Model selection: An integral part of inference’, *Biometrics* **53**, 603–618.

Burnham, K. P. & Anderson, D. R. (2002), *Model Selection and Multimodel Inference*, Springer, New York.

Christoffersen, P. (1998), ‘Evaluation of interval forecasts’, *International Economic Review* **39**, 841–862.

Christoffersen, P. F. & Diebold, F. X. (2000), ‘How relevant is volatility forecasting for financial risk management’, *The Review of Economics and Statistics* **82**, 12–22.

Claeskens, G., Croux, C. & Van Kerckhoven, J. (2006), 'Variable selection for logistic regression using a prediction focussed information criterion', *Biometrics* **62**, 972–979.

Claeskens, G. & Hjort, N. L. (2003), 'The focused information criterion', *Journal of the American Statistician Association* **98**, 900–916.

Corsi, F. (2003), A simple long memory model of realized volatility, Technical report, Manuscript. University of Southern Switzerland.

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

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the the variance of united kingdom inflation', *Econometrica* **50**, 987–1008.

Engle, R. F. (2002), 'New frontiers for arch models', *Journal of Applied Econometrics* **17**, 425–446.

Engle, R. & Gallo, G. (2006), 'A multiple indicator model for volatility using intra daily data', *Journal of Econometrics* **131**, 3–27.

Forsberg, L. & Ghysels, E. (2007), 'Why do absolute returns predict volatility so well?', *Journal of Financial Econometrics* **5**, 31–67.

Gallo, G. (2001), 'Modelling the impact of overnight surprises on intra-daily volatility', *Australian Economic Papers* **40**, 567–580.

Gallo, G. & Velucchi, M. (2007), Ultra–high frequency measures of volatility: an mem–based investigation, Technical report, Dipartimento di Statistica 'G.Parenti', Università di Firenze. WP 2007-01.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2006), 'Predicting volatility : getting the most out of return data sampled at different frequencies', *Journal of Econometrics* **131**, 59–95.

Giot, P. & Laurent, S. (2003), 'Modelling daily value–at–risk using realized volatility and arch type models', *Journal of Empirical Finance* **11**, 379–398.

Hansen, P. & Lunde, A. (2005), 'A forecast comparison of volatility models: Does anything beat a garch(1,1)?', *Journal of Applied Econometrics* **20**, 873–889.

Hjort, N. L. & Claeskens, G. (2003), 'Frequentist model average estimators', *Journal of the American Statistician Association* **98**, 879–899.

Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. (1999), 'Bayesian model averaging: A tutorial', *Statistical Science* **14**, 382–417.

Koopman, S., Jungbacker, B. & Hol, E. (2005), 'Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* **12**, 445–475.

Lamoureux, C. G. & Lastrapes, W. D. (1990), 'Heteroskedasticity in stock return data: Volume versus garch effects', *The Journal of Finance* **45**, 221–229.

Lamoureux, C. G. & Lastrapes, W. D. (1994), 'Endogenous trading volume and momentum in stock-return volatility', *Journal of Business and Economic Statistics* **12**, 253–260.

Leeb, H. & Pötscher, B. M. (2005), 'Model selection and inference: Facts and fiction', *Econometric Theory* **21**, 21–59.

Lopez, J. (1999), 'Methods for evaluating value–at–risk estimates', *Federal Reserve of San Francisco Economic Review* **2**, 3–17.

Pong, S., Shackleton, M., Taylor, S. & Xu, X. (2004), 'Forecasting currency volatility: A comparison of implied volatilities and ar(fi)ma models', *Journal of Banking & Finance* **28**, 2541–2563.

Poon, S. & Granger, C. W. J. (2005), 'Practical issues in forecasting volatility', *Financial Analysis Journal* **61**, 45–56.

Sarma, M., Thomas, S. & Shah, A. (2003), 'Selection of value at risk models', *Journal of Forecasting* **22**, 337–358.

Schwartz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**, 461–464.

Shibata, R. (1984), 'Approximate efficiency of a selection procedure for the number of regression variables', *Biometrika* **71**, 43–49.

Stock, J. H. & Watson, M. W. (2006), An empirical comparison of methods for forecasting using many predictors. Manuscript.

Tauchen, G. & Pitts, M. (1983), 'The price variability-volume relationship on speculative markets', *Econometrica* **51**, 485–506.

Timmermann, A. (2006), *Forecast Combinations, Handbook of Economic Forecasting*, Elsevier Science B.V.

Copyright © 2007
Christian T. Brownlees,
Giampiero M. Gallo