Volatility Forecast Comparison using Imperfect Volatility Proxies

Andrew Patton

London School of Economics

July 2006

1 Motivation

- The efforts devoted to econometric modelling and forecasting generates strong demand for forecast comparison methods

- The study of forecast evaluation and comparison methods has a long history, back to at least Cowles (1933). (See West (2005) for a recent survey.) But most existing methods rely on the target variable being *observable*.

- Many economic forecasting problems involve *unobservable* variables
  - conditional variance or integrated variance
  - default probabilities or "crash" probabilities
  - "true" rates of GDP growth or inflation (opposed to announced rates).

2 Motivation

- Forecast evaluation and comparison for latent variables often involves the use of a "proxy", (i.e., some imperfect estimate of the variable of interest). For example:
  - using squared returns to proxy for the conditional variance
  - using an default indicator variable to proxy for conditional default probabilities
- The use of proxies in forecast evaluation and comparison may or may not lead to complications.
  - See Andersen and Bollerslev (1998), Meddahi (2001) and Hansen and Lunde (2006), for example.

3 "Robust" loss functions

- A property, first considered in Hansen and Lunde (2006), that will guide my analysis of the forecast comparison problem is the following:

Definition 1: A loss function, L, is "*robust*" if the ranking of any two (possibly imperfect) volatility forecasts, h_{1t} and h_{2t}, by expected loss is the same whether the ranking is done using the true conditional variance, σ_t^2, or some conditionally unbiased proxy, $\hat{\sigma}_t^2$.

That is, if $E\left[\hat{\sigma}_t^2|\mathcal{F}_{t-1}\right] = \sigma_t^2$, then

$$E\left[L\left(\sigma_t^2, h_{1t}\right)\right] \gtreqless E\left[L\left(\sigma_t^2, h_{2t}\right)\right] \Leftrightarrow E\left[L\left(\hat{\sigma}_t^2, h_{1t}\right)\right] \gtreqless E\left[L\left(\hat{\sigma}_t^2, h_{2t}\right)\right]$$

4 "Economic" loss functions

- The ideal scenario in forecasting is when the entire decision problem of the forecast user is known to the forecast producer. In such cases we may use the relevant "economic" loss function - see West, *et al.* (1993), Fleming, *et al.* (2001) or Engle, *et al.* (1993) for examples.

- In such cases the forecast becomes just an input to the decision, and the optimal volatility forecast will not generally be the true conditional variance.

- Unfortunately, the economic loss function of the user of a volatility forecast is usually unknown, leading us to rely on "statistical" loss functions.
  - This paper provides guidance on the choice of statistical loss functions for volatility forecasting.

5 Notation

Returns : $r_t|\mathcal{F}_{t-1} \sim F_t\left(0, \sigma_t^2\right)$

Standardised returns : $\varepsilon_t \equiv r_t/\sigma_t \sim F_t\,(0,1)$

Variance : $V_{t-1}\,[r_t] = E_{t-1}\left[r_t^2\right] = \sigma_t^2$

Volatility proxy : $\hat{\sigma}_t^2$, such that $E_{t-1}\left[\hat{\sigma}_t^2\right] = \sigma_t^2$

'Optimal' volatility forecast : $h_t^* = \arg\min_{h_t} \; E_{t-1}\left[L\left(\hat{\sigma}_t^2, h_t\right)\right]$

6 Outline of talk

1. Comparisons using squared returns as a proxy

2. Comparisons using more efficient volatility proxies

3. A class of “robust” loss functions

4. Application to forecasting IBM stock return volatility

5. Conclusions and some extensions

7 Related literature

Forecast evaluation and comparison surveys

Clements (2005)
West (2005)
Diebold and Lopez (1996)

Volatility forecasting surveys

Andersen, Bollerslev, Christoffersen and Diebold (2005)
Shephard (2005)
Poon and Granger (2003)
Bollerslev, Engle and Nelson (1994)

The use of volatility proxies

Pagan and Ullah (1988, *J. Applied Econometrics*)
Andersen and Bollerslev (1998, *IER*)
Meddahi (2001, working paper)
Andersen, Bollerslev and Meddahi (2005, *Econometrica*)
Hansen and Lunde (2006, J. Econometrics)

8 Loss function "robustness" in the literature

- Meddahi (2001) showed that the R^2 from the Mincer-Zarnowitz regression:

$$\hat{\sigma}_t^2 = \beta_0 + \beta_1 h_{it} + e_{it}$$

yields a robust ranking of volatility forecasts.

- Hansen and Lunde (2006) showed that the R^2 from the MZ regression in logs is *not* robust. Further, Hansen and Lunde (2006) provide a sufficient condition for a loss function to be robust:

$$\frac{\partial^3 L\left(\hat{\sigma}_t^2, h\right)}{\partial h \partial \left(\hat{\sigma}_t^2\right)^2} = 0$$

9 Very brief summary of results

- I build on the the work of Andersen and Bollerslev (1998), Meddahi (2001), and Hansen and Lunde (2006) to show two main results:

  1. I analytically derive the problems cause by noise for the 9 most common loss functions, revealing some to be worse than others.

     – Using squared daily returns, the range and realised variance as proxies

  2. I propose a necessary and sufficient class of loss functions for use with a conditionally unbiased, but imperfect, proxy.

     – I derive the homogeneous sub-set of this class of functions, which nests the MSE and QLIKE loss functions, and provide moment conditions for their use in forecast comparison tests

10 Diebold-Mariano (1995) - West (1996) test

- This is the most widely used test for forecast comparison. Let

$$\begin{aligned} d_t &= L\left(\hat{\sigma}_t^2, h_{1,t}\right) - L\left(\hat{\sigma}_t^2, h_{2,t}\right) \\ \text{eg} \quad d_t &= \left(\hat{\sigma}_t^2 - h_{1,t}\right)^2 - \left(\hat{\sigma}_t^2 - h_{2,t}\right)^2 \end{aligned}$$

- If two forecasts yield equal expected loss, for some loss function, then

$$\begin{aligned} H_0 &: E\left[d_t\right] = 0 \\ \text{vs.} \quad H_a &: E\left[d_t\right] \neq 0 \end{aligned}$$

- This test can be conducted as a t-test, with the standard error appropriately adjusted for serial dependence (Diebold-Mariano) and/or estimation error in the forecasts (West).

11 Loss functions used in DMW tests

$$
\begin{aligned}
MSE &: L\left(\hat{\sigma}_t^2, h_t\right) = \left(\hat{\sigma}_t^2 - h_t\right)^2 \\
QLIKE &: L\left(\hat{\sigma}_t^2, h_t\right) = \log h_t + \frac{\hat{\sigma}_t^2}{h_t} \\
MSE\text{-}LOG &: L\left(\hat{\sigma}_t^2, h_t\right) = \left(\log \hat{\sigma}_t^2 - \log h_t\right)^2 \\
MSE\text{-}SD &: L\left(\hat{\sigma}_t^2, h_t\right) = \left(\hat{\sigma}_t - \sqrt{h_t}\right)^2 \\
MSE\text{-}prop &: L\left(\hat{\sigma}_t^2, h_t\right) = \left(\frac{\hat{\sigma}_t^2}{h_t} - 1\right)^2 \\
MAE &: L\left(\hat{\sigma}_t^2, h_t\right) = \left|\hat{\sigma}_t^2 - h_t\right| \\
MAE\text{-}LOG &: L\left(\hat{\sigma}_t^2, h_t\right) = \left|\log \hat{\sigma}_t^2 - \log h_t\right| \\
MAE\text{-}SD &: L\left(\hat{\sigma}_t^2, h_t\right) = \left|\hat{\sigma}_t - \sqrt{h_t}\right| \\
MAE\text{-}prop &: L\left(\hat{\sigma}_t^2, h_t\right) = \left|\frac{\hat{\sigma}_t^2}{h_t} - 1\right|
\end{aligned}
$$

12 A necessary condition for robustness

- If a loss function is "robust"

$$E\left[L\left(\sigma_t^2, h_{1t}\right)\right] \gtreqless E\left[L\left(\sigma_t^2, h_{2t}\right)\right] \Leftrightarrow E\left[L\left(\hat{\sigma}_t^2, h_{1t}\right)\right] \gtreqless E\left[L\left(\hat{\sigma}_t^2, h_{2t}\right)\right]$$

then it follows directly that the optimal forecast under that loss function must be the conditional variance.

- We can thus check a necessary condition for robustness by determining whether the loss function implies $h_t^* = \sigma_t^2$.

13 Optimal forecasts under MSE loss

- MSE loss is the most commonly employed loss function. The optimal forecast under MSE loss is the true conditional variance:

$$h_t^* \equiv \arg\min_h E_{t-1}\left[\left(r_t^2 - h\right)^2\right]$$

$$FOC \qquad h_t^* = E_{t-1}\left[r_t^2\right] = \sigma_t^2$$

- Thus this loss function satisfies the necessary condition. (It also satisfies Hansen and Lunde's sufficient condition.)

14 Optimal forecasts under MAE loss

- One of the most commonly employed alternative loss functions is the absolute-error criterion $L\left(r_t^2, h_t\right) = \left|r_t^2 - h_t\right|$, which yields:

$$
\begin{aligned}
h_t^* &= Median_{t-1}\left[r_t^2\right] \\
&= \sigma_t^2 \cdot \frac{\nu - 2}{\nu} \cdot Median\left[F_{1,\nu}\right], \text{ if } r_t|\mathcal{F}_{t-1} \sim t\left(0, \sigma_t^2, \nu\right), \ \nu > 2 \\
&= \sigma_t^2 \cdot Median\left[\chi_1^2\right], \text{ if } r_t|\mathcal{F}_{t-1} \sim N\left(0, \sigma_t^2\right) \\
&\approx 0.45\sigma_t^2
\end{aligned}
$$

thus this loss function does *not* satisfy the necessary condition for robustness. MAE is a *non-robust* loss function.

15 Optimal forecasts under MSE-SD loss

- Another commonly used loss function is the MSE on standard deviations:

$$L\left(r_t^2, h_t\right) = \left(|r_t| - \sqrt{h_t}\right)^2$$

$$\begin{aligned} h_t^* &= \left(E_{t-1}\left[|r_t|\right]\right)^2 \\ &= \frac{\nu - 2}{\pi} \frac{\Gamma\left(\frac{\nu-1}{2}\right)^2}{\Gamma\left(\frac{\nu}{2}\right)^2} \sigma_t^2 \quad \text{if returns are } t \text{ distributed} \\ &= \frac{2}{\pi}\sigma_t^2 \approx 0.64\sigma_t^2 \quad \text{if returns are normally distributed} \end{aligned}$$

- For both the MAE and the MSE-SD loss functions the distortion is exacerbated when returns have excess kurtosis.

16 Optimal forecasts under various loss functions

	Distribution of daily returns		
Loss function	$r_t \vert \mathcal{F}_{t-1} \sim \left(0, \sigma_t^2\right)$	$t\left(0, \sigma_t^2, 6\right)$	$N\left(0, \sigma_t^2\right)$
MSE, QLIKE	σ_t^2	σ_t^2	σ_t^2
MSE-LOG	$\exp\left\{E_{t-1}\log\varepsilon_t^2]\right\}\sigma_t^2$	$0.22\sigma_t^2$	$0.28\sigma_t^2$
MSE-SD	$\left(E_{t-1}\left\vert r_t \right\vert]\right)^2$	$0.56\sigma_t^2$	$0.64\sigma_t^2$
MSE-prop	$Kurt_{t-1}\left[r_t\right]\sigma_t^2$	$6.00\sigma_t^2$	$3.00\sigma_t^2$
MAE	$Median_{t-1}\left[r_t^2\right]$	$0.34\sigma_t^2$	$0.45\sigma_t^2$
MAE-SD	$Median_{t-1}\left[r_t^2\right]$	$0.34\sigma_t^2$	$0.45\sigma_t^2$
MAE-prop	n/a	$2.73\sigma_t^2$	$2.36\sigma_t^2$

17 Using better volatility proxies

- What if we employ volatility proxies that are known to have less noise?

- Consider the following simple DGP: there are m equally-spaced observations per trade day, and let $r_{i,m,t}$ denote the i^{th} intra-daily return on day t.

$$\begin{aligned} r_t &= d \ln P_t = \sigma_t dW_t \\ \sigma_\tau &= \sigma_t \ \forall \ \tau \in (t-1, t] \end{aligned}$$

$$\begin{aligned} r_{i,m,t} &\equiv \int\limits_{(i-1)/m}^{i/m} r_\tau d\tau = \sigma_t \int\limits_{(i-1)/m}^{i/m} dW_\tau \\ \text{so } \left\{ r_{i,m,t} \right\}_{i=1}^{m} &\sim iid \ N\left(0, \frac{\sigma_t^2}{m}\right) \end{aligned}$$

- One alternative volatility proxy is "realized volatility", see Andersen, *et al.* (2001a, 2003), and Barndorff-Neilsen and Shephard (2002, 2004):

$$RV_t^{(m)} \equiv \sum_{i=1}^{m} r_{i,m,t}^2$$

- Another commonly-used alternative to squared returns is the intra-daily range, see Parkinson (1980) and Feller (1951):

$$RG_t \equiv \sup_{\tau} \log P_\tau - \inf_{\tau} \log P_\tau,\ t-1 < \tau \leq t$$

- Efficiency comparison under this DGP:

$$\begin{aligned} MSE_{t-1}\left[r_t^2\right] &= 2\sigma_t^4 \\ MSE_{t-1}\left[RV_t^{(m)}\right] &= 2\sigma_t^4/m \\ MSE_{t-1}\left[RG_t^{*2}\right] &\approx 0.4073\sigma_t^4 \end{aligned}$$

18 Optimal forecasts - analytical, constant vol

Loss function	Range	Realised volatility Daily: $m = 1$	30-min: $m = 13$	5-min: $m = 78$	True: $m \to \infty$
MSE, QLIKE	σ_t^2	σ_t^2	σ_t^2	σ_t^2	σ_t^2
MSE-LOG	$0.85\sigma_t^2$	$0.28\sigma_t^2$	$0.91\sigma_t^2$	$0.98\sigma_t^2$	σ_t^2
MSE-SD	$0.92\sigma_t^2$	$0.56\sigma_t^2$	$0.96\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MSE-prop	$1.41\sigma_t^2$	$3.00\sigma_t^2$	$1.15\sigma_t^2$	$1.03\sigma_t^2$	σ_t^2
MAE	$0.83\sigma_t^2$	$0.46\sigma_t^2$	$0.95\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MAE-SD	$0.83\sigma_t^2$	$0.46\sigma_t^2$	$0.95\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MAE-prop	$1.19\sigma_t^2$	$2.36\sigma_t^2$	$1.10\sigma_t^2$	$1.02\sigma_t^2$	σ_t^2

19 Optimal forecasts - simulation, GARCH SV

Loss function	Range	Realised volatility Daily: $m = 1$	30-min: $m = 13$	5-min: $m = 78$	True: $m \to \infty$
MSE, QLIKE	$0.99\sigma_t^2$	σ_t^2	σ_t^2	σ_t^2	σ_t^2
MSE-LOG	$0.83\sigma_t^2$	$0.28\sigma_t^2$	$0.92\sigma_t^2$	$0.98\sigma_t^2$	σ_t^2
MSE-SD	$0.91\sigma_t^2$	$0.63\sigma_t^2$	$0.96\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MSE-prop	$1.40\sigma_t^2$	$3.02\sigma_t^2$	$1.16\sigma_t^2$	$1.03\sigma_t^2$	σ_t^2
MAE	$0.82\sigma_t^2$	$0.46\sigma_t^2$	$0.94\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MAE-SD	$0.82\sigma_t^2$	$0.46\sigma_t^2$	$0.94\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MAE-prop	$1.18\sigma_t^2$	$2.37\sigma_t^2$	$1.10\sigma_t^2$	$1.01\sigma_t^2$	σ_t^2

20 Optimal forecasts - simulation, log-normal SV

Loss function	Range	Realised volatility Daily: $m = 1$	30-min: $m = 13$	5-min: $m = 78$	True: $m \to \infty$
MSE, QLIKE	$0.99\sigma_t^2$	σ_t^2	σ_t^2	σ_t^2	σ_t^2
MSE-LOG	$0.83\sigma_t^2$	$0.28\sigma_t^2$	$0.92\sigma_t^2$	$0.98\sigma_t^2$	σ_t^2
MSE-SD	$0.91\sigma_t^2$	$0.63\sigma_t^2$	$0.96\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MSE-prop	$1.40\sigma_t^2$	$3.03\sigma_t^2$	$1.16\sigma_t^2$	$1.03\sigma_t^2$	σ_t^2
MAE	$0.82\sigma_t^2$	$0.46\sigma_t^2$	$0.94\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MAE-SD	$0.82\sigma_t^2$	$0.46\sigma_t^2$	$0.94\sigma_t^2$	$0.99\sigma_t^2$	σ_t^2
MAE-prop	$1.18\sigma_t^2$	$2.37\sigma_t^2$	$1.10\sigma_t^2$	$1.02\sigma_t^2$	σ_t^2

21 Optimal forecasts - simulation, two-factor SV

		Realised volatility			
Loss function	Range	Daily: $m=1$	30-min: $m=13$	5-min: $m=78$	True: $m\to\infty$
MSE, QLIKE	σ_t^2	$1.01\sigma_t^2$	σ_t^2	σ_t^2	σ_t^2
MSE-LOG	$0.35\sigma_t^2$	$0.12\sigma_t^2$	$0.37\sigma_t^2$	$0.41\sigma_t^2$	σ_t^2
MSE-SD	$0.57\sigma_t^2$	$0.40\sigma_t^2$	$0.58\sigma_t^2$	$0.62\sigma_t^2$	σ_t^2
MSE-prop	$9.79\sigma_t^2$	$20.6\sigma_t^2$	$9.03\sigma_t^2$	$6.70\sigma_t^2$	σ_t^2
MAE	$0.31\sigma_t^2$	$0.17\sigma_t^2$	$0.32\sigma_t^2$	$0.35\sigma_t^2$	σ_t^2
MAE-SD	$0.31\sigma_t^2$	$0.17\sigma_t^2$	$0.32\sigma_t^2$	$0.35\sigma_t^2$	σ_t^2
MAE-prop	$3.47\sigma_t^2$	$6.60\sigma_t^2$	$3.33\sigma_t^2$	$2.98\sigma_t^2$	σ_t^2

22 SV models used in the simulations

- For the simulations we used the same models and parameter values as used in Gonçalves and Meddahi (2005):

1. GARCH diffusion, as in Anderson and Bollerslev (1998):

$$\begin{aligned} d\log P_t &= 0.0314dt + \nu_t\left(-0.576dW_{1t} + \sqrt{1-0.576^2}dW_{2t}\right) \\ d\nu_t^2 &= 0.035\left(0.636 - \nu_t^2\right)dt + 0.144\nu_t^2 dW_{1t} \end{aligned}$$

2. Log-normal diffusion, as in Anderson, Benzoni and Lund (2002):

$$\begin{aligned} d\log P_t &= 0.0314dt + \nu_t\left(-0.576dW_{1t} + \sqrt{1-0.576^2}dW_{2t}\right) \\ d\log\nu_t^2 &= -0.0136\left(0.8382 + \log\nu_t^2\right)dt + 0.1148dW_{1t} \end{aligned}$$

23 SV models used in the simulations, cont'd

3. Two-factor diffusion, as in Chernov, Gallant, Ghysels and Tauchen (2003):

$$
\begin{aligned}
d \log P_t &= 0.030 dt + \nu_t \Big(-0.30 dW_{1t} - 0.30 dW_{2t} \\
&\quad + \sqrt{1 - 0.3^2 - 0.3^2} dW_{3t} \Big) \\
\nu_t^2 &= \text{s-exp}\left\{ -1.2 + 0.04 \nu_{1t}^2 + 1.5 \nu_{2t}^2 \right\} \\
d\nu_{1t}^2 &= -0.00137 \nu_{1t}^2 dt + dW_{1t} \\
d\nu_{2t}^2 &= -1.386 \nu_{2t}^2 dt + \left(1 + 0.25 \nu_{2t}^2\right) dW_{2t}
\end{aligned}
$$

$$
\text{where} \quad \text{s-exp}\{x\} = \begin{cases} \exp\{x\}, & x \le x_0 \\ \exp\{x_0\} \sqrt{1 - x_0 + x^2/x_0}, & x > x_0 \end{cases}
$$

24 Generalising these results

- Using a 2^{nd}-order mean-value expansion for L, the first-order condition is:

$$0 = E_{t-1}\left[\frac{\partial L\left(\hat{\sigma}_t^2, h_t^*\right)}{\partial h}\right] = \frac{\partial L\left(\sigma_t^2, h_t^*\right)}{\partial h} + \frac{\partial^3 L\left(\ddot{\sigma}_t^2, h_t^*\right)}{\partial\left(\sigma_t^2\right)^2 \partial h} \cdot \frac{1}{2} V_{t-1}\left[\hat{\sigma}_t^2\right]$$

1. If $\partial^3 L / \partial\left(\sigma_t^2\right)^2 \partial h = 0$ for all $\left(\sigma^2, h\right)$, then $h_t^* = \sigma_t^2$. This is a key result of Hansen and Lunde (2006)

2. If $\partial^3 L / \partial\left(\sigma_t^2\right)^2 \partial h > 0$ for all $\left(\sigma^2, h\right)$, then we must have $\partial L / \partial h < 0$, implying $h_t^* < \sigma_t^2$. Eg: MSE-SD and MSE-log loss functions.

3. If $\partial^3 L / \partial\left(\sigma_t^2\right)^2 \partial h < 0$ for all $\left(\sigma^2, h\right)$, then we must have $\partial L / \partial h > 0$, implying $h_t^* > \sigma_t^2$. Eg: MSE-prop loss function.

25 A class of robust loss functions

- Both the MSE and QLIKE loss functions yielded the conditional variance as the optimal forecast.

- This leads to the question: Is there a general class of such loss functions?

- The following proposition suggests a class of loss functions, related to the linear-exponentional family of densities of Gourieroux, *et al.* (1984), and to Gourieroux, *et al.* (1987).

Assumptions:

A1: $E\left[\hat{\sigma}_t^2|\mathcal{F}_{t-1}\right] = \sigma_t^2$

A2: $\hat{\sigma}_t^2|\mathcal{F}_{t-1} \sim F_t \in \tilde{F}$, the set of all absolutely continuous distribution functions on $\mathbb{R}_+$.

A3: L is twice continuously differentiable with respect to h and $\hat{\sigma}^2$, and has a unique minimum at $\hat{\sigma}^2 = h$.

A4: There exists some $h_t^* \in int\,(\mathcal{H})$ such that $h_t^* = E_{t-1}\left[\hat{\sigma}_t^2\right]$, where $\mathcal{H}$ is a compact subset of $\mathbb{R}_{++}$.

A5: L and F_t are such that: (a) $E_{t-1}\left[L\left(\hat{\sigma}_t^2, h\right)\right] < \infty$ for some $h \in \mathcal{H}$; (b) $E_{t-1}\left[\partial L\left(\hat{\sigma}_t^2, \sigma_t^2\right)/\partial h\right] < \infty$; and (c) $\left|E_{t-1}\left[\partial^2 L\left(\hat{\sigma}_t^2, \sigma_t^2\right)/\partial h^2\right]\right| < \infty$ for all t.

.

Proposition 2:

Let assumptions A1 to A5 hold. Then a loss function L is "robust" *if and only if* it takes the following form:

$$L\left(\hat{\sigma}^2, h\right) = \tilde{C}(h) + B\left(\hat{\sigma}^2\right) + C(h)\left(\hat{\sigma}^2 - h\right)$$

where B and C are twice continuously differentiable, C is a strictly decreasing function on $\mathcal{H}$, and $\tilde{C}$ is the anti-derivative of C.

26 Sub-sets of robust loss functions - 1

Proposition 3:

(i) The "MSE" loss function is the *only* robust loss function that depends solely on the forecast error, $\hat{\sigma}^2 - h$.

(ii) The "QLIKE" loss function is the *only* robust loss function that depends solely on the standardised forecast error, $\hat{\sigma}^2 / h$.

27 A parametric family of loss functions for volatility forecast comparison

- We now seek to find a parametric family of loss functions within the broader class of robust loss functions, that nests both MSE and QLIKE loss functions.

- Note that both MSE and QLIKE loss functions have first-order conditions that can be written as:

$$E_{t-1}\left[\frac{\partial L\left(\hat{\sigma}_t^2, h_t^*\right)}{\partial h}\right] \;=\; 0 = {h_t^*}^{k-2}\left(E_{t-1}\left[\hat{\sigma}_t^2\right] - h_t^*\right),\; k \in \mathbb{R}$$

Proposition 4:

(i) The following family of functions

$$L\left(\hat{\sigma}^2, h; k\right) = \begin{cases} \left(\hat{\sigma}^{2(k-1)} - h^{k-1}\right)/\left(k-1\right) - \left(\hat{\sigma}^{2k} - h^{k}\right)/k, & \text{for } k \notin \{0,1\} \\ h - \hat{\sigma}^2 + \hat{\sigma}^2 \log\left(\hat{\sigma}^2/h\right), & \text{for } k = 1 \\ \hat{\sigma}^2/h - \log\left(\hat{\sigma}^2/h\right) - 1, & \text{for } k = 0 \end{cases}$$

satisfy $L\left(h, h; k\right) = 0$ for all $h \in \mathcal{H}$, and are of the form in Proposition 2.

(ii) The family of loss functions in part (i) corresponds to the entire sub-set of homogeneous robust loss functions. The degree of homogeneity is equal to k.

Aside: Recall that homogeneity of degree k implies

$$L\left(a\hat{\sigma}^2, ah\right) = a^k L\left(\hat{\sigma}^2, h\right), \forall a > 0$$

28 Units of measurement and forecast rankings

- The choice of units in many economic and financial problems is arbitrary (prices in dollars versus cents, returns in percentages versus decimals)

Proposition 5:

(i) The ranking of any two (possibly imperfect) volatility forecasts by expected loss is invariant to a re-scaling of the data if the loss function is robust and homogeneous.

(ii) The ranking of any two (possibly imperfect) volatility forecasts by expected loss may *not* be invariant to a re-scaling of the data if the loss function is robust but not homogeneous.

Proof: (ii) Consider the following example: $\sigma_t^2 = 1\,\forall t,\ (h_{1t}, h_{2t}) = (\gamma_1, \gamma_2)\ \forall t$, and $\hat{\sigma}_t^2$ is such that $E_{t-1}\left[\hat{\sigma}_t^2\right] = 1$ a.s. $\forall t$. As a robust but non-homogeneous loss we will use the one generated by the following specification for C':

$$C'(h) = -\log(1+h)$$

Given this set-up, we have

$$\begin{aligned} E\left[L\left(a\hat{\sigma}_t^2, ah_{it}\right)\right] &= \frac{1}{4}\left[a\gamma_i\left(3a\gamma_i+2\right) - 2\left(1+a\gamma_i\right)^2\log\left(1+a\gamma_i\right)\right] \\ &\quad + a\left[a\gamma_i - \left(1+a\gamma_i\right)\log\left(1+a\gamma_i\right)\right]\left(1-\gamma_i\right) + const \end{aligned}$$

Then define

$$d_t(\gamma_1, \gamma_2, a) \equiv L\left(a\hat{\sigma}_t^2, a\gamma_1\right) - L\left(a\hat{\sigma}_t^2, a\gamma_2\right)$$

Then note that

$$\begin{aligned} E\left[d_t(0.33, 1.5, 1)\right] &= -0.0087 \\ \text{but}\quad E\left[d_t(0.33, 1.5, 2)\right] &= +0.0061 \end{aligned}$$

Proposition 6(ii):

$$\text{Let } \; d_t(k) = L\left(\hat{\sigma}_t^2, h_{1t}; k\right) - L\left(\hat{\sigma}_t^2, h_{2t}; k\right)$$

Sufficient conditions for $E\left[d_t(k)^2\right] < \infty$ are:

1) inf $\mathcal{H}_i \equiv c_i > 0$ for $i = 1, 2$,

2) $E\left[h_{it}^p\right] < \infty$, $i = 1, 2$, and $E\left[\hat{\sigma}_t^q\right] < \infty$,

where p and q are as follows:

$p = \max[0, 2k]$,	$q = \max[4 + \delta, 4k]$,	$k \notin \{0, 1\}$
$p = 2(e+1)/e \approx 2.74$,	$q = 4(e+1)/e \approx 5.47$,	$k = 1$
$p = 2/e + \delta \approx 0.74 + \delta$,	$q = 4 + \delta$,	$k = 0$

for $\delta > 0$.

29 Forecasting IBM return volatility

- Daily and intra-daily data on IBM from January 1993 to December 2003, 2772 observations

- I consider two simple but widely-used volatility models:

$$\textit{Rolling window} \quad : \quad h_{1t} = \frac{1}{60} \sum_{j=1}^{60} r_{t-j}^2$$

$$\textit{RiskMetrics} \quad : \quad h_{2t} = \lambda h_{2t-1} + (1-\lambda)\, r_{t-1}^2,\ \lambda = 0.94$$

- First 272 observations are used for estimation, last 2500 observations are used for forecast comparison

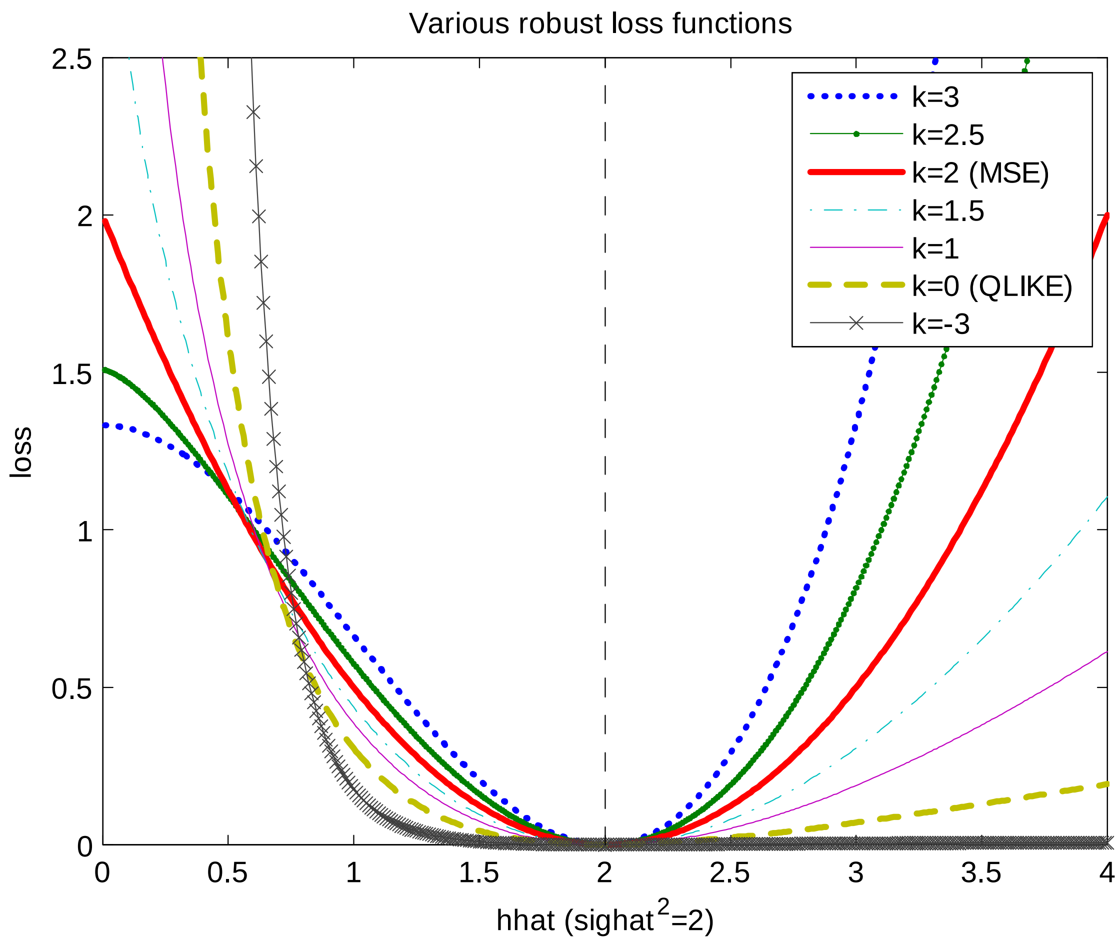

Figure 1: *Loss functions for various choices of* k. *True* $\hat{\sigma}^2=2$ *in this example, with the volatility forecast ranging between 0 and 4.*

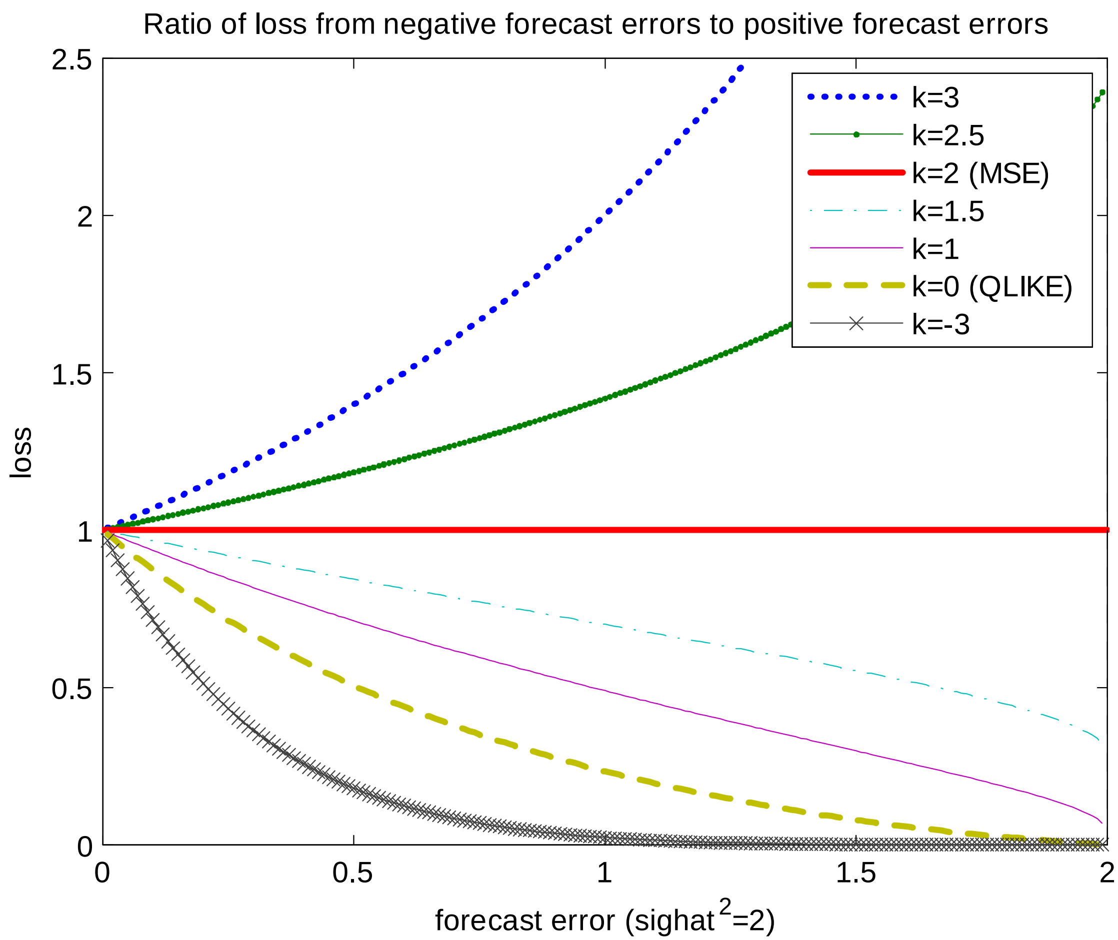

Figure 2: *Ratio of losses from negative forecast errors to positive forecast errors, for various choices of* b*. True* $\hat{\sigma}^2{=}2$ *in this example, with the volatility forecast ranging between 0 and 4.*

30 Mincer-Zarnowitz regression results

- MZ regressions: $\hat{\sigma}_t^2 = \beta_0 + \beta_1 h_{it} + e_{it}$

		Volatility proxy	
		Daily squared return	*5-min realised vol*
Rolling window	$\hat{\beta}_0$	2.13^*	2.33^*
	$(s.e.)$	(0.48)	(0.40)
	$\hat{\beta}_1$	0.55^*	0.53^*
	$(s.e.)$	(0.09)	(0.07)
	χ_2^2-stat	25.63^*	43.86^*
RiskMetrics	$\hat{\beta}_0$	2.39^*	2.43^*
	$(s.e.)$	(0.46)	(0.42)
	$\hat{\beta}_1$	0.50^*	0.51^*
	$(s.e.)$	(0.09)	(0.09)
	χ_2^2-stat	32.99^*	35.93^*

- So it is clear that we are comparing two *imperfect* forecasts.

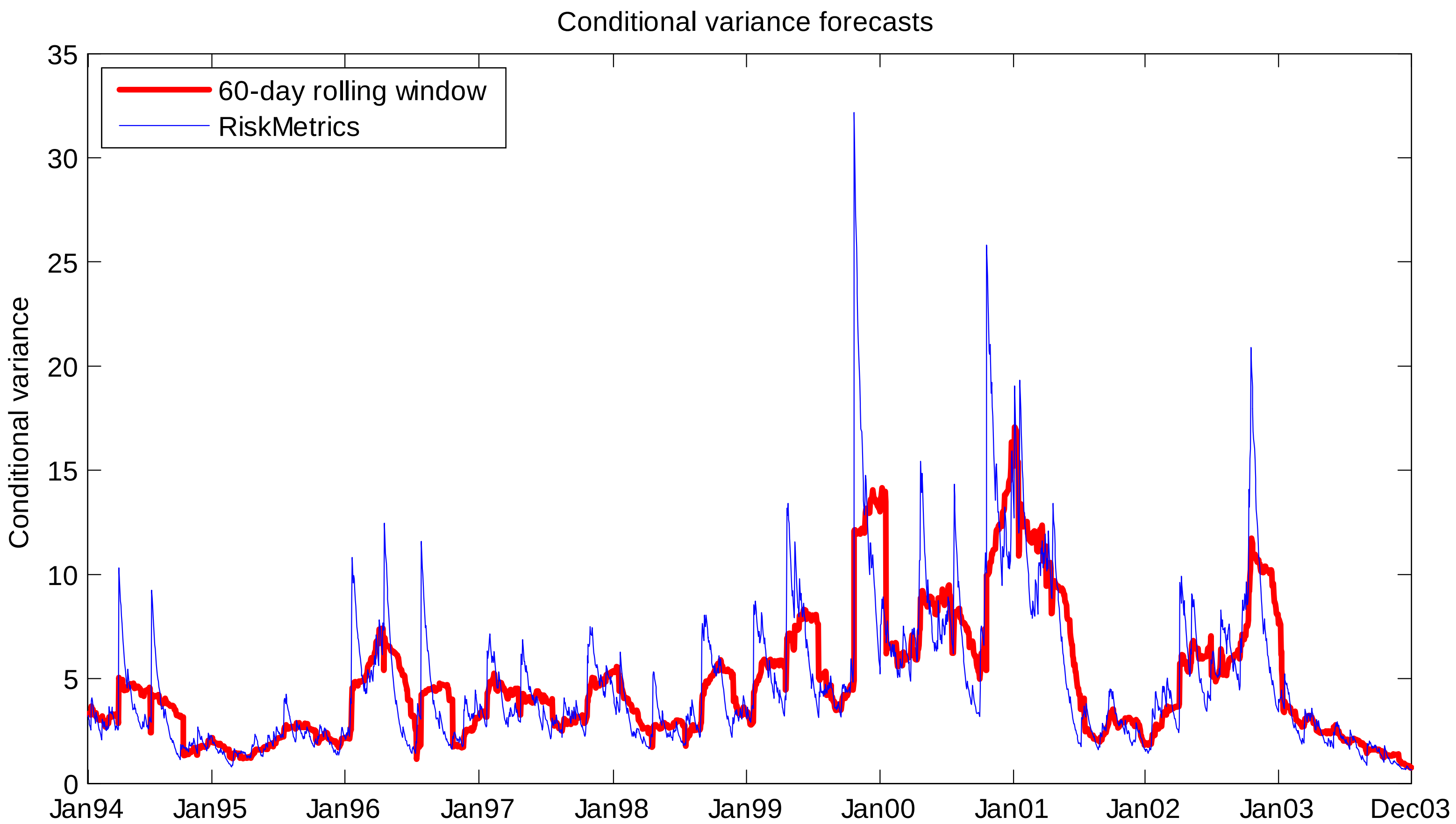

Figure 3: *Conditional variance forecasts from the two simple models, January 1994 to December 2003.*

31 DMW forecast comparison tests

t-statistics	*Volatility proxy*			
	Daily	*65-min*	*15-min*	*5-min*
Loss function	*squared return*	*realised vol*	*realised vol*	*realised vol*
$k = 3$	-1.58	-1.66	-1.30	-1.35
$k = 2$ (MSE)	-0.59	-0.80	-0.03	-0.13
$k = 1$	1.30	1.04	1.65	-1.55
$k = 0$ (QLIKE)	1.94	2.21*	2.73*	2.41*
$k = -3$	-0.17	0.25	1.63	0.65

- Under QLIKE loss, RiskMetrics significantly out-performs the rolling window forecasts.

- Under MSE loss, the rolling window forecasts are weakly out-performs the RiskMetrics forecasts.

35 Conclusions

- We have shown some of the problems that arise when an imperfect proxy is employed to compare volatility forecasts, extending the work of Andersen and Bollerslev (1998), Meddahi (2001) and Hansen and Lunde (2006).
  - More accurate volatility proxies were shown to alleviate these problems, but they do not completely remove them.

- A necessary and sufficient condition on the form of loss functions used for volatility forecast comparison was presented, ruling out some previously-used loss functions
  - A new parametric family of loss functions was proposed, which nests MSE and QLIKE, and works with noisy volatility proxies.