Abstract:

I compare the forecasts of returns from the mean predictor (optimal under MSE), with the pseudo-optimal and optimal predictor for an asymmetric loss function under the assumption that agents have asymmetric LINLIN loss function. I consider both univariate and multivariate cases. For the multivariate case, I generalize the LINLIN loss function to a multivariate LINLIN loss function and use a normal and t-diagonal-BEKK GARCH(1,1) model to predict the time varying variances. The results strongly suggest not to use the conditional mean predictor under any kind of asymmetry. In general, forecasts can be improved considerably by the use of optimal predictor versus the pseudo-optimal predictor suggesting that the loss reduction due to using the optimal predictor can actually be important for a practitioner as well.

Keywords: Loss function; GARCH models; Volatility forecasting; Time series; Multivariate time series.
“An assumption of symmetry about the conditional mean ... is likely to be an easy one to accept ... an assumption of symmetry for the cost function is much less acceptable.”

Granger, Newbold (1986, p.125)

1. Introduction

In the literature, a widely used forecast evaluation criteria is the MSE, which is a symmetric quadratic loss function. MSE penalizes the positive errors and negative errors of the same magnitude equally. However, in finance we know that realistically, forecasters do not necessarily have a quadratic cost nor a symmetric loss function. Over prediction can be more (less) costly than under prediction depending on the situation. For example, if we consider the positive relationship between the volatility of underlying stock prices and call option prices (Brailsford, Faff, 1996), an under prediction of volatility will result in a downward bias to the estimates of the call option price which is probably of more concern to a seller than a buyer while the converse holds for the over predictions of volatility (Mc Millian, Speight, Apgwilym, 2000). Despite this fact, MSE has become a very popular loss function mostly due to mathematical convenience.

Symmetric loss functions are convenient because the analytical expression for the optimal predictor is straightforward: it is the conditional mean. However, if we use a general loss function, deriving the closed form for the optimal predictor analytically can be very challenging, and even not viable. Studies have avoided using general asymmetric loss functions mainly because most of the time the closed form for the optimal predictor does not exist. Under asymmetric loss, the optimal predictor is no longer the conditional mean. Granger (1969) showed that the optimal predictor under asymmetric loss is the
conditional mean plus a constant bias term. Granger (1969) assumed a constant conditional prediction error variance. Christoffersen and Diebold (henceforth abbreviated CD) (1996, 1997) considered the same problem, and generalized Granger’s result. They showed that for conditionally Gaussian processes if an agent has an asymmetric loss function, adding a constant bias term is not sufficient and that time varying second order moments become relevant for optimal prediction. They derived the analytical expression for the optimal predictor for two specific asymmetric loss functions, the LINLIN and the LINEX. Through a Monte Carlo simulation for a conditionally Gaussian GARCH(1,1) process under LINLIN loss CD (1996) show that 35% loss reduction is attainable when the optimal predictor is used.

Although CD (1996, 1997) have important practical implications, there is no empirical study that illustrates the gain in adding a time-varying bias term to the conditional mean to obtain the optimal predictor when agents have asymmetric loss. In this paper I bridge this gap by illustrating the loss associated with using the optimal, the pseudo-optimal and the conditional mean predictor under asymmetric loss. One of the motivations behind this study is the fact that the realized average loss reduction might be considerably smaller than the theoretically expected losses in a world of parameter uncertainty and may not be of interest to a practitioner. I demonstrate that in practice incorporating time varying second moments can actually be of interest to a financial forecaster reducing their loss significantly. The second contribution of this paper to the existing literature is deriving the optimal, pseudo-optimal predictors for a multivariate asymmetric loss function, and using them to evaluate multivariate forecasts from multivariate GARCH models.
I consider returns on three representative exchange rates and returns on five representative market indices that are commonly used in empirical work. I consider different predictors of the variance since there is no unanimous agreement on the best variance model in the literature. I choose the most popular models that are believed to characterize conditional volatility in asset returns. I consider normal-GARCH(1,1), t-GARCH(1,1), EGARCH(1,1) and a nonparametric model for univariate variance estimation and forecasting and normal and t-diagonal BEKK GARCH (1,1) models for the multivariate case. I empirically illustrate that the loss associated by using the optimal predictor which incorporates time varying second order moments, is much smaller than the loss associated with using the conditional mean predictor when agents have asymmetric loss function for forecast horizons of one week and four weeks. Moreover, when I compare the difference in loss between the optimal predictor and the pseudo-optimal predictor, in 5 out of the 7 series the optimal predictor outperforms the pseudo-optimal predictor. However, the magnitude of loss reduction is very sensitive to the degree of asymmetry, the conditional variance parameters being used, and the forecast horizon. The multivariate case supports the use of optimal predictor especially for moderate and high degrees of asymmetry as well. When agents have asymmetric loss function, the gain using the optimal predictor versus the pseudo-optimal predictor is clearer compared to the univariate case.

Section 2, is on univariate forecasting under asymmetric loss where I introduce the LINLIN asymmetric loss function, and the univariate variance models I use. Section 3 is on multivariate forecasting under asymmetric loss where I derive the optimal and the pseudo-optimal predictors for a multivariate asymmetric generalized loss function and
use it to compare forecasts from normal and t-diagonal-BEKK GARCH(1,1) multivariate variance models. Section 4 describes the data. Section 5 presents the results and Section 6 concludes.

2. Univariate Forecasting with asymmetric Loss

2.1. LINLIN Loss Function:

The LINLIN loss function was used by Granger (1969). It is linear on each side of the origin; however, positive errors are penalized differently than the negative errors because the lines have different slopes in each side of the origin. The LINLIN loss function is

\[
L(y_{t+h} - \hat{y}_{t+h}) = \begin{cases} 
  a |y_{t+h} - \hat{y}_{t+h}| & \text{if } (y_{t+h} - \hat{y}_{t+h}) > 0 \\
  b |y_{t+h} - \hat{y}_{t+h}| & \text{if } (y_{t+h} - \hat{y}_{t+h}) \leq 0
\end{cases}
\]

where \(L(\cdot)\) is a loss function defined on \(h\)-step-ahead prediction error, \(y_{t+h}\) is the realized value of \(y\), \(t+h\) periods ahead and \(\hat{y}_{t+h}\) is the predicted value of \(y_{t+h}\). The ratio \(a/b\) measures the cost of under predicting relative to the cost of over predicting. If \(a/b=2\), it means that the loss associated with a positive error is twice as much the loss associated with negative error of the same magnitude. If \(a=b\), the average loss is identical to MAE.

CD (1997) derived the optimal predictor, a pseudo-optimal predictor and the expected losses associated with each predictor for the LINLIN asymmetric loss function under the assumption of conditional normality. Given \(y_{t+h} | \Omega_t \sim N(\mu_{t+h}, \sigma_{t+h})\), they showed the optimal predictor for \(y_{t+h}\), is \(\mu_{t+h} + \sigma_{t+h} \Phi^{-1}(a/(a+b))\), and the pseudo-optimal predictor is \(\mu_{t+h} + \sigma_h \Phi^{-1}(a/(a+b))\), where \(\sigma_h^2\) is the \(h\)-step-ahead homoscedastic prediction error variance and \(\Phi(z)\) is the N-(0,1) c.d.f. Under
asymmetric loss, if the conditional heteroscedasticity is ignored, the associated conditionally expected loss will be greater then the case when optimal predictor is used. The pseudo-optimal predictor coincides with the optimal predictor when $\sigma_h^2 = \sigma_{t+h|t}^2$. Notice that for a given series, and for certain degree of asymmetry the difference in mean losses associated with using the optimal versus the pseudo optimal predictor, depend on the deviation of the square root of conditional variance from its mean ($\sigma_{t+h|t} - \sigma_h$). Thus the difference between the mean losses is proportional to this variation. If this variation is large, then incorporating the time varying second order moments can be very crucial in terms of reducing the losses.

In order to mimic forecasting in real time, estimation is done in rolling windows of fixed length of one week and one-step-ahead ($h=1$) and four-step-ahead ($h=4$) forecasts are constructed. For the series I consider, the conditional means are constant. I obtain the conditional variance using different models. I do not limit the analysis only to conditionally Gaussian processes. Although CD assumed conditional normality, it is a common fact that financial return series have excessive kurtosis. To take this fact into account I also estimate a GARCH(1,1) model with t-distributed innovations and an EGARCH(1,1) model with a generalized error distribution. The EGARCH(1,1) model with generalized error distribution allows both for asymmetry in the conditional variance and conditional non-normality. In order to prevent any possible misspecification problems due to parameterization, I also use a nonparametric model. I compute the value of the optimal, the pseudo-optimal and the conditional mean predictors and corresponding losses over all the rolling windows of $h=1$ and take the simple average to calculate the average losses associated with them. Although, the computation is straight
forward for $h=1$, some distributional and second moment related problems have to be considered for $h=4$.

When forecasting out-of-sample with GARCH models, although the Gaussian assumption holds for $h=1$, the distribution of the prediction error is not known for $h>1$. For this case I used Cornish–Fisher Asymptotic Expansion to approximate the prediction error distribution [See Bailie and Bollerslev (1992)] when the returns are assumed to have conditional normal distribution. The approximated four-step-ahead quantile values are used in the optimal and pseudo-optimal predictor formulas for the expression $\Phi^{-1}(a/(a+b))$. Another difficulty when forecasting out-of-sample with GARCH models for longer horizons ($h>1$) is the calculation of the $\text{var}(y_{t+h})$. Again the formulas are readily available for a general ARMA(k,l)-GARCH(p,q) model for a general $h$ [See Bailie and Bollerslev (1992) for more detail]. However, expressions of the $h$-step-ahead conditional variance has to be analytically computed for EGARCH, and multivariate GARCH models. The results don’t exist for the nonparametric model as well. This is beyond the scope of this paper. For practical purposes, for the $h$-step-ahead prediction, when $h>1$ the losses associated with each predictor are calculated for the GARCH(1,1)-n using the existing theoretical results. I consider different degrees of asymmetry to compare the loss associated with using different predictors. Specifically, I fix $b=1$ and change the values of $a$. I consider cases up to where $a/b=20$. This asymmetric penalization scheme is plausible in finance.

2.2. **Univariate Variance Models:**

GARCH Models:
The most commonly used model for time-varying volatility is the G/ARCH model of Engle (1982) and Bollerslev (1986). A GARCH(1,1) model for the return on a financial asset, \( r_t \), can be written as 
\[
\sigma_t^2 = \gamma + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2
\]
where \( \gamma > 0, \alpha \geq 0 \) and \( \beta \geq 0 \). I assume \( z_t \) has finite first and second moments. For a normal GARCH(1,1) model, denoted n-GARCH(1,1), I assume an independent normal innovation. Since financial returns are known to be heavy tailed, and often the conditional normality assumption is inappropriate, I also estimate GARCH(1,1) models with conditional Student-t distributions, denoted by t-GARCH(1,1), for each return series. The one-step-ahead optimal and the pseudo-optimal predictors for the GARCH(1,1) model under conditional-t-distribution assumption are:
\[
y_{t+1} = \mu_{t+1|t} + \sigma_{t+1} F^{-1}(a/(a+b),\nu) \quad \text{and} \quad y_{t+1} = \mu_{t+1|t'} + \sigma_{t+1} F^{-1}(a/(a+b),\nu),
\]
where \( \nu \) is the data specific degrees of freedom and \( F^{-1} \) is the inverse of conditional t cumulative distribution function.

Another feature of the financial data is the skewness inherent in asset return volatility. Especially with the stock return data, it is found that an unexpected drop in price increases the predictable volatility more than an unexpected increase in price of similar magnitude, which is commonly referred as the “leverage effect”[Black (1976)]. In order to capture this skewness inherent in asset return volatility, I estimate the most commonly used asymmetric variance model, the EGARCH(1,1)-model introduced by Nelson (1991). I estimate an EGARCH(1,1) model with a generalized error distribution. The conditional variance for the EGARCH(1,1) model can be written as
\[
\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha |r_{t-1}| \sigma_{t-1} + \gamma (r_{t-1} / \sigma_{t-1})
\]
where \( \gamma \) is the parameter that captures the leverage effect. If \( \gamma \neq 0 \), then the conditional variance is asymmetric. This model
has two advantages; first it allows for the asymmetry and captures the leverage effect. Second it does not assume conditional normality, thus allows for heavy-tailed distributions.

Nonparametric variance Model:

Nonparametric models are popular because they do not introduce any parametric assumption for the underlying distribution and thus prevent the bias problem due to misspecification. Predicting exchange rate returns nonparametrically goes back to Diebold and Nason (1990). The idea behind using nonparametric techniques to estimate and predict exchange rates is to exploit any non-linearities that may be present in the financial return data. Pagan and Schewert (1990) also use this technique for in sample prediction as well.

I fit a nonparametric model for the asset returns conditioning on the lagged returns. I choose a Gaussian multivariate kernel, and I consider the optimal bandwidth to be fixed for a given data set [see Nadaraya (1964) and Watson (1964)]. To predict the second order conditional moment I use the formula in Pagan and Schwert (1990). When forecasting out of sample, I replace the time varying conditional mean and conditional standard deviations in the CD (1997) optimal prediction formula by their nonparametric estimators. In the pseudo-optimal prediction formula, I replace the constant conditional standard deviation with the sample standard deviation of the in sample nonparametric residuals.

3. Multivariate forecasting with asymmetric loss

3.1 Multivariate Loss function:
The theory of forecasting with asymmetric loss as originally presented by Granger (1969), and further developed by CD (1996, 1997), only considered the prediction of a single variable based on its own passed values. In this section I extend the theory to a multivariate framework in which more than one series is to be forecasted. Let $Y_{t+h}$ be an $n \times 1$ vector of variables to be forecasted at horizon $h$. $\hat{Y}_{t+h}$ be the $n \times 1$ vector of forecasts and $e_{t+h} = Y_{t+h} - \hat{Y}_{t+h}$ is the $n \times 1$ vector of forecast errors. I then have the following extension of CD’s (1997) Proposition 1.

**Proposition 1:** If $Y_{t+h} | \Omega_t \sim N(\mu_{t+h|t}, \Sigma_{t+h|t})$ is conditionally multivariate normal and $L(e_{t+h})$ is any loss function defined on the vector of $h$-step-ahead prediction error, then the optimal predictor is of the form $\hat{Y}_{t+h} = \mu_{t+h|t} + \alpha_{t+h|t}$, where $\alpha_{t+h|t}$ depends only on the loss function and the conditional prediction error variance-covariance matrix $\Sigma_{t+h|t} = \text{var}(Y_{t+h} | \Omega_t) = \text{var}(e_{t+h} | \Omega_t)$.

Proof: See Appendix.

Following Zellner (1986), it may be reasonable to assume the loss function is additively separable in the $n$ prediction errors and can be written as

$L(Y_{t+h} - \hat{Y}_{t+h}) = \sum_{i=1}^{n} L_i (y_{i,t+h} - \hat{y}_{i,t+h})$. Possible choices for $L_i(\cdot)$ are the linlin and linex loss functions. For the linlin loss function, if $y_{i,t+h}$ is conditionally normal, then the optimal predictor vector can be written as $\hat{y}_{i,t+h} = \mu_{i,t+h} + \sigma_{i,t+h} \cdot \Phi^{-1}(a_i/(a_i + b_i))$.

Proof: See Appendix.

### 3.2 Multivariate Variance Models:

A univariate GARCH model can be generalized to an $n$-dimensional multivariate GARCH model as $r_t | \Psi_{t-1} \sim N(0, \Sigma_t)$, where $r_t$ is the $n$-dimensional zero mean random variable,
$\Sigma_t$ is the variance covariance matrix that depend on information set available at $t-1$. $\Sigma_t$ depends on $q$ lagged values of squares and cross products of $r_t$ and $p$ lagged values of $\Sigma_t$.

The extension of a univariate GARCH model to a $n$-dimensional multivariate GARCH model require some restrictions on the conditional variance-covariance matrix $\Sigma$. There are different parameterizations of the variance covariance matrix. Among the most popular ones are the VEC, the diagonal representation (Engle, Granger, Kraft (1984), Bollerslev, Engle and Wooldridge(1988) and the BEKK model of Engle and Kroner (1995). The advantage of BEKK representation is that it is easy to impose restrictions on the conditional variance-covariance matrix (Engle, Kroner, (1995)) that ensures positive definiteness.

For the multivariate case I estimate the $n$-diagonal -BEKK multivariate GARCH(1,1) model. The BEKK variance model can be written as:

$$\Sigma_t = C'C + A' \varepsilon_{t-1} \varepsilon_{t-1}' A + B' \Sigma_{t-1} B.$$ 

By the specification of the model, the conditional variance-covariance matrix is guaranteed to be positive-definite. I forecast the one period ahead conditional variance-covariance matrix by iterating the matrix equation

$$\hat{\Sigma}_{t+1} = \hat{C}' \hat{C} + \hat{A}' \hat{\varepsilon}_{t-1} \hat{\varepsilon}_{t-1}' \hat{A} + \hat{B}' \hat{\Sigma}_{t-1} \hat{B}$$

and compute the values of the optimal predictor and the pseudo-optimal predictor for the multivariate LINLIN loss function over a rolling window.

4. Data

To determine whether using the optimal predictor in the presence of asymmetric loss is empirically useful, I use data representative of that used in financial forecasting. I estimate models and predict returns for three exchange rates between the U.S. Dollar and
the Canadian Dollar, the Japanese Yen and the British Pound. I also estimate models and predict returns for five major market indexes: the Dow Jones Industrial Average, and the S&P 500, the NASDAQ, the NIKKEI, and FTSE. The frequency of the data is weekly. The data run from January 1995, to August 6, 2006. The sample is used for estimation using a rolling window with a fixed length of one week and the one-step-ahead ($h=1$) and four-step-ahead ($h=4$) rolling forecasts are constructed.

5. Results

I first estimate n-GARCH (1,1), models for each return series over the whole sample using a rolling window with a fixed length of one week. Although the efficient market hypothesis suggests that returns should be serially uncorrelated, I check for possible serial correlation in the returns. I find that all the return series on exchange rates are serially uncorrelated. Table 1 shows the estimated GARCH (1,1) models for returns on three exchange rates and diagnostic statistics for the standardized residuals for the last window. The estimates for the conditional variance parameters are all significant. The skewness coefficients of the standardized residuals show that they are slightly negatively skewed. The kurtosis coefficients are three for Canada and U.K. and six for Yen/$ series. The Jarque-Bera statistics for Canada and UK. are not significant suggesting normality assumption can not be rejected while for the Yen/$ the Jarque-Bera statistic is significant suggesting that the data does not have normal distributions. The Ljung-Box Portmanteau tests for the serial correlation in the standardized and squared standardized residuals up to 10 lags indicate that the residuals are white noise. For the Yen/$ series I
re-estimated the model using t-GARCH(1,1) model since the Jarque-Bera statistic suggests that there is non-normality.

Table 2 shows the estimated GARCH models for the returns on indices and diagnostic statistics for the standardized residuals for the last rolling window. I find that the return series for S&P 500 are AR(1) in the mean. The estimates for the conditional variance parameters are all significant. For the Nasdaq series the estimated variance parameters, with the n-GARCH(1,1) model are \((\alpha, \beta) = (0.11, 0.89)\) and t-GARCH(1,1) model are \((\alpha, \beta) = (0.07, 0.93)\). The variance parameters sum to one, indicating that Nasdaq return series have an IGARCH property, which suggests that the unconditional variance for this series is not finite. Since the computation of pseudo-optimal predictor requires an expression for the unconditional variance, this series is dropped from the analysis. For the rest of the series the skewness coefficients of the standardized residuals show that they are slightly negatively skewed. The Jarque-Bera statistics are all highly significant. The Ljung-Box Portmanteau tests for the serial correlation in the standardized and squared standardized residuals up to 10 lags indicate that the residuals are white noise. In order to account for heavy tails and possible asymmetry in the conditional variance, I re-estimated the models using t-GARCH(1,1) and EGARCH(1,1) models. The results are reported in Table 2 for the last rolling window.

Given the models estimated above, for each rolling window with a fixed length of one week, I use the estimated variances from different models to compute the optimal, pseudo-optimal and conditional mean predictors and the losses associated with them for forecast horizons one and four which corresponds to weekly and monthly forecasting.
Figure 1 shows the ratio of the average losses between the optimal predictor and the conditional mean, and also the ratio of the average losses between the pseudo-optimal predictor and the conditional mean for each of the exchange rate return series from one-step-ahead forecasts. For all the series, it is striking that there is considerable gain by using the optimal or the pseudo optimal predictor versus the conditional mean predictor. The conditional mean predictor performs the worst, even for very low degrees of asymmetry. For the exchange rate series, with the exception of the Yen/$ series, the optimal predictor and the pseudo-optimal predictor perform almost the same suggesting there is no or very small gain using the optimal versus the pseudo optimal predictor, regardless of the variance model. For the Yen/$ series, with the exception of the results from non-parametric model, there is evidence of reduction in loss due to using the optimal predictor, the loss reduction is around 5% for asymmetry level of five and increases as the asymmetry level increases and is around 15% for the asymmetry level of 20.

Figure 2 shows the ratio of the average losses between the optimal predictor and the conditional mean and also the ratio of the average losses between the pseudo-optimal predictor and the conditional mean for the returns on the five market indices from the univariate variance models from one-step-ahead forecasts. Regardless of the series and the variance models being used, again the conditional mean predictor performs the worst. The optimal predictor generally out-performs the pseudo optimal predictor with the exception of Nikkei series, and as the degree of asymmetry increases, so does the loss reduction. The EGARCH(1,1) model seems to provide the highest values for the achieved loss reduction for all the models and the non-parametric model the worst and
the results from t-GARCH(1,1) model suggest slightly more loss reduction than the n-
GARCH(1,1) models. For the DJIA the results from series results from n-GARCH and t-
GARCH (1,1) suggest that the gain from using the optimal predictor is more than 15%
for moderate degrees of asymmetry (a=5) and is around 20% for higher degrees of
asymmetry. The results from the EGARCH model are further promising; 20% loss
reduction is achievable by using the optimal predictor for moderate asymmetry degrees
(five), and the loss reduction reaches to 35% as the asymmetry level increases. The
results from the nonparametric model suggest only 6% reduction even with the highest
level of asymmetry. For the FTSE, the results from the n-GARCH(1,1) and t-
GARCH(1,1) model suggests that there is more than 10% loss reduction when optimal
predictor is used, even for moderate degrees of asymmetry, and more than 20% loss
reduction is achievable when $a \geq 6$. The results from the EGARCH(1,1) model looks
more promising, suggesting up 26% loss reduction is possible using the optimal versus
the pseudo optimal predictor. The results from the nonparametric model, suggest only
2% loss reduction due to using the optimal predictor. The results on the Nikkei series
suggest that the pseudo-optimal predictor performs at least as good as the optimal
predictor, and some times slightly better. For the S&P 500 once again all the models
suggest that the optimal predictor out-performs the pseudo-optimal predictor. Results
from the parametric models imply a possible loss reduction of 20% or more even for
moderate degrees of asymmetry, reaching 30-35% for high degree of asymmetry. We
obtain the highest loss reduction value from the EGARCH(1,1) model, reaching 35% for
high asymmetry levels. This achieved loss reduction value is very close to the Monte
Carlo results of CD (1996). The nonparametric model suggests 1-5% loss reduction depending on the degree of asymmetry.

Figure 3 presents the results of the average loss ratios from 4-step-ahead forecasts using a n-GARCH(1,1) model for all the series considered. The optimal predictor generally outperforms the pseudo-optimal predictor with the exception of the $/£ series. The apparent loss reduction due to using the optimal predictor is interesting because when we use forecast horizon four, we are subject to more uncertainty since we are approximating the conditional distribution of the series as well as the conditional variance four periods ahead as forecast horizon increases we are likely to have larger forecast errors. Even under these circumstances it is interesting that the loss reduction due to using the optimal predictor is not negligible.

For the multivariate case I estimate both a normal and t-diagonal-BEKK multivariate GARCH(1,1) models for the Canadian, Japanese and U.K. FX return series. Consistent with the univariate normal case, the conditional mean predictor performs the worst and the optimal predictor performs slightly better than the pseudo-optimal predictor. The reduction in loss is around 4% (Figure 9) from optimal predictor when a diagonal BEKK GARCH(1,1) model is used with n-distributed errors for a high degree of asymmetric loss. This loss reduction, except for yen/$ series, is higher than the loss reduction the univariate variance models suggest. This is not a surprise because for a given series multivariate BEKK model exploits the information present in the rest of the series considered and makes use of a larger information set to estimate each variance, thus is more efficient then the univariate variance models. We obtain similar results from the diagonal t-BEKK GARCH(1,1) model but the loss reduction is slightly smaller.
6. Conclusion

I consider the mean losses associated with using the optimal predictor, pseudo-optimal predictor and the conditional mean predictor when agents have asymmetric loss function. I use both univariate and multivariate variance models. I also derive the optimal, and the pseudo-optimal predictor for additively separable multivariate loss function.

My results provide strong empirical evidence to the Granger (1969) regardless of the series and the variance models being used. For all series, loss associated with using the conditional mean predictor versus using the pseudo-optimal predictor is considerably higher even for moderate degrees of asymmetry, regardless of the variance model being used. The results suggest that under any kind of asymmetry the conditional predictor should not be used at all.

The results also provide empirical evidence to CD (1996, 1997); similar, but not as strong argument holds for the comparison between the optimal versus the pseudo-optimal predictor, as the results are sensitive to the series, variance parameters and asymmetry level being used. The loss reduction due to using the optimal predictor versus the pseudo-optimal predictor in general seems to be higher for the index return series, than the exchange rate series, and significantly large specially when EGARCH(1,1) model, that incorporates the “leverage effect”, is used, reaching up to 35%. This realized loss reduction from using the optimal versus the pseudo-optimal predictor is very close to the Monte-Carlo results of CD (1996) for expected losses at short horizons.
When we compare the realized loss reduction of all the series with the Monte-Carlo simulations of CD (1996), however, realized loss reduction is smaller than the expected. This is not a surprise, since in empirical work we are subject to parameter uncertainty. This also explains the reason why we get different results from different variance models. For the indices for example we see that, with the exception of Nikkei, the highest loss reduction is achieved by using the optimal predictor versus the pseudo-optimal predictor when we use the EGARCH(1,1) model, which captures the “leverage effect” and is around 35%, and generally the realized loss reduction from t-GARCH(1,1) models that account for the heavy tails are slightly larger than n-GARCH(1,1) models. While for the exchange rate series except for Yen/$ series, loss reduction from using the optimal predictor can be “statistically” small. In a few cases we also get, the pseudo-optimal predictor very slightly out-performing the optimal predictor. This might be due to the sampling variability or model misspecifications. However, the optimal predictor does tend to out perform the pseudo-optimal predictor in most of the cases. The difference can be very small 3-5% or as large as a 35% reduction in loss. Clearly, the results suggest the use of the optimal predictor, for the index series even for moderate degrees of asymmetry.

The results on exchange rates are more sensitive to the series, the variance parameters being used and the asymmetry level. The results in general suggest that using the optimal predictor that incorporates time varying second moments results in considerable loss reduction and can practically be important. The forecaster for should decide if the realized loss reduction is financially important for the agent as the use of the
optimal versus the pseudo-optimal predictor depends on the financial importance of the loss reduction to the individual.

Is this reduction in the mean loss financially important? A parallel argument to the difference between economic significance and statistical significance can be carried out here. Clearly, if you are a hedge fund manager, even 1% loss reduction would be of critical importance. However for different utility functions, or moderate degrees of asymmetry, agents might be indifferent between using the optimal or the pseudo-optimal predictor.
Table 1. Estimated n-GARCH(1,1) and t-GARCH(1,1) models for returns on exchange rates for the last rolling window.

<table>
<thead>
<tr>
<th>Model</th>
<th>parameters</th>
<th>FX Canada</th>
<th>FX Japan</th>
<th>FX U.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>psi</td>
<td>0.014</td>
<td>0.049</td>
<td>0.063</td>
</tr>
<tr>
<td>n-GARCH(1,1)</td>
<td></td>
<td>(0.007)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td></td>
<td>alpha</td>
<td>0.034</td>
<td>0.059</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.016)</td>
<td>(0.03)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>beta</td>
<td>0.94</td>
<td>0.921</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>-0.029</td>
<td>-0.832</td>
<td>-0.127</td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td>3.318</td>
<td>5.98</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td>Jaque-Bera</td>
<td>2.274</td>
<td>254.32</td>
<td>1.667</td>
</tr>
<tr>
<td></td>
<td>Q(10)</td>
<td>3.94</td>
<td>8.547</td>
<td>12.96</td>
</tr>
<tr>
<td></td>
<td>Q^2(10)</td>
<td>9.4</td>
<td>9.1317</td>
<td>5.274</td>
</tr>
<tr>
<td>t-GARCH(1,1)</td>
<td>psi</td>
<td>0.067</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.041)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>alpha</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>beta</td>
<td>0.937</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nu</td>
<td>7.239</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.759)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>-0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td>7.073</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jaque-Bera</td>
<td>442.807</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q(10)</td>
<td>9.097</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q^2(10)</td>
<td>9.978</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Estimated n-GARCH(1,1) and t-GARCH(1,1) models for returns on market indexes.

<table>
<thead>
<tr>
<th>Model</th>
<th>parameters</th>
<th>Dow-Jones</th>
<th>FTSE</th>
<th>NIKKEI</th>
<th>S&amp;P 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-GARCH(1,1)</td>
<td>$\psi$</td>
<td>.696</td>
<td>.166</td>
<td>.252</td>
<td>.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.276)</td>
<td>(.064)</td>
<td>(.259)</td>
<td>(.053)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>.199</td>
<td>.143</td>
<td>.0325</td>
<td>.091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.101)</td>
<td>(.041)</td>
<td>(.019)</td>
<td>(.03)</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.690</td>
<td>.833</td>
<td>.939</td>
<td>.898</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.109)</td>
<td>(.04)</td>
<td>(.044)</td>
<td>(.03)</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>-.509</td>
<td>-.443</td>
<td>-.209</td>
<td>-.387</td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td>4.768</td>
<td>3.632</td>
<td>3.67</td>
<td>3.822</td>
</tr>
<tr>
<td></td>
<td>Jarque-Bera</td>
<td>90.621</td>
<td>25.827</td>
<td>13.64</td>
<td>27.756</td>
</tr>
<tr>
<td></td>
<td>$Q$ (10)</td>
<td>11.930</td>
<td>1.645</td>
<td>3.81</td>
<td>9.947</td>
</tr>
<tr>
<td></td>
<td>$Q^2$ (10)</td>
<td>2.248</td>
<td>8.837</td>
<td>9.404</td>
<td>6.265</td>
</tr>
<tr>
<td>t-GARCH(1,1)</td>
<td>$\psi$</td>
<td>.540</td>
<td>.165</td>
<td>.195</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.28)</td>
<td>(.09)</td>
<td>(.183)</td>
<td>(.055)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>.147</td>
<td>.134</td>
<td>.034</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.05)</td>
<td>(.039)</td>
<td>(.02)</td>
<td>(.021)</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.759</td>
<td>.841</td>
<td>.945</td>
<td>.932</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>8.0402</td>
<td>11.559</td>
<td>13.11</td>
<td>10.137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.307)</td>
<td>(5.417)</td>
<td>(5.123)</td>
<td>(4.074)</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>-.571</td>
<td>-.441</td>
<td>-.320</td>
<td>-.388</td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td>5.042</td>
<td>3.628</td>
<td>4.24</td>
<td>3.985</td>
</tr>
<tr>
<td></td>
<td>Jarque-Bera</td>
<td>119.262</td>
<td>25.53</td>
<td>42.446</td>
<td>34.296</td>
</tr>
<tr>
<td></td>
<td>$Q$ (10)</td>
<td>11.971</td>
<td>1.649</td>
<td>3.997</td>
<td>9.076</td>
</tr>
<tr>
<td></td>
<td>$Q^2$ (10)</td>
<td>2.734</td>
<td>8.867</td>
<td>5.215</td>
<td>7.252</td>
</tr>
<tr>
<td>EGARCH(1,1)</td>
<td>$\psi$</td>
<td>-.008</td>
<td>-.074</td>
<td>.373</td>
<td>-.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.053)</td>
<td>(.050)</td>
<td>(.172)</td>
<td>(.042)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>.158</td>
<td>.181</td>
<td>.015</td>
<td>.154</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.061)</td>
<td>(.062)</td>
<td>(.05)</td>
<td>(.06)</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.926</td>
<td>.953</td>
<td>.824</td>
<td>.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.029)</td>
<td>(.018)</td>
<td>(.086)</td>
<td>(.02)</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>-.178</td>
<td>-.1147</td>
<td>-.12</td>
<td>-.136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.039)</td>
<td>(.032)</td>
<td>(.042)</td>
<td>(.03)</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>-.467</td>
<td>-.428</td>
<td>-.183</td>
<td>-.389</td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td>3.84</td>
<td>3.449</td>
<td>3.609</td>
<td>3.696</td>
</tr>
<tr>
<td></td>
<td>Jarque-Bera</td>
<td>34.664</td>
<td>20.373</td>
<td>11.0252</td>
<td>23.607</td>
</tr>
<tr>
<td></td>
<td>$Q$ (10)</td>
<td>12.290</td>
<td>2.124</td>
<td>3.780</td>
<td>9.58</td>
</tr>
<tr>
<td></td>
<td>$Q^2$ (10)</td>
<td>2.87</td>
<td>8.683</td>
<td>4.74</td>
<td>4.196</td>
</tr>
</tbody>
</table>
Figure 1.
Ratio of rolling one-step-ahead average losses from different variance models for exchange rate series.
Figure 2.
Ratio of rolling one-step-ahead average losses from different variance models for index returns.
Figure 3.
Ratio of rolling four-steps-ahead average losses from n-GARCH(1,1) models for index and exchange rate return series.
Figure 4. Ratio of Average Losses for exchange rates, n-BEKK GARCH(1,1).

Figure 5. Ratio of Average Losses for exchange rates, t-BEKK GARCH(1,1).
Appendix:

A.1. 

**Proof:** The optimal predictor $\hat{Y}_{t+h}$ minimizes the expected loss

$$E_t[L(Y_{t+h} - \hat{Y}_{t+h})] = \int_{-\infty}^{\infty} L(Y_{t+h} - \hat{Y}_{t+h}) \phi(S_t^{-1/2} (Y_{t+h} - \mu_{t+h})) dY_{t+h}$$

where $\phi()$ denotes the multivariate standard normal density and the integral sign denotes an $n$-fold integral over the elements of $Y_{t+h}$. Let $X_{t+h} = Y_{t+h} - \mu_{t+h}$ denote the observations deviation from its conditional mean. Changing variables, where determinant of the Jacobian of the transformation is one, the objective function can be expressed in deviation from mean as

$$E_t[L(X_{t+h} - \alpha_{t+h})] = \int_{-\infty}^{\infty} L(X_{t+h} - \alpha_{t+h}) \phi(S_t^{-1/2} X_{t+h} | \Omega_t) dX_{t+h}$$

where $\alpha_{t+h} = \hat{Y}_{t+h} - \mu_{t+h}$ is chosen to be the optimal predictor of $X_{t+h}$. The objective function does depend on the conditional mean $\mu_{t+h}$, and therefore, the optimal predictor only depends on the loss function $L()$ and the conditional variance-covariance matrix $\Sigma_t$. Given that $\alpha_{t+h}$ is the optimal predictor of $X_{t+h}$, the optimal predictor of $Y_{t+h}$ is $\mu_{t+h} + \alpha_{t+h}$.

A.2. 

The first order conditions are

$$\frac{\partial E[L(Y_{t+h} - \hat{Y}_{t+h})]}{\partial \hat{Y}_{t+h}} = b_i f_i(Y_{t+h}) dy_{t+h} - a_i \int_{-\infty}^{\infty} f_i(Y_{t+h}) dy_{t+h} = 0 \quad i = 1, ..., n$$

or $b_i F_i(\hat{Y}_{t+h}) - a_i [1 - F_i(\hat{Y}_{t+h})] = 0$, which can be solved as $\hat{Y}_{t+h} = F^{-1}(a_i/(a_i + b_i)).$

If $Y_{t+h}$ is conditionally normal, then $F_i(Y_{t+h}) = \Phi((Y_{t+h} - \mu_{i,t+h})/\sigma_{i,t+h})$ and the optimal predictor vector can be written as $\hat{Y}_{i,t+h} = \mu_{i,t+h} + \sigma_{i,t+h} \Phi^{-1}(a_i/(a_i + b_i))$. 

26
References:


However, as CD (1997) point out, the conditionally Gaussian assumption can be relaxed. The optimal predictor is obtained by substituting the appropriate conditional CDF.

However, as CD (1997) point out, the conditionally Gaussian assumption can be relaxed. The optimal predictor is obtained by substituting the appropriate conditional CDF.

CD (1996) finds about 35% loss reduction in conditionally expected loss for short horizons, from using the optimal versus the pseudo-optimal predictors.