Numerical Methods in Q. Fin. - Model Risk

Janos C. R. Füting

March 21st, 2024

Outline

What is model risk?

Knightian Uncertainty

The notion of Knightian uncertainty was introduced in economics by Frank Knight in 1921. We distinguish risk and uncertainty as follows:

- **Risk** is what we know we don't know
- **Uncertainty** is what we don't know we don't know
- **Risk** we can measure, **Uncertainty** we can not

All models are wrong, but some are useful.
- George Box

Risk and Uncertainty in Statistics

1. Parameter estimates are at the minimum subject to sampling error (measurable)
2. We may select the wrong model among our set of candidates (measurable)
3. The true model may not be among our set of candidates (immeasurable, true uncertainty)

Model Risk literature focuses on the second point, the possible risk arrising due to choosing a specific model among a selection of available models, but does not deal with true model uncertainty (in general).

Measuring Model Risk

Recall: Definition of a Risk Measure

Let X a r.v. on $(\Omega, \mathcal{F}, P)$ with distribution function $F_X(x) = P(X \leq x)$.

Definition

A ***risk measure*** is a mapping $\rho : \mathcal{L}_\rho \to \mathbb{R}$ defined on some space of random variables $\mathcal{L}_\rho$ and satisfying the following properties:

- *law invariance:* $\rho(X) = \rho(Y)$ whenever $F_X \equiv F_Y$
- *positive homogeneity:* $\rho(aX) = a\rho(X)$ for any $a \geq 0$
- *translation invariance:* $\rho(X + b) = \rho(X) - b$ for any $b \in \mathbb{R}$

Barrieu & Scandolo Approach

Let our model of interest be the one that gives us the distibution of our portfolio $X_0 \in \mathcal{L} \subset \mathcal{L}_\rho$ and define

$$\underline{\rho}(\mathcal{L}) = \inf_{X \in \mathcal{L}} \rho(X) \qquad \bar{\rho}(\mathcal{L}) = \sup_{X \in \mathcal{L}} \rho(X)$$

Definition

The ***absolute measure of model risk*** is given by

$$AM = AM(X_0, \mathcal{L}) = \frac{\bar{\rho}(\mathcal{L})}{\rho(X_0)} - 1$$

and the ***relative measure of model risk*** is given by

$$RM = RM(X_0, \mathcal{L}) = \frac{\bar{\rho}(\mathcal{L}) - \rho(X_0)}{\bar{\rho}(\mathcal{L}) - \underline{\rho}(\mathcal{L})}$$

In the special case of the risk measure of interest beeing the VaR or ES for some specified level of probability α, we can derive some general bounds that rely only on the assumption of a distbution with finite variance. Illustrated here is the case of standardised variables:

$$VaR_\alpha(X_0) = 1 \Bigg/ \sqrt{\frac{1-\alpha}{\alpha}} \qquad \Rightarrow \qquad AM(X_0, \mathcal{L}) = \frac{\sqrt{\frac{1-\alpha}{\alpha}}}{VaR_\alpha(X_0)}$$

$$\sup ES_\alpha(X_0) = 1 \Bigg/ \sqrt{\frac{1-\alpha}{\alpha}} \qquad \Rightarrow \qquad AM(X_0, \mathcal{L}) = \frac{\sqrt{\frac{1-\alpha}{\alpha}}}{ES_\alpha(X_0)}$$

Alternative Approaches

Nothing in the previous definitions is particularly natural, these measures are heuristic (especially in practical implementation), so naturally there are many competing measures:

$$M_1 = \bar{\rho}(\mathcal{L}) - \underline{\rho}(\mathcal{L})$$

$$M_2 = \bar{\rho}(\mathcal{L}) - \rho(X_0)$$

$$M_3 = \rho(X_0) - \underline{\rho}(\mathcal{L})$$

$$M_4 = \mu\left(|\rho(X) - \rho(X_0)|^p\right)^{\frac{1}{p}}$$

$$M_5 = \frac{\bar{\rho}(\mathcal{L}) - \rho(X_0)}{\bar{\rho}(\mathcal{L})}$$

$$M_6 = \frac{\rho(X_0) - \underline{\rho}(\mathcal{L})}{\underline{\rho}(\mathcal{L})}$$

$$M_7 = \frac{\bar{\rho}(\mathcal{L})}{\underline{\rho}(\mathcal{L})}$$

$$M_8 = \frac{\mu(|\rho(X) - \mu(\rho(X))|)}{\mu(\rho(X))}$$

$$M_9 = \mu\left((\rho(X) - \rho(X_0))_{\pm}^{p}\right)^{\frac{1}{p}}$$

$$M_{10} = ES_{1-\alpha}(X - \rho(X_0))$$

$$M_{11} = \rho^{\star}(f(X - \rho(X)))$$

$\rho^{\star}$ some RM, f absolute value, positive part, etc.

Assessing Model Fit in Practice

PP and QQ Plots

- Plots of model-implied versus actual probabilities and quantiles are simple yet powerfull tools to asses wether our model describes the data well.
- They can be used both for unconditional distributions, or by calculating the conditional proability integral transform and reprojecting into margins that have similar characteristics as the conditional distributions.

PP and QQ Plots

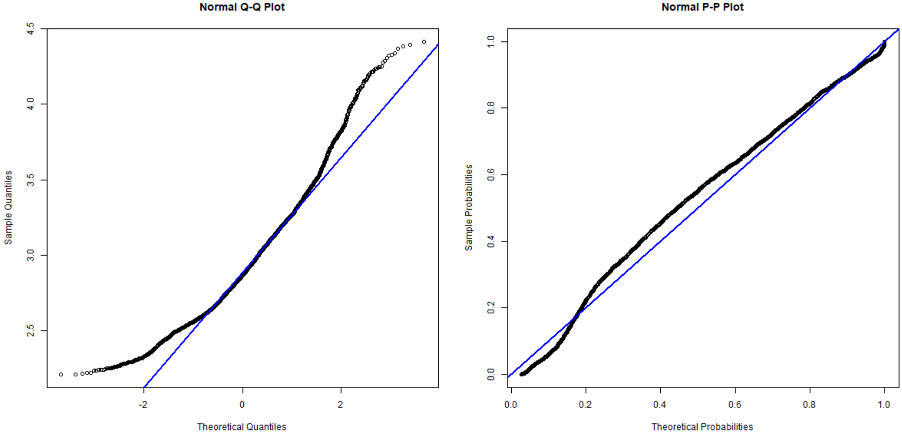

Figure: Normal-PP and -QQ plots of log of VIX closing values, 2007-01-03/2023-03-18.

Distribution Tests

The idea of comparing distributions can be made more formal, commonly used:

- The Komogorov-Smirnov test with its statistic $\sup_x |F_{data}(x) - F_{model}(x)|$ tests the equivalence of the data disribution and the model implied distribution directly.
- We can use the model implied (condtitional) CDF to perform a Jarque-Berra test on $\Phi^{-1}(F_{model,t}(x_t))$. This has better power than directly testing $F_{model}(x)$ for uniformity.
- In a discrete setting (or using a binning from a continous distribution) we can use a χ^2 test:

$$\sum \frac{(N_i - Np_i)^2}{Np_i} \sim \chi^2$$

where N is the sample size, and N_i and p_i are the bin count and probability respectively.

Assessing Residuals

Especially but not only in a timeseries context it is important to not only compare unconditional fit but also to assess the behaviour of residuals:

- correlation of squared residuals with covariates/time indicate heteroskedasticity
- improper modelling may lead to (squared) residuals that are correlated over time
- residuals with non-finite second/fourth moments are can mean that the fitting process is unreliable

Special Case: VaR Forecasts

A $\alpha\%$-VaR gives rise to a $\alpha\%$-confidence interval intervall and exceeding the VaR can be thought of as a Bernoulli variable with parameter α .

The procedure for testing is simple:

1. In a rolling or expanding window, estimate your model and generate h-step ahead forecasts for the period following your estimation window.
2. Compare the behaviour of your VaR forecasts with your realised variables:
   - do they get "hit" α percent of the time?
   - is the hit/non-hit classification consistient with the hypothesis that they are Bernoulli with probability α?
   - are hits equally spaced over time?

Mitigating Model Risk

Mitigating Model Risk

The basic idea underlying the many appraoches to model risk mitigation is to combine models and/or their oputput, instead of relying only on information from a single "best" model.
There are a variety of approaches to dealing with model risk and improving accuracy that can be roughly categorised among three axes:

- point, interval, or density combination
- combining models vs combining forecasts
- frequentist vs Bayesian methods

The methods go by many names with sometimes slightly varying approaches depending on the field doing the research:
forecast combination, model averaging, opinion pooling, ensemble methods, etc

The "forecast combination puzzle"

A name given to the fact that very simple heuristics for selecting forecast weights, especially the equal weighting heuristic which completely disregards information on past forecaster performance and correlation of errors, tends to not be outperformed by sophisticated approaches that should be (assymptotically) superior.

Point Forecast Combination

Heuristics

There is a large literature on the combination of point forecasts, a lot of which focus on the effectiveness of simple heuristics:

- mean forecast
- median or mode forecast, to be more stable to outliers
- trimmed or winsorised mean forecast, similarly for stability

but with no clear consensus on which heuristic performs best.

While conceptionally and numerically very simple in their implementation, some care has to be taken in selecting the pool of the forecasts. Including forecasts of low accuracy or having insufficient variability in forecasts can hurt performance.

Initial proposals for ”optimal” weights included

$$w_i^{opt} = \frac{\Sigma_{i,i}^{-1}}{\sum_j \Sigma_{j,j}^{-1}}$$

and

$$w^{opt} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$$

but these can fare poorly in practice depending on the loss function we care about and the joint distribution of forecasts, so the literature quickly moved on to regression based approaches.
Nevertheless there is a bunch of other such ”performance based” approaches with decent empirical sucess.

When moving to a regression framework there are three choice for the model equation:

$$
\begin{aligned}
y_t &= \mathbf{w}^T\mathbf{f} + \varepsilon \qquad s.t. \quad \mathbf{w}^T\mathbf{1} = 1 \\
y_t &= \mathbf{w}^T\mathbf{f} + \varepsilon \\
y_t &= w_0 + \mathbf{w}^T\mathbf{f} + \varepsilon
\end{aligned}
$$

optionally with the constraint $w_i \geq 0$. There is an active debate about which model should be used.

Once the model is determined, we have various choices of penalty functions to minimise, such as:

- $\sum(y_t - \omega f_t)^2$ (OLS)
- $\sum(y_t - \omega f_t)^2 + \lambda|\omega|_p$ (LASSO and Ridge)
- $\sum|y_t - \omega f_t|$ (median regression)
- and many more (e.g. more general quantile regression)

To solve the regression problem we either solve the *normal equation* of the form

$$\begin{aligned} X^T X\mathbf{w} &= X^T\mathbf{y} \qquad \text{OLS} \\ (X^T X + \lambda\mathbf{I})\mathbf{w} &= X^T\mathbf{y} \qquad \text{Ridge} \end{aligned}$$

or optimise directly using gradient descent (e.g. LASSO)

$$\nabla_w \ell(w) = \lambda \text{sign}(w) + \nabla_w l(w)$$

with l the quadratic loss function.

A few points of caution when relying on standard software implementations of linear regression:

- The analytical solution for the normal equation involves inversion and can be instable since forecasts for the same target tend to be correlated (and the better the forecasts the more correlated). Hence it may be preferable numerical to solve the normal equation using e.g. a least squares solver.
- Another common approach to deal with the correlation between forecasts is to perform a principle component decomposition and then treat the components as the forecasts in the regression problem, empirically this outperforms regular OLS, is simple to implement, and robust to multicolinearity and $N > T$ situations.

Example Application: Comining Survey Forecasts

Macroeconomic Forecasts - Intro

The Survey of Professional Forecastsers is a quarterly survey of forecasters in industry and academia conducted by the Federal Reserve. It aims to forecast several macroeconomic variables at a quarterly frequency:

- interest rates
- market returns
- economic growth
- inflation, etc. etc.

The participants are typically asked for point and interval probability forecasts that are aggregated by the FED themselves, but the history of individual forecasts is also available. As a matter of fact aggregating exactly this type of forecast has been one of the motivating example of the literature.

Table: Forecast combinations of SPF next quarter inflation forecasts using rolling window with length 10, 20, 30, 40 quarters. Data from Q3 1981 to Q1 2024. Estimators fall back to naive if encountering numerical infeasability.

	avg. err.	MSE	MAE	autocorr	mean error	MSE	MAE	autocorr
naive	-0.04	9.46	2.01	-0.25	0.05	9.75	2.08	-0.22
opt_i	-0.04	9.46	2.01	-0.25	0.05	9.75	2.08	-0.22
opt	-18.59	5e+4	20.66	-0.01	0.07	9.81	2.10	-0.22
OLS+0	-506.78	3e+7	565.11	0.40	-204.64	1e+6	224.07	0.62
OLS	-5e+5	5e11	7e+4	-0.01	2188.87	4e+9	1e+4	-0.47
naive	0.78	9.75	2.34	0.03	1.39	9.96	2.46	0.12
opt_i	0.78	9.75	2.34	0.03	1.39	9.96	2.46	0.12
opt	0.78	9.76	2.35	0.03	1.39	9.96	2.46	0.12
OLS+0	-158.41	8e+4	168.42	0.74	-104.02	7e+5	112.70	0.81
OLS	624.58	9e+7	781.95	0.72	316.26	3e+7	450.98	0.67

Among non-linear methods, methods based on machine learning have been shown to beat linear combinations. The following approaches can be found in the literature:

- GAM type models $y = \beta_0 + \sum \beta_j \hat{y}_j + \sum \delta_i g_i(\gamma_i \hat{y}_i)$, where we try to estimate the the non-linear mapping g_i along with the coefficients.
- Stacked Generalisation / "stacking", where multiple models are trained on a singular timeseries, potentially by training the same architecture on folds of the data, and then aggregated by means of a "meta"-model.
- Cross learning methods, where models and meta-models are trained using information from multiple timeseries. Potentially by first extracting series of features from the historical data and then training the models on the data and the extracted features.

Probabilistic Combination - Score Function Approach

When combining densities the problem becomes less straightforward and we have to intodoce the concept of a "score" to replace the loss function while respecting the fact that our forecasts are probabilistic.

A score is a function that tells us how good/bad our probabilistic forecasts have been doing, popular

- log score : $L(p, y) = -\log\left(\sum p_m \mathbb{I}_{y \in b_m}\right)$
- Brier score : $B(p, y) = \frac{1}{M} \sum (p_m - \mathbb{I}_{y \in b_m})^2$
- quadratic score : $Q(p, y) = -2\left(\sum a p_m \mathbb{I}_{y \in b_m}\right) + \left(\sum p_m^2\right)$
- Ranked score : $R(p, y) = \sum (P_m - \mathbb{I}_{y \leq b_{m+}})^2$

the problem then simply becomes a numerial optimisation with target $\omega : p_m = \omega f_t$ s.t. $\sum \omega = 1$, and we can impose additional penalties for deviation from equal weights.

Probabilistic Combination - Continously Ranked Probability Score

Most score functions are relatively agnostic to the part of the distribution we are particularly interest in. But as some applications (financial risk management in particular) are very sensitive to specific parts of the forecast distribution while beeing very in-sensitive to others, it is advantagous to consider a scoring rule that reflects The Continously Ranked Probability Score (CRPS) is given by

$$CRPS = \int_{-\infty}^{\infty} S(F(z), \mathbb{I}_{y \leq b_{m+}})u(z)\mathrm{d}z$$

where S is the quadratic or Brier score we saw before.
Here $u(z)$ is a weight function that we choose so as to put more weight on regions that interest us and less on regions that don't. The simplest choice beeing an indicator function.

Probabilistic Combination - Bayesian Model Averaging

- The basic idea is to, instead of just having priors over parameters as in the usual Bayesian setup, also put priors over the model space.
- if models are nested this is particularly easy since we can encode model space as a vector in $\{0,1\}^p$ with p beeing the numer of parameters.
- We can incorporate the various models directly when sampling our posterior predictive distribution.
- The (philosophical) problem is that with growing sample size BMA converges to one model as the distribution over the model space degenerates.

Probabilistic Combination - Nonlinear Pooling

Problems with Linear Pooling

Linear pooling of probabilistic forecasts has some undesirable consequences:

- a linear opinion pool lacks *calibration* (i.e. a forecast probability is not a consistent estimator for the true probability, even if the individual forecasts are)
- a linear opinion pool lacks *sharpness* (i.e. the pooled forecast is overly variable and underconfident)

One strategy to deal with the sharpness issue is to pool over quantiles, which results in a distribution with the same mean as the probability pooling but is always sharper, unimodal, and closed with respect to location and scale families.
Another strategy are non-linear pools.

A generalised linear pool takes the form

$$F(y|I) = g^{-1}\left(\sum w_i g(F_i(y|I))\right), \quad w_i > 0, \quad \sum w_i = 1$$

where g is a continous, strictily monotonic, invertible function. Examples for g are

- $g(x) = x$ (classical linear pool)
- $g(x) = 1/x$ (harmonic pool)
- $g(x) = \log(x)$ (geometric pool)

A Beta-Transformed Linear Pool (BLP) takes the form

$$F(y|I) = B_{\alpha,\beta}\left(\sum_i w_i F_i(y|I)\right), \quad w_i > 0, \quad \sum_i w_i = 1$$

with $B_{\alpha,\beta}$ the CDF of the beta distribution.

While at first estimation may seem rather complicated, this is a generalised linear model and can be estimated as such.

Thank you for your attention!

References

- Barrieu & Scandolo (2013) - Asessing Financial Model Risk
- Müller & Righi (2019) - Model Risk Measures: A Review and New Proposals on Risk Forecasting
- Diebold et al. (2023) - On the aggregation of probability assessments: Regularized densities for Eurozone inflation and real interest rates
- Wang et al. (2022) - Forecast combinations: an over 50-year review