Real-World Data

MFM Practitioner Module: Risk & Asset Allocation

John Dodson

October 19, 2011

Outline

Robustness

Non-Parametric Estimators

The term robustness in statistics can sometimes refer to non-parametric techniques that do not require assumptions about the characterization of the random variables involved.

- Such techniques usually lean on the Law of Large Numbers, and hence require very large samples to be effective.

Robust Estimators

A more precise meaning has evolved that focuses on estimators that may be based on parametric characterizations, but which can produce reasonable results for data that does not come from that class of characterizations or stress-test distributions.

- We can make this desire concrete in term of the the influence function associated with an estimator.

Robust Estimators

Influence Function

Previously, we have discussed estimators as functions of samples. If instead we consider the estimator as a functional of the density from which the sample is drawn, we can consider its (functional) derivative with respect to an infinitesimal perturbation in the density given by

$$f_X(x) \to (1-\epsilon) f_X(x) + \epsilon\delta(x-y)$$

Thus, with $\tilde{\theta}$ the functional induced by the estimator $\hat{\theta}$,

$$\mathrm{IF}\left[y, f_X, \hat{\theta}\right] = \lim_{\epsilon \to 0} \frac{\tilde{\theta}\left[(1-\epsilon) f_X(x) + \epsilon\delta(x-y)\right] - \tilde{\theta}\left[f_X\right]}{\epsilon}$$

If this derivative is bounded for all possible displacements, y, we say the estimator is robust.

Robust Estimators

Robustness of the MLE

For the maximum likelihood estimator, the influence function turns out to be proportional to

$$\mathrm{IF}\left[y, f_X, \hat{\theta}\right] \propto \left. \frac{\partial \log f_{X|\theta}(y)}{\partial \theta} \right|_{\theta=\hat{\theta}}$$

For some characterizations, the parameter MLE's are robust. For some they are not.

- for $X \sim \mathcal{N}(\mu, \Sigma)$, $\hat{\mu}$ and $\hat{\Sigma}$ are not robust
- for $X \sim \mathrm{Ca}(\mu, \Sigma)$, they are

Meucci demonstrates that, even for the empirical characterization, the influence functions for the sample mean and the sample covariance are not bounded; therefore these sample estimators are *never* robust.

M-Estimators

Location and Dispersion

Recall the general elliptic location and dispersion MLE's from week 5,

$$\hat{\mu} = \sum_{i=1}^{N} \frac{w_i}{\sum_j w_j} x_i$$

$$\hat{\Sigma} = \sum_{i=1}^{N} \frac{w_i}{N} (x_i - \hat{\mu})(x_i - \hat{\mu})' \quad \text{with}$$

$$w_i \triangleq h\left(\text{Ma}^2\left(x_i, \hat{\mu}, \hat{\Sigma}\right)\right) \quad \forall i = 1, \ldots, N$$

where the function $h(\cdot)$ is the value of a particular functional on the density. The idea with M-estimators is to choose $h(\cdot)$ exogenously in order to bound the influence function by design.

M-Estimators

We know that $h(\cdot) = 1$ corresponds to the MLE for normals and also to the sample estimators, which do not have bounded influence functions. A weighting function that falls towards zero for large arguments is more likely to be robust. Some examples include

- Trimmed estimators for which
$$h(z) = \begin{cases} 1 & z < z_0 = Q_{\chi^2_{\dim X}}(p) \\ 0 & \text{otherwise} \end{cases}$$
- Cauchy estimators for which $h(z) = \frac{1+\dim X}{1+z}$
- Schemes such as Huber's or Hampel's for which
$$h(z) = \begin{cases} 1 & z < z_0 = \left(\sqrt{2} + \sqrt{\dim X}\right)^2 \\ \sqrt{\frac{z_0}{z}} e^{-\frac{(\sqrt{z}-\sqrt{z_0})^2}{2b^2}} & \text{otherwise} \end{cases}$$

These estimators can be evaluated numerically by iterating to the fixed point.

Missing Data

Sometimes data for one or more securities for a particular historical horizon is missing or suspect. Rather than removing the entire period from your dataset, it might make sense to try to estimate around these holes. In the MLE setting, the established approach to this is called the EM algorithm.

EM alogrithm

This is based on replacing the objective function with a conditional expectation in an iterative scheme.

$$\hat{\theta}^{(j+1)} = \arg\max_{\theta} \mathrm{E}_{Y_{\text{mis}}^{(N)} | y^{(N)}, \hat{\theta}^{(j)}} \left[\log L \left(\theta, y^{(N)} + Y_{\text{mis}}^{(N)} \right) \right]$$

For the normal, this update equation has a tractable form. See §4.6.2 for details.

Missing Data

Notice that in the EM algorithm, we effectively treat the dataset as a combination of valid observations, $y^{(N)}$, and random variables, $Y^{(N)}_{\text{mis}}$, representing possible values for the missing observations.

Bridge models

Once we are comfortable with the results of the iteration, $\hat{\theta}$, we can use this to draw any number of "samples of the sample" if we need to work with a complete dataset.

- While conditional expectation features in the parameter estimation, it is not advisable to use the conditional expectation for re-sampling, because that could cause one to subsequently under-estimate (co)variances.

We will re-visit re-sampling schemes next semester.

Overlapping Data

Market data can be be overlapped for several reasons.

- Stale quotations erroneously carry forward old information
- Official closings may not be synchronous across markets and time zones
- You may *choose* to overlap your data because your investment horizon is longer than your sampling period

In all cases, your data are no longer i.i.d.

- standard errors based naïvely on Fisher information will be too small
- estimated correlations will be biased towards zero compared to instantaneous correlations

N.B.: The correlations can be fixed by dividing them by the fraction of the period that was overlapped.

Other Techniques

Rolling Window

As you revise your estimates, you should consider dropping your oldest observations as you add new ones, so that the size of your dataset remains constant.

Exponential Decay

If you think your parameters might be slowly varying, you can emulate a simple Kalman filter by using a weight scheme of the form $w_i = \frac{1-\lambda}{1-\lambda^N}\lambda^{i-1}$ for $0 < \lambda < 1$.

Zero Drift

An extreme version of shrinkage which can be relevant to finance is to assume that the (objective) drift of all or some asset value processes is zero. We will need to amend this assumption in any case once we gathered the manager views.

Implied Distribution

At-the-money (European) implied volatility from the options market can be interpreted as the market consensus about the variance of future returns *under the assumption of log-normality*.

Breeden-Litzenberger

But if you have implied volatilities for a range of strikes (or rather the call premia c and a discount factor d), you can use the result

$$f^{\mathbb{Q}}_{S_T|\mathcal{F}_t}(K) = \frac{1}{d_t(T)} \frac{\partial^2 c_t(T,K)}{\partial K^2}$$

to make a much for comprehensive statement about the (risk-netrual) implied distribution.

Implied Distribution

This result gives us a recipe for the (risk-neutral) expected value of any function of the terminal underlying price $h(\cdot)$.

$$\mathrm{E}^{\mathbb{Q}}\left[h(S_T)|\mathcal{F}_t\right] = h\left(F_t(T)\right) + \int_0^\infty h''(K)\left(\frac{c_t(T,K)}{d_t(T)} - (F_t(T)-K)^+\right)\,dK$$

where $F_t(T) = c_t(T,0)/d_t(T)$ is the forward price of the underlying.

Implied Variance

One application of this result is the (risk-neutral) implied price variance.

$$\mathrm{var}^{\mathbb{Q}}\left[S_T|\mathcal{F}_t\right] = 2\int_0^\infty \frac{c_t(T,K)}{d_t(T)}\,dK - F_t(T)^2$$