2020 March 16

Scribed by Roh

hyunwoo@uchicago.edu

What makes an asset useful?

1 Paper Review (Continued)

1.1 Differential Mutual Information Timescale as a Measure of Incremental Diversification

To motivate our measure of incremental diversification, we start with the simple case of two assets π and A.

Case 1: Two assets, i.i.d Gaussian

Returns of π respect to A are assumed to be independent draws across time from the same distribution r_π respect to r_A with mean μ and variance σ^2. We assume that (r_π, r_A) is jointly Gaussian and the corr between r_π and r_A is ρ . As previously discussed the lower the correlation, the higher the return per unit of risk we can obtain by combining π and A .

Given that Gaussian random variables are fully determined by their first two moments, the uncertainty remaining in r_A after knowing r_π is well characterized by the conditional variance $\mathrm{Var}(r_A \mid r_\pi)$,which in this case is easily shown to be

$$\mathrm{Var}(r_A \mid r_\pi) = \sigma^2(1-\rho^2) \tag{1}$$

Recall that The one-period returns of the new portfolio are easily found to be independent draws across time from the random variable

$$r_{\pi'} = wr_\pi + (1-w)r_A$$

Then,

$$\begin{aligned}
\mathrm{Var}(r_{\pi'}) &= \mathrm{Var}\left(wr_\pi + (1-w)r_A\right) \\
&= \mathrm{Var}(wr_\pi) + \mathrm{Var}\left((1-w)r_A\right) + \mathrm{Cov}\left(wr_\pi, (1-w)r_A\right) \\
&= \sigma^2\left(w^2 + (1-w)^2\right) + 2\rho\sigma^2\left(w(1-w)\right) \\
&= \sigma^2\left(w^2 + 1 - 2w + w^2 + 2\rho w - 2\rho w^2\right) \\
&= \sigma^2\left(1 - 2w(1-w)(1-\rho)\right) \qquad (2)
\end{aligned}$$

When $\rho \geq 0$, $\mathrm{Var}(r_{\pi'})$ (in Equation 1) increases with ρ, whereas $\mathrm{Var}(r_A \mid r_\pi)$ decreases with ρ, and the two requirements for diversification, namely high expected return per unit of risk and unrelated returns, are consistent. However, when $\rho < 0$, $\mathrm{Var}(r_{\pi'})$ still increases as a function of ρ but $\mathrm{Var}(r_A \mid r_\pi)$ now also increases with ρ. In other words, when $\rho < 0$, the potential to increase to return per unit of risk by combining π and A increases as the correlation decreases, but π **and** A **shares more underlying driving factors.** It might seem as though $\mathrm{Var}(r_{\pi'})$ should be preferred over $\mathrm{Var}(r_A \mid r_\pi)$ to measure diversification in such a case. However, the previous discussion on $\mathrm{Var}(r_{\pi'})$ only holds when both r_π and r_A have the same expected return and variance. When either the variances or the expectations differ, we can no longer draw simple conclusions as to whether the return per unit of risk can be increased solely based on $\mathrm{Var}(r_{\pi'})$. On the other hand the conditional variance, which is the general case reads

$$\mathrm{Var}(r_A \mid r_\pi) = \mathrm{Var}(r_A)(1-\rho^2) \tag{3}$$

still provides valuable insight on shared information between A and π, namely that knowing $r_\pi$never increases the uncertainty about r_A; the uncertainty is preserved when the two assets are decorrelated ($\rho = 0$), and decrease otherwise. A natural measure of the diversification A adds to π is therefore

$$D(A;\pi) = \mathrm{Var}(r_A \mid r_\pi) \tag{4}$$

Remark. Expected returns being equal, equation (8) as a measure of incremental diversification penalizes equally new assets with $\rho \approx -1$ and new assets with $\rho \approx 1$, which could be perceived as a limitation, as the former can be used to construct portfolios with much higher return per unit of risk than the latter. This should however not pose a problem in practice, as two assets that have the same expected return are unlikely to have a correlation close to -1; this would be an arbitrage opportunity.

Case 2: 'N' assets, i.i.d Gaussian

This intuitive measure of incremental diversification easily extends to the multi-assets case. If we consider a pool of n assets and factors P, with corresponding returns and factor values drawn independently (across time) from a Gaussian random vector x that is also assumed to be jointly Gaussian with r_A, Equation (8) can be extended to quantify the incremental diversification A adds to the pool P as follows:

$$D(A;\pi) = \mathrm{Var}(r_A \mid x)$$

It immediately follows from Gaussian identities that

$$D(A;\pi) = \frac{\det\left(\mathrm{Cov}\left([x, r_A], [x, r_A]\right)\right)}{\det\left(\mathrm{Cov}(x, x)\right)}$$

where we assume that x is non-degenerate, from which we recover equation in the two assets special case. Let's see how it recover when bivariate case.

$$\begin{aligned}
&= \frac{\det\left(\mathrm{Cov}\left([r_\pi, r_A], [r_\pi, r_A]\right)\right)}{\det\left(\mathrm{Cov}(r_\pi, r_\pi)\right)} \\
&= \frac{\det\begin{pmatrix} \sigma_\pi^2 & \sigma_{\pi,A} \\ \sigma_{\pi,A} & \sigma_A^2 \end{pmatrix}}{\sigma_\pi^2} = \sigma_\pi^2\sigma_A^2 - \sigma_{\pi,A}^2 \\
&= \sigma_A^2 - \rho^2\sigma_A^2 \\
&= \mathrm{Var}(r_A)(1-\rho^2)
\end{aligned}$$

Note: What does Determinant of Covariance Matrix give?

We will see why we use the determinant of the covariance matrix instead of having the covariance matrix itself when writing down the multivariate normal distribution. This becomes straight forward if we look in to bivariate case first,

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

We can see that we write

$$\sigma_x\sigma_y\sqrt{1-\rho^2} = \det(\Sigma)$$

where

$$\Sigma = \begin{pmatrix} \sigma_\pi^2 & \sigma_{\pi,A} \\ \sigma_{\pi,A} & \sigma_A^2 \end{pmatrix}$$

In the univariate case you don't have a determinant because Σ consists of just one term. You don't have another variable, so we don't need to take into account any interaction between them.

In relation to differential entropy

Let's see a connection between the determinant of the covariance matrix of (Gaussian distributed) data points and the differential entropy of the distribution. Let's say we have a (large) set of points from which we assume it is Gaussian distributed. If you compute the determinant of the sample covariance matrix then we measure (indirectly) the differential entropy of the distribution up to constant factors and a logarithm.

The differential entropy of a Gaussian density is defined as

$$H[p] = \frac{k}{2}(1+\log(2\pi)) + \frac{1}{2}\log|\Sigma|$$

where k is the dimensionality of your space. How to show this?

Entropy for normal distribution:

$$\begin{aligned}
H[x] &= -\int_{-\infty}^{\infty} N(x \mid \mu, \Sigma)\log(N(x \mid \mu, \Sigma))dx = E_p[-\log p(x)] \\
&= -E\left[\log\left((2\pi)^{-\frac{D}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}\right)\right] \\
&= \frac{D}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{2}\underbrace{E\left[(x-\mu)^T\Sigma^{-1}(x-\mu)\right]}_{=D} \\
&= \frac{D}{2}(1+\log(2\pi)) + \frac{1}{2}\log|\Sigma|
\end{aligned}$$

We can see that the determinant of covariance matrix shows up in the differential entropy.

Case 3: Beyond Gaussianity

In the non-Gaussian case, the conditional variance $\mathrm{Var}(r_A \mid x)$ might very well be a function of x, so that a more suitable candidate to quantify incremental diversification is obtained by taking the expectation with respect to x:

$$D(A;\pi) = E_x\left[\mathrm{Var}(r_A \mid x)\right]$$

Expected conditional variance as a measure of incremental diversification only captures the first two moments of the joint-distribution. This is sufficient for Gaussian distributions as they are fully determined by their first two moments. **However, non-Gaussian distributions typical exhibit tail behaviors that are not captured by the first two moments; such tail behaviors play a role in our intuitive understanding of risk, and should therefore be embedded in our measure of incremental diversification.**

Another way to look at this is that, although knowing x might not reduce the variance of r_A, if it does affect higher moments of r_A, then A should be regarded as more related to the reference pool than if r_A and x were independent. **Our measure of incremental diversification should therefore be capable of differentiating statistical independence from decorrelation, which as per Proposition** 3.1 **conditional variance cannot.**

Proposition. *Let x and y be two squared-integrable random variables. Then*

$$E_y\left[\mathrm{Var}(x|y)\right] \le \mathrm{Var}(x),$$

and the inequality is an equality if and only if

$$E\left[y \mid x\right] = E\left[y\right] \qquad a.s.,$$

or equivalently, if and only if $\mathrm{Cov}(y, f(x)) = 0$ *for any f.*

The canonical measure of the amount of information in a random variable with probability measure P and admitting pmf or pdf $p(x)$, is the notion of entropy (expressed in bits) defined as

$$H[x] = E_p\left[-\log_2 p(x)\right]$$

Unless stated otherwise, throughout the rest of this paper we assume P admits a pdf. When we need both cases, we will use the expression continuous entropy or differential entropy to emphasize that P admits a pdf, and discrete entropy or Shanon entropy when P admits a pmf, in which case we will use the notation H instead of h.

A related notion is that of conditional entropy,

$$h(y|x) = h(x, y) - h(x)$$

,which measures the amount of information contained in random variable y that is not already contained in random variable x. In the multi-assets Gaussian case, this measure of incremental diversification reads

$$h(r_A \mid x) = \frac{1}{2}(1 + \log(2\pi)) + \frac{1}{2}\log \mathrm{Var}(r_A \mid x)$$

In other words, in the Gaussian case, conditional entropy and conditional variance are equivalent measures of incremental diversification as one is fully determined by other and is an increasing function of the other. **In general, however, conditional entropy is a more general measure of incremental diversification in the following sense.**

Proposition. *Let x and y be two random variables having finite entropies $h(x)$ and $h(y)$. Then*

$$h(y \mid x) \leq h(y)$$

and the inequality is an equality if and only if x and y are independent.

Unlike expected conditional variance that cannot differentiate decorrelation from independence, conditional entropy, as a measure of incremental diversification, is informative about the full distribution tails, and is maximized (for a given entropy $h(r_A)$) when r_A is independent from x, which we recall implies, but is not equivalent to, $\mathrm{Cov}(r_A, f(x)) = 0$ for any f.

Case 3: Beyond Temporal Independence

Conditional entropy as a measure of incremental diversification satisfies both Fact 1 and Fact 2. To see why, we note that, in the Gaussian case, $\mathrm{Var}(r_A \mid x)$ **is also the variance of the residual return of the best replicating portfolio**,

$$r_A \mid x = r_A^*$$

and using Equation (4)

$$\begin{aligned} \mathrm{Cov}(r_A, r_A^*) &= \mathrm{Var}(r_A^*) \\ &= \mathrm{Cov}(r_A, x)\,\mathrm{Cov}(x, x)^{-1}\,\mathrm{Cov}(x, r_A) \end{aligned}$$

this implies

$$\mathrm{Var}(r_A^*) = \mathrm{Cov}(r_A, x)\,\mathrm{Cov}(x, x)^{-1}\,\mathrm{Cov}(x, r_A)$$

,we obtain

$$h(r_A \mid x) = \frac{1}{2}\log_2\left[1 - \mathrm{Cov}(r_A, r_A^*)\sqrt{\frac{\mathrm{Var}(r_A^*)}{\mathrm{Var}(r_A)}}\right] + h(r_A)$$

However, conditional entropy as a measure of incremental diversification does not satisfy Fact 3, as we have been ignoring the temporal aspect of our time series because of our i.i.d assumption across time. To see why, we consider

$$y_t = f(x_{t-i}), \qquad i > 0$$

and note that under our memoryless assumption on the reference pool characteristic time series x_t, y_t is independent from x_t and consequently has the highest conditional entropy for a given $h(y_t)$. **The main issue here is that, as a measure of incremental diversification, conditional entropy does not capture similarities across time.** Independence of returns corresponding to the same time period, x_t and y_t should not be the ideal diversification scenario, independence of the underlying stochastic process $\{x_t\}$ and $\{y_t\}$ should be.

The notion of entropy of random variables is extended to discrete-time stochastic process by the notion of entropy rate which is defined as

$$h(\{x_t\}) = \lim_{T\to\infty} \frac{1}{T} h(x_1, ..., x_T)$$

when the limit exists. The notion of conditional entropy is then extended to define conditional entropy rate as

$$h(\{y_t\} \mid \{x_t\}) = h(\{y_t, x_t\}) - h(\{x_t\})$$

(to be continued)

Reference:

1. https://arxiv.org/pdf/1806.08444.pdf
2. https://math.stackexchange.com/questions/889425/what-does-determinant-of-covariance-matrix-give
3. https://stats.stackexchange.com/questions/89952/why-do-we-use-the-determinant-of-the-covariance-matrix-when-using-the-multivaria
4. https://math.stackexchange.com/questions/2029707/entropy-of-the-multivariate-gaussian