IC, TC and All That

paleologo@gmail.com

November 28, 2025

"TC" is a confusing acronym. It can mean Total Costs, or Tom Cruise, whose Total Costs per movie are extraordinarily high, or Traction Control of motorcycles ridden by Tom Cruise in movies with high Total Costs. For the duration of this note, though, TC will stand for "Transfer Coefficient". The Transfer Coefficient was introduced in a influential paper **?**. The most important things to say about the TC are that: 1) it describes a real need; 2) it is sometimes misunderstood; 3) more work is needed. This note describes the concept, the assumptions, and the open problems. It assumes some basic knowledge of Modern Portfolio Theory, at the level of my blue book (ironically titled by the publisher "Advanced Portfolio Management").

1 Information Coefficient

Richard Grinold was a business professor at UC Berkeley in the 70s and 80s. He started working at Barra (acquired by MSCI in 2004) sometimes in the late 80s, and then moved to BGI (Barclays Global Investors, also called "Barra Graduate Institute", given the large number of Barra alumni in their midst). BGI was a pioneer of quantitative investing. It was acquired by Blackrock in 2010. Around that time Grinold retired. Grinold is perhaps most famous for the Fundamental Law of Active Investing (or FLAM), which is to quantitative researchers what "Ring around the rosie" is to preschoolers: everyone knows it and sings it, but few understand its true meaning. So it seems a good jump-off point.

Start with the simplest setting. There are N assets, and two periods. You purchase assets in period one, and receive a random payoff in period two. The random returns r_i are independent from each other, and have mean α_i and variance σ_i^2. The assumption of independence is not unrealistic. Consider the returns to be idiosyncratic returns. Alternatively, you could interpret the asset as factors with independent returns, which always exist, given a factor model. If this statement makes no sense to you right now, ignore it for now and let me know if you would like to know more (it's ok not to).

A researcher has estimates of the expected returns $\hat{\alpha}_i$. The mean-variance portfolio allocation is

$$w_i^\star = \lambda \hat{\alpha}_i / \sigma_i^2$$

where λ is a positive constant chosen so that the portfolio meets a certain budget constraint. What is the Sharpe Ratio of the portfolio? The PnL is

$$\lambda \sum_i r_i(\hat{\alpha}_i/\sigma_i^2) = \lambda \sum_i (\hat{\alpha}_i/\sigma_i)(r_i/\sigma_i).$$

The volatility of the portfolio is

$$\sqrt{\sum_i (\lambda \sigma_i w_i^\star)^2} = \lambda \sqrt{\sum_i (\hat{\alpha}_i/\sigma_i)^2}.$$

The Sharpe is the ratio of the two.

$$\mathrm{SR} = \frac{\mathrm{PnL}}{\mathrm{Vol}} = \frac{\sum_i (\hat{\alpha}_i/\sigma_i)(r_i/\sigma_i)}{\sqrt{\sum_i (\hat{\alpha}_i/\sigma_i)^2}} \tag{1}$$

Divide and multiply by $\sqrt{\sum_i (r_i/\sigma_i)^2}$:

$$\mathrm{SR} = \frac{\sum_i (\hat{\alpha}_i/\sigma_i)(r_i/\sigma_i)}{\sqrt{\sum_i (\hat{\alpha}_i/\sigma_i)^2}\sqrt{\sum_i (r_i/\sigma_i)^2}} \sqrt{\frac{1}{N}\sum_i (r_i/\sigma_i)^2}\sqrt{N} \tag{2}$$

It is reasonable to assume that the sums in the numerator, $\sum_i (\hat{\alpha}_i/\sigma_i)$ and $\sum_i (r_i/\sigma_i)$, are zero, or very close to zero. Positive and negative positions should approximately be equal in number; and risk-adjusted positions should be approximately be zero. The fraction can be interpreted as a correlation between risk-adjusted alphas and returns. Grinold gives this correlation the name *Information Coefficient*:

$$\boxed{IC = \mathrm{cor}(\alpha/\sigma, r/\sigma)}$$

It makes sense that the Sharpe Ratio should be proportional to this coefficient. Consider the other factor in the formula. For large N, the average

$$N^{-1}\sum_i (r_i/\sigma_i)^2$$

converges to one by the Law of Large Numbers. The formula for the Sharpe Ratio takes the form

$$\mathrm{SR} \simeq \mathrm{cor}(\hat{\alpha}/\sigma, r/\sigma)\sqrt{N} \tag{3}$$

The Sharpe Ratio is the product of two factors. The first one is an *intensive* quantity. It does not depend on the number of assets. It is indicative of skills. The other one is an *extensive* quantity: it scales like the square root of the investment universe. FLAM is intuitive and is useful in many ways. It points the attention to what really matters in cross-sectional alphas. If you have one-period out predictions for your cross-section, a research strategy suggested by FLAM is

1. Get a decent factor model produce idio returns for you, and predicted volatilities.

2. Generate predicted returns and rescale them:

$$\hat{\alpha}_i \leftarrow \frac{\hat{\alpha}_i}{\sqrt{\sum_i (\hat{\alpha}_i/\sigma_i)^2}} \tag{4}$$

3. Regress

$$\frac{r_i}{\sigma_i} = \beta \frac{\hat{\alpha}_i}{\sigma_i} + \epsilon_i$$

   In this regression, the error terms ϵ_i have approximately unit variance. The formula for beta is

$$\hat{\beta} = IC$$

You have reduced your research program to the pursuit of two subproblems. The first one is signal research, which is what I outlined above. The second one is monetization. Take your alphas and make a strategy out of it that will somehow have good real-life Sharpe, in spite of execution costs, financing costs, and whatnot. This program, based on a separation of concerns, is to a large but not *universal* extent followed by most quantitative researchers. If we take a small step further, the signal research program can be generalized a little:

1. "Alphas" are a function from a parameter space Θ and a set of data $X_t \in \mathbb{R}^{N\times M}$ to $\mathbb{R}^N$. The data panel X_t contains M-dimensional vectors of "features" for each asset.

$$\hat{\alpha} : \Theta \times \mathbb{R}^{N\times M} \to \mathbb{R}^N \tag{5}$$

   The function $\hat{\alpha}(\cdot,\cdot)$ encompasses an embarrassingly large array of choices, from linear combinations of columns to the latest and greatest learning architecture with weights representable as a vector in Θ.

2. Compute the empirical IC each period, and average the IC:

$$\hat{\text{IC}}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \text{cor}\left(\frac{\hat{\alpha}(\theta, X_t)}{\sigma}, \frac{r}{\sigma}\right) \tag{6}$$

3. Maximize $\hat{\text{IC}}(\theta)$, and/or compute its value over a "covering" of the parameter space Θ, i.e., $\theta_1, \ldots, \Theta_K$. Compute a confidence interval using some methodology[1].

Let us summarize what we assumed and what we did not assume, either explicitly or implicitly.

We *assumed* that:

1. Returns have finite variances (and therefore finite means).

[1] I cover the subject in my book "The Elements of Quantitative Investing" (Wiley, 2025).

2. Returns are temporally independent across periods
3. Returns are identically distributed in each period.
4. Volatilities are known exactly.
5. Investors are mean-variance optimizers.

We did *not* assume that:

1. Predicted returns $\hat{\alpha}_i$ are accurate.
2. Returns are normally distributed.

2 Transfer Coefficient

Back to the process. We have *signals* $\hat{\alpha}$. Rewrite the expected PnL as

$$E(\sum_i (\alpha_i + \eta_i)(s_i/\hat{\alpha}_i^2) = \sum_i \alpha_i \hat{\alpha}_i/\sigma_i^2$$

$$\begin{aligned}
\mathrm{SR} =& \frac{\sum_i (\alpha_i/\sigma_i)(\hat{\alpha}_i/\sigma_i)}{\sqrt{\sum_i (s_i/\sigma_i)^2}} && (7)\\
=& \frac{\sum_i (\alpha_i/\sigma_i)(\hat{\alpha}_i/\sigma_i)}{\sqrt{\sum_i (\hat{\alpha}_i/\sigma_i)^2}\sqrt{\sum_i (\alpha_i/\sigma_i)^2}}\sqrt{\sum_i (\alpha_i/\sigma_i)^2} && (8)\\
=& \mathrm{cor}(\alpha_i/\sigma_i, s_i/\sigma_i) SR && (9)
\end{aligned}$$

We assumed that holdings took the form $\hat{\alpha}_i/\sigma_i^2$ but in fact the relationship holds for general holdings $w_i(\hat{\alpha}_i, \sigma_i)$:

$$\begin{aligned}
\mathrm{SR} =& \frac{\sum_i \alpha_i w_i}{\sqrt{\sum_i (w_i \sigma_i)^2}} && (10)\\
=& \frac{\sum_i (\alpha_i/\sigma_i)(w_i \sigma_i)}{\sqrt{\sum_i (\alpha_i/\sigma_i)^2}\sqrt{\sum_i (w_i \sigma_i)^2}}\sqrt{\sum_i (\alpha_i/\sigma_i)^2} && (11)\\
=& \mathrm{cor}(\alpha_i/\sigma_i, w_i \sigma_i)\, \mathrm{SR}_{\mathrm{opt}} && (12)
\end{aligned}$$

The quantity SR_{opt} is the highest possible Sharpe Ratio, achieved when forecasts are proportional to the true expected returns. We compute the corresponding $\mathrm{IC}_{\mathrm{opt}}$ in two ways:

$$\text{From the definition of IC:} \qquad \mathrm{IC}_{\mathrm{opt}} = \mathrm{cor}(\alpha/\sigma + \eta, \alpha/\sigma), \qquad \eta \sim N(0,1) \tag{13}$$

$$= \frac{N^{-1} \sum_{i=1}^{N} \frac{\alpha_i^2}{\sigma_i^2}}{\sqrt{N^{-1} \sum_{i=1}^{N} \frac{\alpha_i^2}{\sigma_i^2}}} \tag{14}$$

$$= \sqrt{N^{-1} \sum_{i=1}^{N} \frac{\alpha_i^2}{\sigma_i^2}} \tag{15}$$

$$\text{From } SR_{\mathrm{opt}}: \qquad \mathrm{IC}_{\mathrm{opt}} = \frac{1}{\sqrt{N}} \mathrm{SR}_{\mathrm{opt}} \tag{16}$$

$$= \sqrt{N^{-1} \sum_{i=1}^{N} \frac{\alpha_i^2}{\sigma_i^2}} \tag{17}$$

Summing up, any Sharpe Ratio is equal to

$$\mathrm{SR} = \mathrm{TC}(\alpha, w)\mathrm{IC}_{\mathrm{opt}}\sqrt{N} \tag{18}$$

$$\text{where} \qquad \mathrm{TC}(\alpha, w) := \mathrm{cor}(\alpha_i/\sigma_i, w_i\sigma_i) \tag{19}$$

The Transfer Coefficient is always smaller than one. It is the *necessary* degradation of the Sharpe Ratio when we construct a portfolio under certain conditions, among which:

- having either inaccurate return predictions.
- having inaccurate volatilities.
- Having accurate alpha and vol predictions, and yet constructing a portfolio that does not maximize Sharpe, either by solving a Sharpe maximization with constraints, or by not solving a Sharpe maximization problem at all.

There are two statements that this result does not imply. First, it does *not* follow that adding constraints always results in Sharpe degradation. In fact, when we have inaccurate predictions, side constraints can *ex ante* improve the Sharpe Ratio. As a simple example, consider a wild case. The investor has grossly inaccurate set of volatility predictions. As a result, the sizing of optimal positions from mean-variance optimization is grossly inaccurate. The optimizer may want to add a side constraint of the form $\sum_i w_i^2 \leq L$. This constraint regularizes the optimal weights, thus mitigating the estimation error in the volatilities. Secondly, it does say that

$$\mathrm{SR} = \mathrm{TC}(\alpha, w)\mathrm{cor}(\alpha/\sigma, r/\sigma)\sqrt{N} \tag{20}$$

and *not* that

$$\mathrm{SR} = \mathrm{TC}(\hat{\alpha}, w)\mathrm{cor}(\hat{\alpha}/\sigma, r/\sigma)\sqrt{N} \tag{21}$$

The "hat" makes all the difference! When returns forecasts are inaccurate, then the formula does not hold. Yet, most people interpret the TC in the form of Equation (21). Equation (20) holds instead, and because it requires knowledge of the true expected returns, which we cannot know, most of all of its empirical utility vanishes.

In summary, we *assumed* that:

1. Returns have finite variances (and therefore finite means).
2. Returns are temporally independent across periods...
3. ... and identically distributed in each period.
4. **We know both the true expected returns and the volatilities of the assets.**

We did *not* assume that:

1. Investors are mean-variance optimizers.
2. Returns are normally distributed.

3 What Is To Be Done?

The original intent of the Transfer Coefficient is to establish a link between a pristine strategy and a strategy corrupted by the real world, as exemplified in Figure 1. This link, however, is not as solid as we thought it was. Was should we do? One approach would be to compare the Sharpe Ratios of the two strategies as if they were independent. For large T, the empirical Sharpe Ratio $\hat{SR}^{(i)}$ is normally distributed with mean $\mu_i := \mathrm{SR}^{(i)}$ and standard deviation $\hat{\sigma}_i := \sqrt{(1+\hat{\mu}_i^2/2)/T}$. Asymptotically, under the null hypothesis $\mu_1 = \mu_2$ the statistic

Figure 1: Left: pristine strategy. Right: implemented and corrupted strategy.

$$Z = \frac{\hat{SR}^{(1)} - \hat{SR}^{(2)}}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \tag{22}$$

is normally distributed: $Z \sim N(0,1)$ and we can develop a test of hypothesis.

The strategies, however, are not independent and their excess returns are paired, i.e., we have $(r_{1,t}, r_{2,t})$ for $t = 1, \ldots, T$. The correlation between the strategies ρ is typically positive, and we can take advantage of that. The idea is to derive a statistic on the difference of Sharpe Ratios under the assumption of paired observations. Such a statistic was derived by Jobson and Korkie[2] in the large-sample asymptotic case The thrust of the paper is to compute a Taylor expansion of the statistic; the name for this technique is "delta method", and it is described in any asymptotic statistics textbook[3]. In applications, though, the sample size nowhere close to large. Many investors have an observable track record of one or two years of daily returns; and these returns are heavy-tailed. A bootstrap-based approach is preferable. It is also much simpler than the analytical one. The steps are:

1. For $i = 1, ..., B$:
   (a) Resample with replacement the pairs $(r_{1,t}, r_{2,t})$ to get T pairs $(r^\star_{1,t}, r^\star_{2,t})$.
   (b) Compute the difference of the Sharpe Ratios from resampled data
   $$\Delta S^\star_i := \hat{\mathrm{SR}}^\star_1 - \hat{\mathrm{SR}}^\star_2$$
2. Let $\Delta S^*_{(\gamma)}$ denote the empirical γ-quantile of $\{\Delta S^\star_i\}_{i=1}^B$. A $(1-\alpha)$ confidence interval for ΔS is then
   $$\left[\Delta S^\star_{(\alpha/2)}, \Delta S^\star_{(1-\alpha/2)}\right]$$

And that's it. Still, it's worth outlining the Jobson-Korkie (henceforth, JK) approach, because the formulas give some insight in what to expect from the bootstrap calculations.

3.1 The Analytical Approach

Define the population moments

$$\mu_i := E[r_{i,t}], \qquad \sigma_i^2 := \mathrm{var}(r_{i,t}), \qquad \sigma_{1,2} := \mathrm{cov}(r_{1,t}, r_{2,t}), \tag{23}$$

and the correlation $\rho_{1,2} := \sigma_{1,2}/(\sigma_1\sigma_2)$.

The (population) Sharpe ratios are

$$\eta_i := \frac{\mu_i}{\sigma_i}, \qquad i = 1, 2. \tag{24}$$

We consider the null hypothesis

$$H_0: \ \eta_1 = \eta_2 \quad \iff \quad \Delta\eta := \eta_1 - \eta_2 = 0. \tag{25}$$

Assumptions of the Jobson–Korkie (JK) test:

- $(r_{1,t}, r_{2,t})$ are i.i.d. over t,
- $(r_{1,t}, r_{2,t})$ are jointly normally distributed with finite second moments.

[2] J. D. Jobson and B. M. Korkie, "Performance Hypothesis Testing with the Sharpe and Treynor Measures", *J. Finance* **36**, 889–908 (1981).

[3] E.g., A. W. van der Vaart, "Asymptotic Statistics", *Cambridge University Press* (2012).

Define the sample means, variances and covariance:

$$\hat{\mu}_i := \frac{1}{T}\sum_{t=1}^{T} r_{i,t}, \quad \hat{\sigma}_i^2 := \frac{1}{T}\sum_{t=1}^{T}(r_{i,t} - \hat{\mu}_i)^2, \quad \hat{\sigma}_{1,2} := \frac{1}{T}\sum_{t=1}^{T}(r_{1,t} - \hat{\mu}_1)(r_{2,t} - \hat{\mu}_2)$$

and the sample Sharpe ratios

$$\hat{\eta}_i := \frac{\hat{\mu}_i}{\hat{\sigma}_i}, \qquad i = 1, 2. \tag{26}$$

We are interested in the difference

$$\Delta\hat{\eta} := \hat{\eta}_1 - \hat{\eta}_2. \tag{27}$$

By the delta method for linear combinations,

$$\sqrt{n}(\Delta\hat{\eta} - \Delta\eta) = \sqrt{n}(\hat{\eta}_1 - \hat{\eta}_2 - (\eta_1 - \eta_2)) \xrightarrow{d} N(0, \omega^2), \tag{28}$$

where

$$\omega^2 = \mathrm{var}(\hat{\eta}_1) + \mathrm{var}(\hat{\eta}_2) - 2\mathrm{cov}(\hat{\eta}_1, \hat{\eta}_2). \tag{29}$$

Under the i.i.d. multivariate normal assumptions, the delta–method algebra leads to the Jobson–Korkie large–sample variance:

$$\omega^2 = 2(1 - \rho_{1,2}) + \eta_1^2 + \eta_2^2 - 2\,\eta_1\eta_2\,\rho_{1,2}^2. \tag{30}$$

In practice we plug in sample estimates:

$$\hat{\omega}^2 := 2(1 - \hat{\rho}_{1,2}) + \hat{\eta}_1^2 + \hat{\eta}_2^2 - 2\,\hat{\eta}_1\hat{\eta}_2\,\hat{\rho}_{1,2}^2, \tag{31}$$

where $\hat{\rho}_{1,2}$ is the sample correlation between $r_{1,t}$ and $r_{2,t}$.

Under the null $H_0 : \eta_1 = \eta_2$ we have $\Delta\eta = 0$, and therefore

$$\sqrt{T}\,\Delta\hat{\eta} \xrightarrow{d} N(0, \omega^2). \tag{32}$$

The Jobson–Korkie test statistic is then

$$Z_{\mathrm{JK}} = \frac{\hat{\eta}_1 - \hat{\eta}_2}{\sqrt{\hat{\omega}^2/T}} = \frac{\hat{\eta}_1 - \hat{\eta}_2}{\sqrt{\frac{1}{T}\left[2(1 - \hat{\rho}_{1,2}) + \hat{\eta}_1^2 + \hat{\eta}_2^2 - 2\,\hat{\eta}_1\hat{\eta}_2\,\hat{\rho}_{1,2}^2\right]}}. \tag{33}$$

For large T, under H_0,

$$Z_{\mathrm{JK}} \sim N(0, 1) \tag{34}$$

and one can run tests, or obtain confidence intervals, from this statistic.

Now one can see the benefit of modeling pairwise correlation of the strategies. The denominator is strictly decreasing in the empirical correlation. For larger correlations, Z_{JK} is larger, thus making it easier to reject the null hypothesis.