The Fundamental Law of Active Management: Time Series Dynamics and Cross-Sectional Properties

Zhuanxin Ding, Ph.D.

Portfolio Manager and Head of Quantitative Strategies
Fuller & Thaler Asset Management
411 Borel Ave, Suite 300
San Mateo, CA 94402
650-931-1507
zding@fullerthaler.com

First Draft October 15, 2009
Revised Feb 24, 2010

ABSTRACT

I derive a generalized version of the fundamental law of active management under some weak conditions. I show that the original fundamental law of Grinold and various extensions are all special cases of the generalized fundamental law presented in this paper. I also show that cross-sectional ICs are usually different from time series ICs even if the time series ICs are all the same across securities. The fundamental law derived in this paper is quite robust to forecast model specification. Our results show that the variation in IC (IC volatility over time) has a much bigger impact to portfolio IR than the breadth N for a typical investment universe. I extend the fundamental law to models with multiple factors and study the impact of missing one or more return or risk factors to portfolio IR. Our results also show that the transfer coefficient as originally defined by Clarke et al. (2002) is not able to capture the impact of constraints to portfolio IR in the presence of IC variation. I redefine the concept of transfer coefficient using the cross-sectional correlation between the total conditional covariance adjusted active weights and alphas so that the resulting transfer coefficient has the desired property.

Since the publication of "The fundamental law of active management" by Grinold (1989) two decades ago, it has been widely used in the quantitative investment community as a tool to assess a portfolio manager's ability to add value. According to Grinold (1989), the fundamental law relates three variables: your skill in forecasting exceptional returns (IC), the breadth of your strategy (N), and the value added of your investment strategy (IR). Grinold (1989) claims that "based on assumptions that are not quite true and simplified with some reasonable approximations" the three variables have the following relationship:

$$\mathrm{IR} = \mathrm{IC}\sqrt{N}\,, \tag{1}$$

where IR is the information ratio, IC is the information coefficient, and N is the breadth. Even though Grinold (1989) did not give a precise definition of breadth N, portfolio managers or analysts usually use the number of stocks in the investment universe as breadth. The derivation of the fundamental law is closely related to another Grinold paper (Grinold (1994)) that shows "Alpha is Volatility Times IC Times Score", i.e.,

$$\alpha_{it} = \sigma_{r_i} \,\mathrm{IC}\, z_{it-1}, \tag{2}$$

where σ_{r_i} is the residual return (will be defined below) volatility and z_{it-1} is the standardized forecast signal (score) that is known at the end of time *t*-1. The theoretical and empirical development on this line of the fundamental law culminated in the book by Grinold and Kahn (2000) titled "Active Portfolio Management." Based on the fundamental law, Grinold and Kahn (2000) conclude that "you (portfolio managers) must play often and play well to win at the investment management game. It takes only a modest amount of skill to win as long as that skill is deployed frequently and across a large number of stocks."

Unfortunately, the theoretically calculated IR number from Grinold's fundamental law seems to always overestimate the IR a portfolio manager can reach. For example, given a forecast signal with a monthly average IC of 0.03 and a selection universe of 1000 stocks, the expected annualized IR from Grinold's formula is 3.29 which is beyond even the most optimistic portfolio manager's dreams. Portfolio managers are left wondering why realized information ratios are only a fraction of their predicted value. Clarke *et al*. (2002, p50) point out "a common rule of thumb in practice is that the theoretical information ratio suggested by the fundamental law should be cut in half." However, for the above mentioned example, the IR estimate will still be too high even if cut by half (IR=1.64). As noted by Grinold (1989, p32) himself "an observed information ratio above 1.5 is rare indeed." Of course, it can be the case that the *N* used in our calculation, which is the number of stocks available in the investment universe, is not what meant to be the right measure of breadth by Grinold. Grinold (1989) provides a detailed discussion on this subject and emphasized the importance of counting only independent bets as breadth. Grinold (2007) provides some further discussion on this topic. Unfortunately, it is still not a straightforward exercise to determine what breadth should be used in practice.

Clarke *et al*. (2002) attribute the reduction in performance to the constraints in the portfolio construction process and proposed the concept of "the transfer coefficient" to account for the leaking of IR from Grinold's original formula. They show that constraints in portfolio construction (constraints such as country or sector exposures, long only, etc.), leads to suboptimal portfolio weights in terms of alpha generation, thus reducing the maximum achievable IR. They developed a framework for measuring the deviation of the optimal constrained weights from optimal non-constrained weights and proposed a generalized fundamental law as follows:

$$\mathrm{IR} = \mathrm{TC}\,\mathrm{IC}\sqrt{N}, \tag{3}$$

where TC is the transfer coefficient, defined as the cross-sectional correlation coefficient between risk-adjusted expected residual returns and risk-adjusted active weights. According to their simulation study, the typical transfer coefficient is in the range of 0.3 to 0.8. So the original IR calculated from Grinold's formula should be about halved. Even so, as discussed above, the TC adjusted IR still appears to be too high.

In order to understand why that happens, we need to examine the assumptions made by Grinold in deriving his fundamental law. The original form of the fundamental law by Grinold is based on the very unrealistic assumption that time series ICs between an individual stock's residual return and its forecast signal are the same across all securities and are a constant over time. Grinold (1989, 1994) and Grinold and Kahn (2000) then used the time series IC and cross-sectional IC interchangeably. In practice, many quantitative managers run a Fama-Mcbeth type cross-sectional regression to get realized ICs at different time periods. The ICs calculated this way are far from constant and often fluctuate around an average IC. As will be shown later in this paper, the cross-sectional IC can be quite different from the time series IC even if all the securities have a same time series IC. Qian and Hua (2004) show that a more appropriate IR to use is average IC divided by the standard deviation of IC

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sigma_{\mathrm{IC}}}, \tag{4}$$

where σ_{IC} is the standard deviation of IC that Qian and Hua (2004) call "the strategy risk." In statistics, the quantity $1/\sigma_{\mathrm{IC}}^2$ is a measure of how close (precise) the realized information coefficient at time t, IC_t, is to the mean IC. In this sense, the Qian and Hua formula states that "Information Ratio equals Skill times Precision."

In a more recent paper, Ye (2008) goes one step further to bridge the gap between the original Grinold (1989) formula and the Qian and Hua (2004) formula. Based on her assumptions, she establishes that

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sqrt{1/N + \sigma_{\mathrm{IC}}^2}}. \tag{5}$$

It is obvious that Equation (1) and Equation (4) are special cases of Equation (5) when $\sigma_{\mathrm{IC}} = 0$ (as assumed by Grinold (1989)) or $\mathrm{N} \to \infty$.

With all these different versions of fundamental laws, it can be confusing for practitioners to decide which one to use. It is crucial to have a full grasp of the different underlying assumptions and the resulting conclusions from these fundamental laws. In this paper, I try to set up a coherent econometric modeling structure and show that all the different forms of fundamental laws discussed above can be special cases of an even more general form of fundamental law based on much weaker assumptions. I will show that time series ICs are usually different from cross-sectional ICs even if time series ICs are the same across all individual securities. They will be the same only under some strong conditions. I will also show that different forms of fundamental laws are a result of either unrealistic assumptions (Grinold (1989)) or mis-specified residual return covariance matrices for the expected residual return used (Grinold (1989), Qian and Hua (2004), and Ye(2008)). When the more relevant conditional residual return covariance matrices are used, we will arrive at the more general form of the fundamental law presented in this paper.

The form of the generalized fundamental law derived in this paper is quite robust to model specification. If one uses the risk adjusted residual returns in the analysis instead of the raw residual returns, one will get the fundamental law in a similar form. Finally I extend the fundamental law to models with more than one factor, and discuss the impact of missing one or more return or risk factors to the portfolio IR. I also show that the transfer coefficient as defined by Clarke *et al*. (2002) will not have the desired property of measuring the impact of constraints to the portfolio IR in the presence of IC variation. I redefine the transfer coefficient as the correlation coefficient between total risk adjusted expected residual returns and total risk adjusted active weights (instead of just the diagonal portion of the covariance matrix). With this modified definition of the transfer coefficient, the resulting constrained portfolio IR is always the product of TC and the unconstrained optimal portfolio IR.

Framework and Notation

I will follow the framework and notation in Clarke, de Silva, and Thorley (2002) and Ye (2008). A variable with subscript *i* ($i = 1, \cdots, \mathrm{N}$) and *t* ($t = 1, \cdots, \mathrm{T}$) represents the variable value for security *i* at the end of time *t*. A variable in bold represents a vector or matrix.

Given a benchmark portfolio, the total excess return (i.e., return in excess of the risk-free rate) on any stock *i* can be decomposed into a systematic portion that is correlated with the benchmark excess return and a residual return that is not by

$$r_{it}^{Total} = \beta_{it} R_{B,t} + r_{it} \ , \tag{6}$$

where

β_{it} = beta of security *i* with respect to the benchmark

$R_{B,t}$ = benchmark excess return

r_{it} = *realized* residual return

The benchmark and the actively managed portfolios are defined by the weights, $w_{B,it}$ and $w_{P,it}$, assigned to each of the *N* stocks in the investable universe respectively. It is shown in Clarke *et al*. (2002) that the portfolio active return, which is defined as the managed portfolio total excess return minus the benchmark total excess return, adjusted for the managed portfolio's beta with respect to the benchmark, can be written as

$$R_{A,t} = R_{P,t} - \beta_{P,t} \ R_{B,t} = \sum_{i=1}^{N} w_{P,it} r_{it} = \sum_{i=1}^{N} \Delta w_{it} r_{it} \ , \tag{7}$$

where Δw_{it} is the active weight defined as the difference between the managed portfolio weight and the benchmark weight at the beginning of time period *t*. [1] Note that the active weights, Δw_{it}, sum to 0 because they are differences in two sets of weights that each sum to 1. Also note that the stock returns, r_{it}, in (7) are residual, not total, excess returns. As pointed out in Clarke, *et al*. (2002), residuals are the relevant component of security returns when performance is measured against a benchmark on a beta-adjusted basis.

We assume that residual returns follow a conditional normal distribution, and define *ex ante* alpha of security *i* ($i = 1, \cdots, \mathrm{N}$) in period *t* as the *expected* residual return conditional on information available at the end of time period $t-1$: I_{t-1}

$$\boldsymbol{\alpha}_t = E(\mathbf{r}_t \mid \mathrm{I}_{t-1}) \ , \tag{8}$$

and we define risk related to the alpha expectation as the conditional covariance of the forecast errors

$$\mathbf{\Omega}_t = E[(\mathbf{r}_t - \boldsymbol{\alpha}_t)(\mathbf{r}_t - \boldsymbol{\alpha}_t)' \mid \mathrm{I}_{t-1}], \qquad (9)$$

where $\boldsymbol{\alpha}_t$ and $\mathbf{r}_t$ are $N \times 1$ vectors with α_{it} and r_{it} as their elements respectively. The assumption of asset return normality is one of the fundamental assumptions under Markowitz's mean-variance portfolio choice theory, and the mean and covariance matrix fully determine a multivariate normal distribution. Under the residual return normality assumption, the covariance of the forecast errors is the relevant measure of risk. There is risk because there is uncertainty, and risk is associated to the part of return that we are not able to predict. If we know the future returns perfectly then there is no uncertainty, hence no risk. The conditional risk associated with our alpha estimate should be smaller than the total risk around the unconditional alpha expectation. If this is not the case, then the forecast provides no additional information and the lagged information set, I_{t-1}, is useless. This is the major difference between the risk model used in this paper and the risk models used in Grinold (1989, 1994), Grinold and Kahn (2000), Clarke *et al*. (2002), Qian and Hua (2004), and Ye (2008). Of course, the assumption of stock return normality may not be valid in practice, and the return and risk models one uses are very likely mis-specified, which may cause theoretically derived results not to reflect what one gets in reality. I will give some discussion later on the impact of missing alpha or risk factors in conditional mean and covariance modeling.

After having specified the conditional mean and covariance matrix, we will then use the mean-variance analysis tool for portfolio construction based on the theory of utility maximization. In each period *t*, the optimal market-neutral portfolio, P_t, is selected to maximize the mean-variance utility function:

$$\begin{aligned} &\underset{\Delta\mathbf{w}_t}{Max} \quad U_t = \alpha_{Pt} - \frac{1}{2}\lambda\sigma_{Pt}^2 = \Delta\mathbf{w}_t'\boldsymbol{\alpha}_t - \frac{1}{2}\lambda\Delta\mathbf{w}_t'\mathbf{\Omega}_t\Delta\mathbf{w}_t, \\ &s.t. \quad \Delta\mathbf{w}_t'\mathbf{1} = 0 \end{aligned} \qquad (10)$$

where

α_{Pt} = expected active return on the portfolio

σ_{Pt}^2 = active risk of the portfolio based on the portfolio holdings

λ = a risk-aversion parameter

$\mathbf{1}$ = $N \times 1$ vector of 1s

The solution for this optimization problem is

$$\Delta\mathbf{w}_t = \frac{1}{\lambda}(\mathbf{\Omega}_t^{-1}\boldsymbol{\alpha}_t - \kappa\mathbf{\Omega}_t^{-1}\mathbf{1}), \qquad (11)$$

where $\kappa = \dfrac{\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1}}$ is a scalar.

A certain value of λ corresponds to a certain value of σ_{Pt} since

$$\Delta\mathbf{w}_t'\mathbf{\Omega}_t\Delta\mathbf{w}_t = \sigma_{Pt}^2. \qquad (12)$$

Substituting (11) into (12) and by some straightforward algebra we have

$$\lambda = \frac{1}{\sigma_{Pt}}\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t - \kappa\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t} \,. \tag{13}$$

The optimal portfolio active weight is then

$$\Delta\mathbf{w}_t = \sigma_{Pt}\frac{\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}{\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}}, \tag{14}$$

and the expected portfolio return

$$\begin{aligned}\alpha_{Pt} &= \Delta\mathbf{w}_t{}'\boldsymbol{\alpha}_t \\ &= \sigma_{Pt}\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})} \,.\end{aligned} \tag{15}$$

If we assume that the target tracking error remains a constant ($\sigma_{Pt} = \sigma_P$) at each rebalance of the portfolio, a typical practice for many quantitative portfolio managers, then the *ex ante* expected information ratio of the portfolio is

$$\begin{aligned}\mathrm{IR} &= \frac{\overline{\alpha}_{Pt}}{\sigma_P} = \frac{1}{T}\sum_{t=1}^{T}\frac{\alpha_{Pt}}{\sigma_P} \\ &= \frac{1}{T}\sum_{t=1}^{T}\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})} \\ &= E\left(\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}\right).\end{aligned} \tag{16}$$

This is a very general result that should hold as long as the residual return has a conditional normal distribution with mean $\boldsymbol{\alpha}_t$ and covariance matrix $\boldsymbol{\Omega}_t$.

From the above discussion, it is clear that the key is how to forecast the alpha and the corresponding covariance matrix. As Kahn (1997) points out "active management is forecasting." Different forecasts will give us different *ex ante* expected information ratios. In the literature, two different approaches are used to forecast alpha. One uses time series models and the other uses a Fama-McBeth type cross-sectional regression approach. As for covariance matrix, many people use a risk model that does not have a direct relationship with the alpha estimation, such as the commercial risk models by BARRA or Northfield. Strictly speaking, a risk model that is detached from the alpha model will be a mis-specified risk model for the reasons discussed above. This mis-specification usually results in the underestimation of risk when one runs an actual portfolio because the very important "strategy risk" is being left out (see Qian and Hua (2004), Qian, Hua, and Sorensen (2007)).

Time Series Dynamics

In the original papers about the fundamental law, Grinold (1989, 1994) concluded that "alpha is volatility times IC times score" without providing the explicit model assumptions and technical derivations of his result. In the endnote of his first paper (1989) he did mention that technical details are available upon request. Detailed discussions were given instead in Chapters 10 and 11 of the book by Grinold and Kahn (2000). Unfortunately, even though their Equation (10.1) is assumed to be for a cross section of *N* assets, the result in (10.16) is derived through a time series model for each of the *N* individual assets. They then use the time series IC and cross-sectional IC interchangeably. The discussion below will show that the result from time series modeling assumptions cannot be applied to cross-sectional modeling structures without some further assumptions.

If we assume that the true forecasting relationship between the lagged information set, I_{t-1}, and the residual returns, r_{it}, is a linear one factor model as follows

$$r_{it} = g_i z_{it-1} + \varepsilon_{it} \tag{17}$$

for security *i* over time *t* = 1, 2, 3, ..., *T*. In the equation, g_i is the time series factor return (g_i is just a regression coefficient and is different from the usual definition of factor return from a cross-sectional regression) for security *i*, z_{it-1} is the factor exposure that becomes known at the end of time *t*-1 that has both time series and cross-sectional mean 0 and standard deviation 1 (as assumed by Grinold and Kahn (2000), p268), $\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon_i})$ is the idiosyncratic noise that cannot be predicted. We further assume

T1) $E(z_{it-1}\varepsilon_{it}) = 0$ for all *i* and *t*,

and

T2) $E(\varepsilon_{it}\varepsilon_{jt}) = 0$ for $i \neq j$.

T1) is a very general assumption for linear regression models stating that the explanatory variable and the residual are not correlated, and T2) assumes that the forecast errors are not correlated across stocks so that the idiosyncratic covariance matrix is diagonal. This is also a common assumption for idiosyncratic noise.

For ease of exposition, we will focus our attention on population quantity and ignore the sample estimation error of the parameters. Basic regression of Equation (17) gives us,

$$\begin{aligned} g_i &= \mathrm{Var}^{-1}(z_i)\,\mathrm{Cov}(z_i, r_i) \\ &= \frac{\mathrm{Cov}(z_i, r_i)}{\sqrt{\mathrm{Var}(z_i)\,\mathrm{Var}(r_i)}}\sqrt{\frac{\mathrm{Var}(r_i)}{\mathrm{Var}(z_i)}}, \\ &= \mathrm{IC}_{ts,i}\sigma_{r_i} / \sigma_{z_i} \end{aligned} \tag{18}$$

where $\mathrm{IC}_{ts,i}$ is the time series correlation between residual return r_{it} and forecast signal z_{it-1}, σ_{r_i} is the standard deviation (volatility) of residual return r_{it}, and σ_{z_i} is the standard deviation (volatility) of z_{it-1} which is 1 by assumption. The time series prediction for alpha from this model is

$$\alpha_{it} = E(r_{it} \mid I_{t-1}) = \mathrm{IC}_{ts,i}\sigma_{r_i} z_{it-1}, \tag{19}$$

and the conditional volatility, or forecast error volatility, is

$$\sigma^2_{\varepsilon_i} = \mathrm{Var}(r_{it} \mid I_{t-1}) = (1 - \mathrm{IC}^2_{ts,i})\sigma^2_{r_i}. \tag{20}$$

It should be noted here that $\sigma_{r_i} \neq \sigma_{\varepsilon_i}$ when $\mathrm{IC}_{ts,i} \neq 0$. As we discussed above for Equation (9), when the forecast signal z_{it-1} contains useful information for predicting residual return r_{it}, then the resulting error variance ($\sigma^2_{\varepsilon_i}$) should be smaller than the original unconditional residual return variance ($\sigma^2_{r_i}$). This is the major difference between the risk estimate here and the risk estimate provided by any commercial risk model which has no connection with alpha estimation.

Substituting the alpha and volatility prediction into Equation (16) we have the *ex ante* expected information ratio as

$$
\begin{aligned}
\mathrm{IR} &= E\left(\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}\right) \\
&= E\left(\sqrt{\sum_{i=1}^{N}\frac{\mathrm{IC}_{ts,i}^2 z_{it-1}^2}{1-\mathrm{IC}_{ts,i}^2} - \kappa\sum_{i=1}^{N}\frac{\mathrm{IC}_{ts,i} z_{it-1}}{(1-\mathrm{IC}_{ts,i}^2)\sigma_{r_i}}}\right).
\end{aligned}
\tag{21}
$$

If we assume that the cross-sectional distribution of $\mathrm{IC}_{ts,i}$ and z_{it-1} are independent, then as N becomes large, we have

$$
\begin{aligned}
\mathrm{IR} &= E\left(\sqrt{NE_{cs}\left(\frac{\mathrm{IC}_{ts,i}^2 z_{it-1}^2}{1-\mathrm{IC}_{ts,i}^2}\right) - N\kappa E_{cs}\left(\frac{\mathrm{IC}_{ts,i} z_{it-1}}{(1-\mathrm{IC}_{ts,i}^2)\sigma_{r_i}}\right)}\right) \\
&= E\left(\sqrt{NE_{cs}\left(\frac{\mathrm{IC}_{ts,i}^2}{1-\mathrm{IC}_{ts,i}^2}\right)E_{cs}z_{it-1}^2 - N\kappa E_{cs}\left(\frac{\mathrm{IC}_{ts,i}}{(1-\mathrm{IC}_{ts,i}^2)\sigma_{r_i}}\right)E_{cs}z_{it-1}}\right) \\
&= \sqrt{\sum_{i=1}^{N}\frac{\mathrm{IC}_{ts,i}^2}{1-\mathrm{IC}_{ts,i}^2}},
\end{aligned}
\tag{22}
$$

where E_{cs} stands for the cross-sectional expectation operator. In deriving Equation (22) we used the assumption that the forecast signal, z_{it-1}, is cross-sectionally normalized to have mean 0 and standard deviation 1. When all the time series ICs are the same, *i.e.* $\mathrm{IC}_{ts,i} = \mathrm{IC}_{ts}$ for all i, we have

$$
\mathrm{IR} = \frac{\mathrm{IC}_{ts}}{\sqrt{1-\mathrm{IC}_{ts}^2}}\sqrt{N} \approx \mathrm{IC}_{ts}\sqrt{N}\,. \tag{23}
$$

The approximation holds when IC_{ts} is small which is typically the case in empirical work.

Equation (23) proved that the original fundamental law of Grinold (1989) holds approximately under the time series model assumption when ICs are the same across all the assets and is small. The reason that the original formula of Grinold (1989) needs to be adjusted by $\sqrt{1-\mathrm{IC}_{ts}^2}$ is that we used the conditional volatility of the residual return instead of the unconditional one. Some interesting observations can be made from Equations (22) and (23). When one has the skill to predict some residual returns perfectly (some $\mathrm{IC}_{ts,i} = 1$) then the IR shall go to infinity no matter what the breadth is. This makes intuitive sense because if one can predict some residual returns perfectly then she/he can make a sure bet on these stocks against the rest of the universe to achieve the desired excess return. The IR will be infinity since the optimization is set in such a way that one can take a leveraged bet. This is not a feature in the original Grinold formula which states that the IR will increase with the square root of N even if $\mathrm{IC}_{ts} = 1$.

If, instead of running a time series regression, we run a "mis-specified" cross-sectional regression for the model in Equation (17),

$$r_{it} = f_t\, z_{it-1} + \xi_{it} \tag{24}$$

for cross-sectional security i = 1, 2 ,..., N at time t. A simple cross-sectional regression gives us

$$\begin{aligned} f_t &= E_{cs,t}(r_{it} z_{it-1}) / E_{cs,t}(z_{it-1}^2) \\ &= \frac{E_{cs,t}(r_{it} z_{it-1})}{\sqrt{E_{cs,t}(r_{it}^2) E_{cs,t}(z_{it-1}^2)}} \sqrt{\frac{E_{cs,t}(r_{it}^2)}{E_{cs,t}(z_{it-1}^2)}} \\ &= \mathrm{IC}_{cs,t} d(\mathbf{r}_t) / d(\mathbf{z}_{t-1}) \\ &= \mathrm{IC}_{cs,t} d(\mathbf{r}_t)\,, \end{aligned} \tag{25}$$

where $E_{cs,t}$ stands for the cross-sectional expectation operator at time t, $\mathrm{IC}_{cs,t}$ is the cross-sectional correlation between residual return r_{it} and forecast signal z_{it-1}, $d(\mathbf{z}_{t-1})$ is the cross-sectional standard deviation (dispersion) [2] of z_{it-1}, which is 1 by assumption, and $d(\mathbf{r}_t)$ is the cross-sectional residual return dispersion at time t.

The expected value of f_t is

$$\begin{aligned} f &= E(f_t) = E(E_{cs,t}(r_{it} z_{it-1})) = E_{cs,t}(E(r_{it} z_{it-1})) \\ &= E_{cs,t}(E((g_i z_{it-1} + \varepsilon_{it}) z_{it-1})) \\ &= E_{cs,t}(g_i) = \frac{1}{N}\sum_{i=1}^{N} g_i \\ &= \frac{1}{N}\sum_{i=1}^{N} \mathrm{IC}_{ts,i}\sigma_{r_i}\,. \end{aligned} \tag{26}$$

On the other hand, if we assume $\mathrm{IC}_{cs,t}$ and $d(\mathbf{r}_t)$ are independent over t, then from Equation (25) we have

$$\begin{aligned} f &= E(f_t) \\ &= E(\mathrm{IC}_{cs,t} d(\mathbf{r}_t)) \\ &= E(\mathrm{IC}_{cs,t}) E(d(\mathbf{r}_t)) \\ &= \mathrm{IC}_{cs}\delta\,, \end{aligned} \tag{27}$$

where $\delta = E(d(\mathbf{r}_t))$ is the expected cross-sectional residual return dispersion. Substituting (26) into (27) we have

$$\mathrm{IC}_{cs} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IC}_{ts,i}\sigma_{r_i} / \delta\,, \tag{28}$$

i.e., the expected cross-sectional IC, IC_{cs}, is a weighted average of time series ICs and they are usually not the same. If the time series ICs are the same across all securities, i.e., $\mathrm{IC}_{ts,i} = \mathrm{IC}_{ts}$ for all i then

$$\text{IC}_{cs} = \text{IC}_{ts} \frac{1}{N} \sum_{i=1}^{N} \sigma_{r_i} / \delta = \text{IC}_{ts} \tilde{\sigma}_r / \delta \quad , \tag{29}$$

where $\tilde{\sigma}_r = \frac{1}{N} \sum_{i=1}^{N} \sigma_{r_i}$ is the cross-sectional average of the residual return standard deviation. So as long as $\tilde{\sigma}_r \neq \delta$, we have the seemly surprising result that the cross-sectional IC_{cs} will be different from the time series IC_{ts} even if the time series ICs are the same across all securities.

In the extreme case that all residual return standard deviations are the same, i.e. $\sigma_{r_i} = \sigma_r$ for all *i*, we have $\tilde{\sigma}_r = \sigma_r = \delta$ and $\text{IC}_{cs} = \text{IC}_{ts}$.[3] So the discussion here shows that the cross-sectional IC is usually different from the time series IC for an identical set of return and factor exposures. They will only be the same under the very strong assumption that the residual return volatilities are the same across all securities.

Given the "mis-specified" cross-sectional model prediction for each individual security,

$$\alpha_{it} = \text{IC}_{cs} \delta z_{it} = \text{IC}_{ts} \tilde{\sigma}_r z_{it} \ , \tag{30}$$

we have the forecast error term as

$$\xi_{it} = \text{IC}_{ts} \sigma_{r_i} z_{it} - \text{IC}_{cs} \delta z_{it} + \varepsilon_{it} = \text{IC}_{ts} (\sigma_{r_i} - \tilde{\sigma}_r) z_{it} + \varepsilon_{it} \, , \tag{31}$$

which is different from ε_{it}. The conditional covariance matrix has the following elements:

$$\omega_{ij} = E(\xi_{it} \xi_{jt}) = \begin{cases} \text{IC}_{ts}^2 (\sigma_{r_i} - \tilde{\sigma}_r)^2 + (1 - \text{IC}_{ts}^2) \sigma_{r_i}^2 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases} \tag{32}$$

Substituting (32) into (16) we have

$$\text{IR} = E\left(\sqrt{\sum_{i=1}^{N} \frac{\text{IC}_{ts}^2 \tilde{\sigma}_r^2 z_{it}^2 - \kappa \text{IC}_{ts} \tilde{\sigma}_r z_{it}}{\text{IC}_{ts}^2 (\sigma_{r_i} - \tilde{\sigma}_r)^2 + (1 - \text{IC}_{ts}^2) \sigma_{r_i}^2}} \right). \tag{33}$$

If we assume that the cross-sectional distribution of σ_{r_i} and z_{it} are independent, then as *N* becomes large, we have

$$\text{IR} = \text{IC}_{ts} \sqrt{\sum_{i=1}^{N} \frac{1}{\text{IC}_{ts}^2 (\sigma_{r_i} / \tilde{\sigma}_r - 1)^2 + (1 - \text{IC}_{ts}^2)(\sigma_{r_i} / \tilde{\sigma}_r)^2}} \, . \tag{34}$$

When all the residual return volatilities are the same we have

$$\text{IR} = \frac{\text{IC}_{ts}}{\sqrt{1 - \text{IC}_{ts}^2}} \sqrt{N} \approx \text{IC}_{ts} \sqrt{N} \, , \tag{35}$$

which is consistent with the result from time series model. When the individual residual return standard deviation varies across securities, the IR we get from the mis-specified cross-sectional model will be different from the IR we get from the time series model.

The discussion above shows that the original fundamental law of Grinold (1989, 1994) only holds under the assumption that the time series ICs are the same across all the securities and the common IC is small. The cross-sectional IC is only the same as the time series IC if an additional assumption is imposed that all residual return standard deviations are the same (Ye (2008) made this assumption).

In practice, the above two assumptions (time series ICs and residual return volatilities are the same across all securities) are overly restrictive and we can almost surely say they do not hold. As an example, I calculated monthly means and standard deviations for time series and cross-sectional ICs for book/price ratio (B/P) and Momentum factors for US stocks in Table 1. The top panels in Figures 1 and 2 show the time series IC distributions for both factors. It can be seen that the time series ICs have a normal-like distribution with high dispersion. The bottom panels in Figures 1 and 2 show the cross-sectional IC distributions for both factors. It can be seen that the cross-sectional ICs are more highly concentrated and are positively skewed.

It is also interesting to see that the average time series ICs for B/P are much higher than the average cross-sectional ICs, especially if the time series B/P is not standardized. The average time series ICs for momentum are negative whether you standardize them in one or both dimensions. The average momentum factor cross-sectional IC is positive only if one does not standardize the exposures in the time dimension.

Further research shows that the basic form of the fundamental law under the time series model assumptions does not change even if I assume the time series ICs to be different across stocks and follow certain cross-sectional distributions (such as a Beta distribution in the range of -1 to 1).

Table 1. Mean and Standard Deviation for Factor IC (Time Series and Cross-Section)

Factors	Time Series			Cross-Section			
	mean	std	n	mean	std	n	t-test
Original Signal							
B/P	0.088	0.176	15232	0.017	0.062	412	1.82
MOM	-0.028	0.152	15232	0.025	0.099	412	-1.58
Both Dimension Normalized							
B/P	0.087	0.175	15232	0.050	0.072	412	0.94
MOM	-0.028	0.152	15232	-0.003	0.085	412	-0.74

Figure 1. Histogram for Time Series and Cross-Sectional Correlation
One dimension standardized

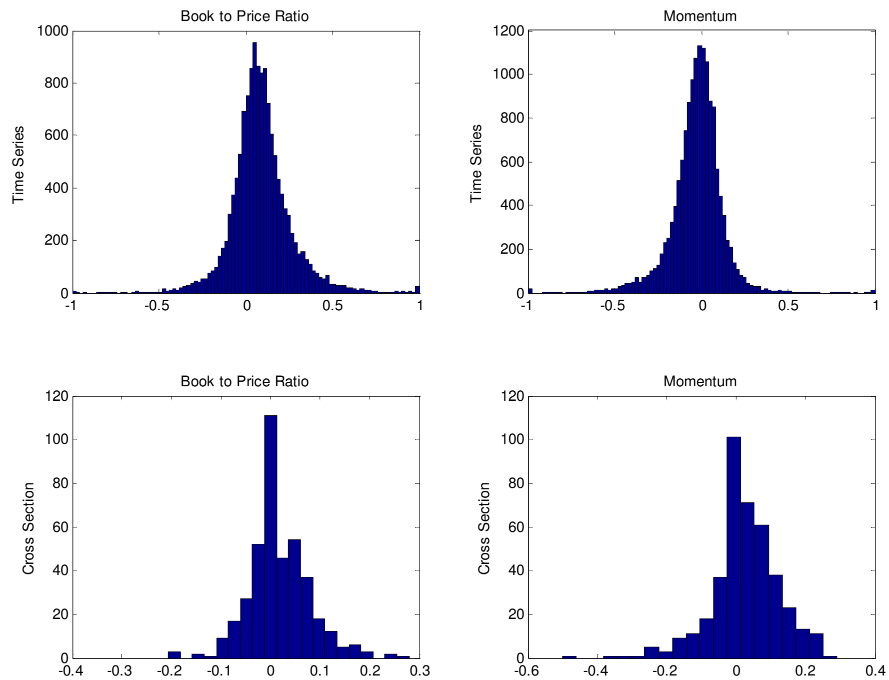
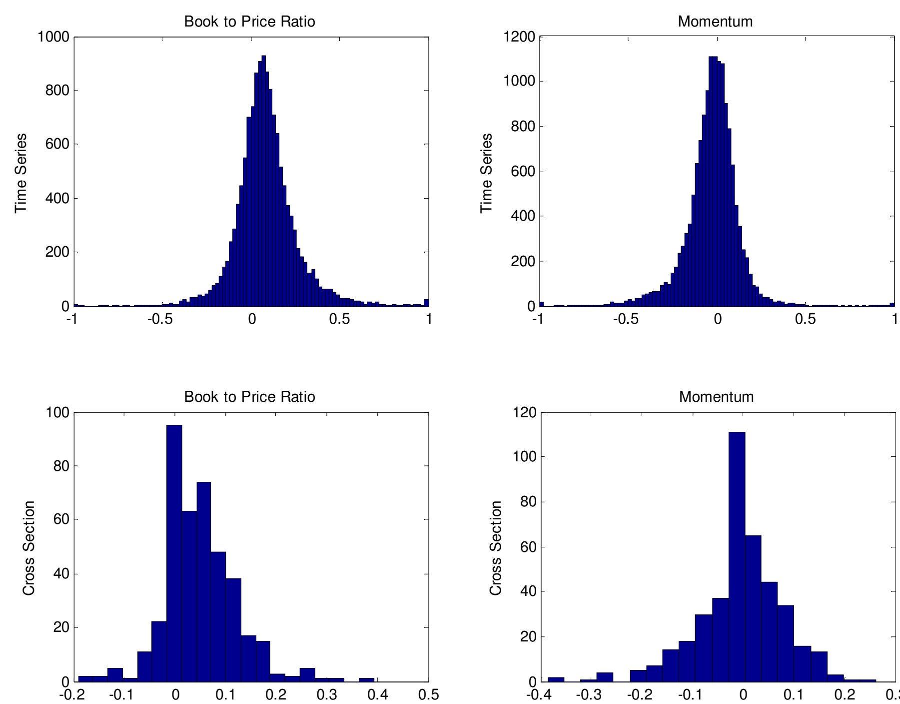


Figure 2. Histogram for Time Series and Cross-Sectional Correlation
Both dimensions standardized



Cross-Sectional Properties

The above discussion shows the assumption that all time series ICs are the same is not realistic. I will show below it is also not necessary in deriving the (generalized) fundamental law. In empirical finance work, many people use a Fama-McBeth type cross-sectional regression in relating the explanatory variables with asset returns. Ibragimov and Müller (2009) find that as long as the cross-sectional coefficient estimators are approximately normal (or scale mixtures of normals) and independent, the Fama-MacBeth method results in valid inference even for a short panel that is heterogeneous over time. Due to the small sample conservativeness result, the approach allows for unknown and unmodelled heterogeneity. Peterson (2009) shows that when the residuals of a given time period are correlated across firms, the Fama-McBeth method produces more efficient estimates than OLS and the standard error will be correct. Another advantage is that the assumptions we have to make to achieve the kind of fundamental law are much weaker than the assumptions we have to make in the time series section.

Assume the basic modeling structures are similar to Equation (17), only this time we have the relationship at time *t* for *i* = 1, 2, 3, ..., *N* assets,

$$r_{it} = f_t z_{it-1} + \varepsilon_{it} \tag{36}$$

where f_t is the cross-sectional factor return at time *t*, z_{it-1} is the factor exposure that becomes known at the end of time *t*-1 that has both time series and cross-sectional mean 0 and standard deviation 1, $\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon_i})$ is the idiosyncratic noise that cannot be predicted. We will make the same assumptions as in time series model concerning z_{it-1} and ε_{it}:

C1) $E(z_{it-1}\varepsilon_{it}) = 0$ for all *i* and *t*,

and

C2) $E(\varepsilon_{it}\varepsilon_{jt}) = 0$ for $i \neq j$.

Under the above assumptions, we have,

$$f_t = \mathrm{IC}_t d(\mathbf{r}_t), \tag{37}$$

where $d(\mathbf{r}_t)$ is the cross-sectional residual return dispersion assumed to be a constant (δ) over time,[4] and IC_t is the cross-sectional IC (all the ICs discussed in this section will be cross-sectional IC unless otherwise specified) between the residual returns and the forecast signals. In empirical work, one needs to get an *ex ante* estimate for the cross-sectional correlation IC_t before making an estimate for the alpha. The most common and simple method just uses historical average as an estimate. After the fact we can estimate the *ex post* realized IC_t using the actual r_{it} and z_{it-1}. As shown in the bottom panels of Figures 1 and 2, usually the cross-sectional factor IC spreads around a mean. For ease of exposition below, we will assume that the cross-sectional factor IC_t follows a normal distribution with mean IC and standard deviation σ_{IC}.[5]

When the alpha model has the linear one factor structure in Equation (36) and under the above assumptions, we have the conditional expectation (on known $\mathbf{z}_{t-1}$) of $\mathbf{r}_t$ as

$$\boldsymbol{\alpha}_t = E(\mathbf{r}_t \mid \mathrm{I}_{t-1}) = \mathrm{IC}\,\delta\mathbf{z}_{t-1}, \tag{38}$$

and the conditional covariance as

$$\boldsymbol{\Omega}_t = E((\mathbf{r}_t - \boldsymbol{\alpha}_t)(\mathbf{r}_t - \boldsymbol{\alpha}_t)' \mid \mathbf{I}_{t-1}) = \sigma_{\mathbf{IC}}^2 \delta^2 \mathbf{z}_{t-1}\mathbf{z}_{t-1}' + \boldsymbol{\Sigma}_t, \tag{39}$$

where $\boldsymbol{\Sigma}_t$ is the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ which should be diagonal according to assumption C2) above [6]

$$\boldsymbol{\Sigma}_t = diag(\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2, \cdots, \sigma_{\varepsilon_n}^2) \tag{40}$$

where $\sigma_{\varepsilon_i}^2 = \sigma_{r_i}^2 - (\mathrm{IC}^2 + \sigma_{\mathrm{IC}}^2)\delta^2$.

Given the above modeling assumptions and by some straightforward algebra, it is shown in Appendix A that the *ex ante* expected portfolio excess return at time *t* to be

$$\alpha_{Pt} = \sigma_{Pt} \frac{\mathrm{IC}}{\sqrt{1/(\phi N) + \sigma_{\mathrm{IC}}^2}}, \tag{41}$$

where $\phi \geq 1$ is a constant that is defined in Appendix A.

So the so-called fundamental law in the more general form should be

$$\mathrm{IR} = \frac{\overline{\alpha}_{Pt}}{\sigma_{Pt}} = \frac{\mathrm{IC}}{\sqrt{1/(\phi N) + \sigma_{\mathrm{IC}}^2}}. \tag{42}$$

The portfolio IR is positively related to the average cross-sectional IC (skill) and the square root of *N* (breadth), but inversely related to the cross-sectional IC standard deviation, $\sigma_{\mathbf{IC}}$ (Qian and Hua (2004) call this *strategy risk*). This result should not be surprising to any student of modern portfolio theory. Basically it states that for a portfolio built upon a sufficiently large universe (large N), the main risk of the portfolio comes from the bet on the alpha factor that has an uncertain (but positive average) payoff stream (strategy risk). As the universe (*N*) becomes larger, the impact of the idiosyncratic risk ($1/(\phi N)$ part in the formula) will diminish. Three interesting special cases emerge from Equation (42):

1) if the cross-sectional IC_t is a constant over time, i.e., $\sigma_{\mathrm{IC}} = 0$, and all the residual return standard deviations (σ_{r_i}) are the same across assets (hence $\phi = 1/(1-\mathrm{IC}^2)$) then we have $\mathrm{IC} = \mathrm{IC}_{ts}$, and the adjusted Fundamental Law of Grinold (1989) we derived in the time series dynamics section: $\mathrm{IR} = \frac{\mathrm{IC}}{\sqrt{1-\mathrm{IC}^2}}\sqrt{N} \approx \mathrm{IC}\sqrt{N}$.
2) when the breadth goes to infinity, or $N >> 1/(\phi\sigma_{\mathrm{IC}}^2)$, then we have the IR formula of Qian and Hua (2004): $\mathrm{IR} = \mathrm{IC}/\sigma_{\mathrm{IC}}$. The formula by Qian and Hua (2004) is interesting in that they got the final result almost right even though they used a conditional covariance matrix that is inconsistent with their alpha forecast assumptions. They realized that there is a "strategy risk" which is a form of systematic risk for their bets. But they missed this risk in their *ex ante* risk model

because they used a third party risk model that is detached from their alpha model. This is common to all quantitative strategies that use a third party risk model. Lee and Stefek (2008) give a very good discussion on this topic. The *ex post* realized portfolio risk is mainly from the "strategy risk" that cannot be diversified away by the optimal portfolio. That is why their *ex ante* target tracking error is so different from the *ex post* tracking error they derived.

3) if all the residual return standard deviations (σ_{r_i}) are the same at time *t* but the IC volatility is not zero (hence $\phi = 1/(1-\mathrm{IC}^2-\sigma_{\mathrm{IC}}^2)$), then we have approximately the IR formula of Ye (2008) $\mathrm{IR} = \dfrac{\mathrm{IC}}{\sqrt{(1-\mathrm{IC}^2-\sigma_{\mathrm{IC}}^2)/N+\sigma_{\mathrm{IC}}^2}} \approx \dfrac{\mathrm{IC}}{\sqrt{1/N+\sigma_{\mathrm{IC}}^2}}$ (empirically factor IC is in the range of 0.02 to 0.05 and IC standard deviation is around 0.1). The approximation results from Ye (2008) using the unconditional residual return standard deviation in her risk model instead of the conditional idiosyncratic error standard deviation that is consistent with the alpha model. In this formula we will also have the property that IR will go to infinity when IC=1 and $\sigma_{\mathrm{IC}} = 0$ no matter what the breadth (*N*) is, while Ye's original formula does not have this feature.

It should be noted that the *ex ante* and *ex post* IR calculation should be very close if the return and risk models are correctly specified (which is a strong assumption!). The difference between the *ex ante* and *ex post* IR should be a result of standard error in parameter estimation. As the sample size gets bigger, the difference should get smaller. If this is not the case, then we can be quite sure that the *ex ante* model specification is incorrect. Since we ignored the sample estimation error in this paper, we should expect the *ex ante* and *ex post* IR to be the same when the model is correctly specified.

As an example, let us look at the realized portfolio excess returns from the above model and calculate the *ex post* IR based on the realized alphas. For ease of exposition, I will assume $\phi = 1/(1-\mathrm{IC}_t^2-\sigma_{\mathrm{IC}}^2)$ (as will be shown in next section, this is true if we use risk-adjusted residual returns in analysis). The realized one period portfolio alpha from the return and risk model is (based on Equation 41)

$$\alpha_{Pt} = \sigma_{Pt} \frac{\mathrm{IC}_t}{\sqrt{(1-\mathrm{IC}_t^2-\sigma_{\mathrm{IC}}^2)/N+\sigma_{\mathrm{IC}}^2}}, \tag{43}$$

where σ_{Pt} is the *ex ante* portfolio tracking error target set as a constant ($\sigma_{Pt} = \sigma_P$). For a specific time period, IC_t can be positive or negative which will result in positive or negative excess return for the portfolio. The portfolio average excess return over time is then

$$
\begin{aligned}
\bar{\alpha}_{Pt} &= \frac{1}{T}\sum_{t=1}^{T}\alpha_{Pt} \\
&= \sigma_P \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{IC}_t}{\sqrt{(1-\mathrm{IC}_t^2-\sigma_{\mathrm{IC}}^2)/N+\sigma_{\mathrm{IC}}^2}} \\
&\approx \sigma_P \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{IC}_t}{\sigma_{\mathrm{IC}}} \qquad \text{when } N \text{ is large} \\
&= \sigma_P \frac{\mathrm{IC}}{\sigma_{\mathrm{IC}}},
\end{aligned} \tag{44}
$$

and the standard deviation of the portfolio average excess return is

$$
\begin{aligned}
Std(\alpha_{Pt}) &= \sigma_P Std\left(\frac{\mathrm{IC}_t}{\sqrt{(1-\mathrm{IC}_t^2-\sigma_{\mathrm{IC}}^2)/N+\sigma_{\mathrm{IC}}^2}}\right) \\
&\approx \sigma_P Std(\mathrm{IC}_t)/\sigma_{\mathrm{IC}} \qquad \text{when } N \text{ is large} \\
&= \sigma_P\ .
\end{aligned} \tag{45}
$$

The ex post realized portfolio IR is then

$$
\begin{aligned}
\mathrm{IR} &= \frac{\bar{\alpha}_{Pt}}{Std(\alpha_{Pt})} \\
&\approx \frac{\mathrm{IC}}{\sigma_{\mathrm{IC}}}.
\end{aligned} \tag{46}
$$

The approximation holds when N is large. Equation (46) is the same as the *ex post* IR formula derived by Qian and Hua (2004).

The interesting extreme case comes when $\sigma_{\mathrm{IC}} = 0$, i.e., the true IC_t is a constant over time as assumed by Grinold (1989) and Clarke *et al* (2002). Then the differences among the *ex post* estimated IC_t are purely a result of sample estimation error. As N gets larger and larger, one gets a more and more precise estimate for IC and the investment risk becomes smaller and smaller. The strategy ultimately becomes a money machine when N is large enough. As discussed in Qian, Hua and Sorensen (2007, p96), the quantity $\sqrt{(1-\mathrm{IC}^2)/N}$ is the standard error of the sample correlation coefficient with a sample of size N. So Equations (44) and (45) become

$$
\begin{aligned}
\bar{\alpha}_{Pt} &= \sigma_P \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{IC}_t}{\sqrt{(1-\mathrm{IC}_t^2)/N}} \\
&\approx \sigma_P \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{IC}_t}{\sqrt{(1-\mathrm{IC}^2)/N}} \\
&= \sigma_P \frac{\overline{\mathrm{IC}}}{\hat{\sigma}_{\mathrm{IC}}}
\end{aligned} \tag{47}
$$

and

$$Std(\alpha_{Pt}) = \sigma_P Std\left(\frac{\mathrm{IC}_t}{\sqrt{(1-\mathrm{IC}_t^2)/N}} \right)$$
$$\approx \sigma_P \frac{\hat{\sigma}_{\mathrm{IC}}}{\sqrt{(1-\mathrm{IC}^2)/N}} \tag{48}$$
$$= \sigma_P \; .$$

So the portfolio excess return mean and standard deviation estimates here still give

$$\mathrm{IR} \approx \frac{\overline{\alpha}_{Pt}}{Std(\alpha_{Pt})} = \frac{\overline{\mathrm{IC}}}{\hat{\sigma}_{\mathrm{IC}}} \; . \tag{49}$$

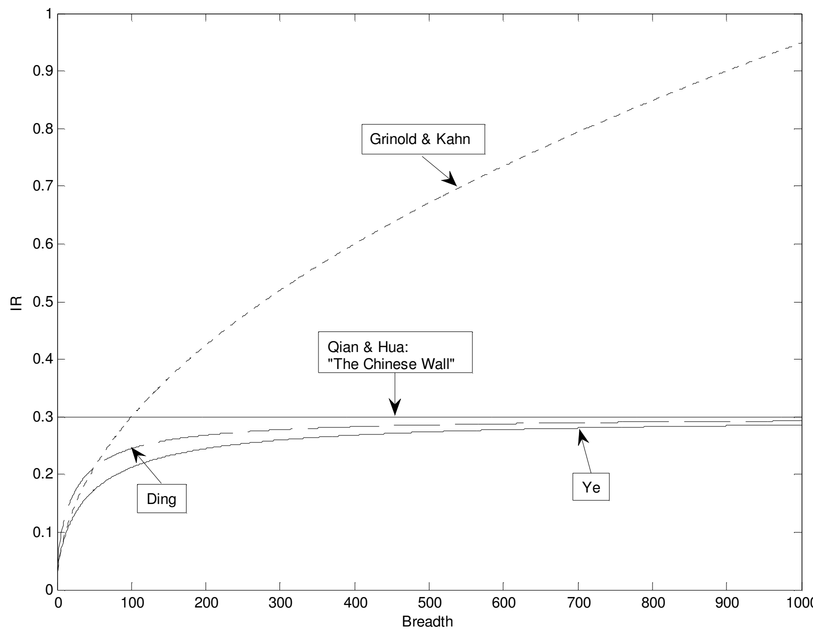
I used $\overline{\mathrm{IC}}$, $\hat{\sigma}_{\mathrm{IC}}$ to distinguish the sample mean and standard deviation from the population values for this special case. The results here show that the *ex post* portfolio excess return is proportional to targeted portfolio tracking error, σ_P, i.e., the more risk one takes, the more return one gets. This is consistent with the fundamentals of financial economics. The *ex post* portfolio excess return is also positively related to one's skill that is represented by the average IC one can achieve, and inversely related to the volatility of the skill, σ_{IC}, i.e., the more volatile the skill, the less excess return one can get. The result also shows that when the risk model, which is represented by the conditional covariance matrix of the forecasting errors, is correctly specified, then the *ex post* realized portfolio tracking error should be very close to the *ex ante* target tracking error one sets.

Figure 3 plots the relationship between portfolio IR and breadth *N* for various forms of the fundamental law discussed above. The parameters are assumed to be IC=0.03, $\sigma_{\mathrm{IC}} = 0.1$ and $\phi = 2$. The portfolio IR based on the Grinold fundamental law increases at the rate of the square root of breadth *N*. As the breadth increases, the portfolio IR will increase without a limit. According to our analysis above, this is true if the manager can pick stocks consistently at certain skill level (so that the cross-sectional IC is a constant over time). In reality, this is hardly the case. A forecast signal's IC changes constantly over time, and $\sigma_{\mathrm{IC}} \neq 0$. Under this more realistic situation, the fundamental law by Qian and Hua (2004) sets a "Chinese Wall" as the limit one can achieve. According to Qian and Hua, as long as $\mathrm{IC}/\sigma_{\mathrm{IC}}$ does not improve, one will not be able to improve the performance even if the breadth increases.

The fundamental law by Ye (2008) bridges the gap between Grinold's original formula and Qian and Hua's limit formula. At the limit as $N \to \infty$, it collapses to Qian and Hua's formula. The *ex ante* IR we derived in Equation (42) is more realistic than Ye's calculation in that it allows the residual returns to have different standard deviations. It can be seen that our IR calculation is higher than Ye's but lower than Qian and Hua's.

Figure 3. Various Forms of the Fundamental Law

(IC=0.03, $\sigma_{\text{IC}} = 0.1, \phi = 2$)



Our discussion above shows that the marginal contribution of breadth (*N*) on portfolio IR diminishes as *N* increases. Here we are using the number of stocks in the selection universe as breadth, which may not be the same as what Grinold uses for breadth in his original paper. Grinold (1989) gives a quite lengthy discussion on the importance of independent bets when determining what *N* is. For example, one should not count two dependent bets as different bets. In practice, it is quite difficult to quantify dependent bets and to make appropriate adjustments. The formula in (42) shows that even if *N* increases, the portfolio IR will not improve much for a typical investment universe of 1000 or 2000 stocks as long as the average IC and volatility of IC stay the same. The important thing is to play often (try to increase *N*) when *N* is small but to play precisely (low σ_{IC}) and to play well (high IC) when N is already large.

In Figure 3, we assumed ϕ to be a constant over time. In reality, it is well known that stock returns exhibit heteroskedasticity so ϕ will be time varying too. Figure 4 shows the estimated ϕ values for Russell 1000, 2000 and 3000 universe from 1978:12 to 2009:08 assuming an IC of 0.03 and σ_{IC} of 0.1. We can observe the following:

1) ϕ is time varying,
2) usually the bigger the sample size, the larger the ϕ is,
3) the minimum value of ϕ is around 1.5, and during most times ϕ is within the range of (1.5, 2),
4) there was a dramatic bubble-burst period for ϕ during the tech bubble time of 1999 to 2002.

Figure 4. ϕ Values for Different Universes over Time

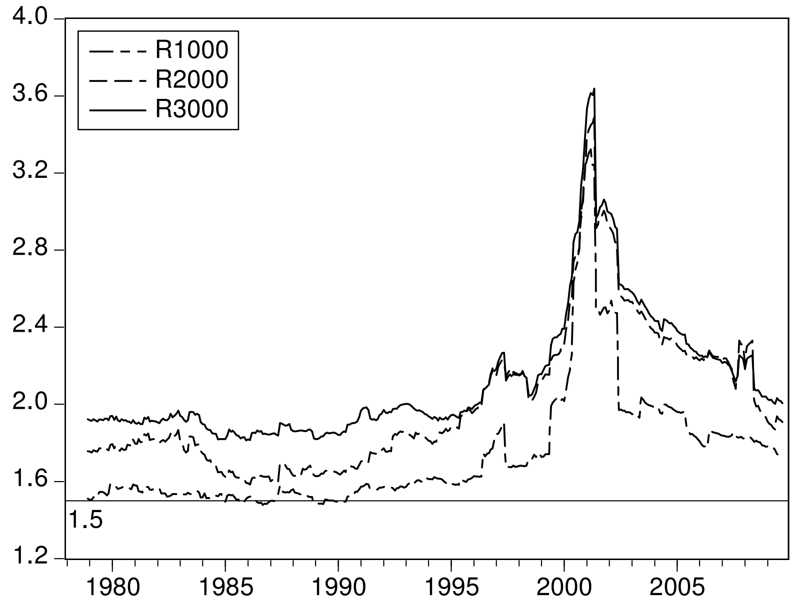


Table 2 shows the average number of stocks ($\overline{N}$), average ϕ ($\overline{\phi}$), $\overline{N}\overline{\phi}$, and $1/\sqrt{\overline{N}\overline{\phi}}$ for Russell universes of stocks. It will be seen later that for most quantitative factors people use, $1/\sqrt{\overline{N}\overline{\phi}}$ is much smaller than the factor IC standard deviation, which suggests that for the most commonly used investment universes the Grinold factor ($1/\sqrt{N}$) has a much smaller impact than the Qian and Hua factor (σ_{IC}). This is also obvious from Figure 3.

Table 2. Average Number of Companies and ϕ for Russell Indices
(1978:12-2009:8)

Index	$\overline{N}$	$\overline{\phi}$	$\overline{N}\overline{\phi}$	$1/\sqrt{\overline{N}\overline{\phi}}$
Russell 1000	949	1.74	1653	0.025
Russell 1000 Growth	507	1.66	843	0.034
Russell 1000 Value	578	1.57	908	0.033
Russell 2000	1833	2.01	3685	0.016
Russell 2000 Growth	1249	1.88	2347	0.021
Russell 2000 Value	1300	1.90	2465	0.020
Russell 3000	2782	2.12	5903	0.013
Russell 3000 Growth	1756	1.95	3425	0.017
Russell 3000 Value	1878	1.99	3729	0.016

Robustness of the Fundamental Law to Model Specification

In deriving the generalized fundamental law in Equation (42), we assumed the true relationship to be a linear one factor model between the residual return and the forecast signal. The residual returns are not risk-adjusted. The cross-sectional heteroskedasticity across residual returns resulted in the ϕ parameter in Equation (42). In practice, people may use risk-adjusted residual return as dependent variable to correct for the cross-sectional heteroskedasticity, i.e.,

$$\tilde{r}_{it} = r_{it} / \sigma_{r_{it}} = \tilde{f}_t z_{it-1} + \tilde{\varepsilon}_{it}, \tag{50}$$

where $\tilde{r}_{it}$ is the risk-adjusted residual return, $\sigma_{r_{it}}$ is the conditional volatility for residual return r_{it} as of time *t*, $\tilde{f}_t$ is the cross-sectional factor return at time *t* (which will be different from the factor return in Equation (36)), z_{it-1} is the factor exposure that has both time series and cross-sectional mean 0 and standard deviation 1, $\tilde{\varepsilon}_{it} \sim N(0, \tilde{\sigma}_{\varepsilon_i}^2)$ is the idiosyncratic noise that cannot be predicted. Under these assumptions, we will have the cross-sectional IC between risk-adjusted residual return $\tilde{r}_{it}$ and z_{it-1} to be the same as $\tilde{f}_t$, i.e.,

$$\widetilde{\mathrm{IC}}_t = \mathrm{corr}(\tilde{r}_{it}, z_{it-1}) = \tilde{f}_t, \tag{51}$$

and

$$\tilde{\sigma}_{\varepsilon_i}^2 = 1 - \widetilde{\mathrm{IC}}^2 - \tilde{\sigma}_{\mathrm{IC}}^2, \tag{52}$$

where $\widetilde{\mathrm{IC}}$ and $\tilde{\sigma}_{\mathrm{IC}}$ are the mean and standard deviation of $\widetilde{\mathrm{IC}}_t$. By using the same algebra in the previous section, we can get

$$\widetilde{\mathrm{IR}} = \frac{\widetilde{\mathrm{IC}}}{\sqrt{(1 - \widetilde{\mathrm{IC}}^2 - \tilde{\sigma}_{\mathrm{IC}}^2)/N + \tilde{\sigma}_{\mathrm{IC}}^2}}. \tag{53}$$

The formula is identical to Equation (42) when $\phi = 1/(1 - \mathrm{IC}^2 - \sigma_{\mathrm{IC}}^2)$, i.e., when the residual standard deviations are the same across all the securities. One thing we have to be aware of is $\widetilde{\mathrm{IC}}$ and $\tilde{\sigma}_{\mathrm{IC}}^2$ in Equation (53) will be different from IC and σ_{IC}^2 in Equation (42).

The above discussion shows that the form of the fundamental law is quite robust to the forecast model specification. In both cases, the most important impact to portfolio IR is the IC volatility over time. One insight from Equations (42) and (53) is that a quant manager should preprocess the residual returns and factor exposures in such a way so that the resulting cross-sectional IC will have a higher average and lower standard deviation. One disadvantage with the model specification in Equation (50) is that one has to estimate the conditional volatility $\sigma_{r_{it}}$ which can involve estimation errors. A GARCH type model will be useful for this purpose.

Multifactor Fundamental Law and the Impact of Missing Factors

The fundamental law we discussed so far only concerns one factor. In practice, analysts or portfolio managers rarely use only one factor. Residual return forecast almost always involves multiple factors. It will be interesting to see the form of fundamental law with multiple factors and study the consequences of missing one or more factors in modeling. In deriving the fundamental laws presented in previous sections, we either made the assumption that the residual return dispersion is a constant over time or used the risk-adjusted residual return in analysis. But this is not necessary if we work on residual security returns and factor returns directly.

If we assume residual returns follow a linear relationship with factor exposures

$$\mathbf{r}_t = \mathbf{Z}_{t-1}\mathbf{F}_t + \boldsymbol{\varepsilon}_t , \tag{54}$$

where $\mathbf{r}_t$ is an $N \times 1$ vector of residual returns, $\mathbf{Z}_{t-1}$ is an $N \times K$ matrix of factor exposures, $\mathbf{F}_t$ is a $K \times 1$ vector of factor returns, and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector of idiosyncratic noise. It is shown in Appendix B under some weak regularity conditions that the *ex ante* expected portfolio IR has the following relationship with the expected factor return (**F**) and factor return covariance ($\boldsymbol{\Sigma}_\mathrm{F}$)

$$\begin{aligned} \mathrm{IR} &= \sqrt{\mathbf{F}'(1/(\tau N)\mathbf{I} + \boldsymbol{\Sigma}_\mathrm{F})^{-1}\mathbf{F}} \\ &\approx \sqrt{\mathbf{F}'\boldsymbol{\Sigma}_\mathrm{F}^{-1}\mathbf{F}} \end{aligned} \tag{55}$$

where $\tau = E_{cs}(1/\sigma_{\varepsilon_i}^2)$ represents part of the risk related to idiosyncratic noise. As in the univariate case, this part of the risk will be diversified away as *N* gets larger, and the remaining dominant risk is the "strategy risk" represented by the factor return covariance that cannot be diversified away. When there is only one factor, Equation (55) reduces to

$$\begin{aligned} \mathrm{IR} &= \frac{f}{\sqrt{(E_{cs}(1/\sigma_{\varepsilon_i}^2))^{-1}/N + \sigma_f^2}} \\ &\approx \frac{f}{\sigma_f} . \end{aligned} \tag{56}$$

So the expected portfolio IR is just the IR of the factor-mimicking portfolio.

If, instead of using the raw residual return in Equation (54), we use the risk-adjusted residual returns, then the multi-factor fundamental law in Equation (55) becomes (see Appendix B)

$$\begin{aligned} \mathrm{IR} &= \sqrt{\mathbf{IC}'(\sigma_\varepsilon^2 / N\,\mathbf{I} + \boldsymbol{\Sigma}_\mathrm{IC})^{-1}\mathbf{IC}} \\ &\approx \sqrt{\mathbf{IC}'\boldsymbol{\Sigma}_\mathrm{IC}^{-1}\mathbf{IC}} , \end{aligned} \tag{57}$$

where $\sigma_\varepsilon^2 = (1 - \sum_{k=1}^{K}(\sigma_{\mathrm{IC},k}^2 + \mathrm{IC}_k^2))$ is the variance for idiosyncratic noise, $\mathbf{IC}$ is the cross-sectional correlation vector between factor exposures and risk-adjusted residual returns, and $\boldsymbol{\Sigma}_\mathrm{IC}$ is the factor IC covariance matrix. Equation (57) reduces to Equation (53) when there is only one factor.

The above conclusion is based on the assumption that the model is correctly specified which is almost surely not the case in practice. A natural question to ask is what happens if the return or risk model is mis-specified. With the fundamental law in multi-factor format, we can easily study the impact of missing one or more return or risk factors. For ease of exposition, I will only present the analysis for a 2-factor system here. More detailed analysis with missing multiple factors can be found in Appendix B. In the analysis below, I will not purposely distinguish risk factors from alpha factors.

Statistically, the only difference should be that the expected IC (or factor return) for risk factor is zero while that for alpha factor is different from zero.

For a 2-factor system, Equation (B15) reduces to

$$
\begin{aligned}
\mathrm{IR} &\approx \sqrt{\frac{1}{1-\rho^2_{\mathrm{IC}_1,\mathrm{IC}_2}}\left(\left(\frac{\mathrm{IC}_1}{\sigma_{\mathrm{IC}_1}}\right)^2+\left(\frac{\mathrm{IC}_2}{\sigma_{\mathrm{IC}_2}}\right)^2-2\left(\frac{\mathrm{IC}_1}{\sigma_{\mathrm{IC}_1}}\frac{\mathrm{IC}_2}{\sigma_{\mathrm{IC}_2}}\right)\rho_{\mathrm{IC}_1,\mathrm{IC}_2}\right)} \\
&= \sqrt{\left(\frac{\mathrm{IC}_1}{\sigma_{\mathrm{IC}_1}}\right)^2+\frac{1}{1-\rho^2_{\mathrm{IC}_1,\mathrm{IC}_2}}\left(\frac{\mathrm{IC}_2}{\sigma_{\mathrm{IC}_2}}-\rho_{\mathrm{IC}_1,\mathrm{IC}_2}\frac{\mathrm{IC}_1}{\sigma_{\mathrm{IC}_1}}\right)^2} \qquad (58) \\
&\geq \frac{\mathrm{IC}_1}{\sigma_{\mathrm{IC}_1}}.
\end{aligned}
$$

where $\rho_{\mathrm{IC}_1,\mathrm{IC}_2}$ is the time series correlation of the two factor ICs.

From Equation (58), it is clear that a mis-specified model, whether it is mis-specified in the return forecast part or the risk forecast part, will almost always hurt the performance. For a missing return factor, the adverse impact comes from both the missing return forecast, IC_2, and the resulting conditional covariance mis-specification, $(1-\rho^2_{\mathrm{IC}_1,\mathrm{IC}_2})$. For a missing risk factor, the adverse impact only comes from the resulting conditional covariance mis-specification $(1-\rho^2_{\mathrm{IC}_1,\mathrm{IC}_2})$. This is not surprising indeed! The only exception is when the missing factor is a risk factor and the risk factor IC is not time-series correlated with the return factor IC (i.e. when $\mathrm{IC}_2=0$ and $\rho_{\mathrm{IC}_1,\mathrm{IC}_2}=0$). When the risk factor is missing, the *ex post* realized portfolio tracking error will be larger than the *ex ante* targeted portfolio tracking error by a factor of $1/\sqrt{1-\rho^2_{\mathrm{IC}_1,\mathrm{IC}_2}}\geq 1$. So if $\rho_{\mathrm{IC}_1,\mathrm{IC}_2}$ is small, then the impact of missing a risk factor is small.

Fundamental Law with Transfer Coefficient

Clarke *et al.* (2002) proposed the concept of "transfer coefficient" to incorporate the impact of additional constraints into the fundamental law. They define the transfer coefficient as the cross-sectional correlation coefficient between the residual return volatility adjusted active weights and alphas

$$
\begin{aligned}
\mathrm{TC} &= \mathrm{corr}(\Delta\tilde{w}_{it}\sigma_{r_i},\alpha_{it}/\sigma_{r_i}) \\
&= \frac{\mathrm{cov}(\Delta\tilde{w}_{it}\sigma_{r_i},\alpha_{it}/\sigma_{r_i})}{d(\Delta\tilde{w}_{it}\sigma_{r_i})d(\alpha_{it}/\sigma_{r_i})}. \qquad (59)
\end{aligned}
$$

This definition has the desired property of measuring the impact of constraints on portfolio IR when the factor IC is a constant so that $\sigma_{\mathrm{IC}}=0$ and the residual return covariance is a diagonal matrix. Under this assumption, the transfer coefficient is the ratio of the constrained portfolio IR and the unconstrained optimal portfolio IR

$$\tilde{\mathrm{IR}} = \mathrm{TC}\,\mathrm{IR}\,, \tag{60}$$

so the transfer coefficient does represent the portion of optimal portfolio IR that can be transferred into the constrained portfolio.

Ye (2008) extended the transfer coefficient into her version of fundamental law with time varying IC. Using her approach, she got the following relationship

$$\tilde{\mathrm{IR}} = \frac{\mathrm{IC}}{\sqrt{1/(N\,\mathrm{TC}^2) + \sigma_{\mathbf{IC}}^2}}\,. \tag{61}$$

One surprising observation from Equation (61) is that the transfer coefficient as derived by Ye (2008) will have diminishing impact as breadth *N* increases. The constrained portfolio IR will approach the unconstrained optimal portfolio IR as *N* increases (both approach $\mathrm{IC}/\sigma_{\mathrm{IC}}$ as $N \to \infty$) no matter what constraints one imposes on the portfolio. This conclusion is quite counter-intuitive to practitioners as it can lead one to believe that any portfolio can have the same IR.

So why does this happen? When the cross-sectional IC is time varying as discussed in Ye (2008) and this paper, the total risk of the residual return is no longer a diagonal covariance matrix. In fact the majority risk comes from the strategy risk which causes the off-diagonal elements of the conditional covariance matrix to be non-zero. The transfer coefficient will not have the desired property if we only use the diagonal portion of the conditional covariance matrix to adjust the weights and alphas in deriving the transfer coefficient. Under this more practical situation, the transfer coefficient needs to be redefined using the total risk adjusted active weights and alphas as follows:

$$\mathrm{TC} = \frac{\Delta\tilde{\mathbf{w}}_t{}'\boldsymbol{\alpha}_t}{\sqrt{\Delta\tilde{\mathbf{w}}_t{}'\boldsymbol{\Omega}_t\Delta\tilde{\mathbf{w}}_t}\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t}}\,, \tag{62}$$

where $\Delta\tilde{\mathbf{w}}_t$ is the active weights of the constrained portfolio. Using this modified transfer coefficient definition, we get the constrained portfolio's expected excess return as,

$$\begin{aligned}\tilde{\alpha}_{Pt} &= \Delta\tilde{\mathbf{w}}_t{}'\boldsymbol{\alpha}_t \\ &= \Delta\tilde{\mathbf{w}}_t{}'\boldsymbol{\Omega}_t^{1/2}\boldsymbol{\Omega}_t^{-1/2}\boldsymbol{\alpha}_t \\ &= \mathrm{Corr}(\boldsymbol{\Omega}_t^{1/2}\Delta\tilde{\mathbf{w}}_t, \boldsymbol{\Omega}_t^{-1/2}\boldsymbol{\alpha}_t)\sqrt{\Delta\tilde{\mathbf{w}}_t{}'\boldsymbol{\Omega}_t\Delta\tilde{\mathbf{w}}_t}\sqrt{\boldsymbol{\alpha}_t{}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t} \\ &= \mathrm{TC}\,\sigma_{Pt}\mathrm{IR}\,,\end{aligned} \tag{63}$$

where σ_{Pt} is the targeted portfolio tracking error and IR is the information ratio for the unconstrained optimal portfolio. So the constrained portfolio information ratio ($\tilde{\mathrm{IR}}$), the transfer coefficient (TC) and the optimal unconstrained portfolio information ratio (IR) have the following relationship

$$\begin{aligned}\tilde{\mathrm{IR}} &= \alpha_{Pt}/\sigma_{Pt} \\ &= \mathrm{TC}\,\mathrm{IR}\,.\end{aligned} \tag{64}$$

The impact of the constraints on portfolio IR will be the same as in Clarke *et al.*'s (2002) original definition. In this way, a transfer coefficient of 0.5 will reduce the portfolio IR by 50% from the unconstrained optimal level.

Empirical Factor IR Comparison

In order to compare the differences between the different forms of the fundamental law, I calculated the IR that can be achieved by various quantitative factors using different formulas. For each factor, I calculate the *ex post* realized cross-sectional correlation (IC) between lagged factor exposures and residual returns, and then calculate the mean and standard deviation of the time series IC. The results are then substituted into various formulas to generate Table 3. For all the factors considered here, σ_{IC} is much more important than $1/\sqrt{\overline{N}\overline{\phi}}$. I calculated $\sigma_{\mathrm{IC}}\sqrt{\overline{N}\overline{\phi}}$ for each factor and they are in the range of 4 to 10 which means σ_{IC} is 4 to 10 times more important than $1/\sqrt{\overline{N}\overline{\phi}}$. From the last four columns of the table, we can see that the expected IR from the Grinold formula is always much higher than the other three while the other three stay very close to each other. This is not surprising given the result in Figure 3 and the above discussion.

Table 3. Factor IR Comparison (monthly, data ends 2009:8)

Factor	Index	$\frac{1}{\sqrt{\overline{N}\overline{\phi}}}$	IC Mean	IC Stdev σ_{IC}	$\frac{\sigma_{IC}}{1/\sqrt{\overline{N}\overline{\phi}}}$	IR GK	IR QH	IR YE	IR DING
Book to Price	R1000	0.024	0.014	0.139	5.67	0.44	0.10	0.10	0.10
	R2000	0.016	0.025	0.113	6.95	1.12	0.22	0.22	0.22
	R3000	0.013	0.020	0.114	8.76	1.06	0.17	0.17	0.17
Cash Flow to Price	R1000	0.024	0.039	0.119	4.88	1.21	0.33	0.32	0.32
	R2000	0.016	0.066	0.122	7.49	2.93	0.54	0.53	0.54
	R3000	0.013	0.058	0.111	8.59	3.17	0.52	0.52	0.52
Earnings to Price	R1000	0.024	0.031	0.140	5.70	0.95	0.22	0.21	0.22
	R2000	0.016	0.067	0.120	7.37	2.96	0.56	0.55	0.55
	R3000	0.013	0.059	0.121	9.35	3.19	0.48	0.48	0.48
Sales to Price	R1000	0.024	0.019	0.129	5.26	0.58	0.15	0.14	0.14
	R2000	0.016	0.026	0.104	6.41	1.16	0.25	0.24	0.25
	R3000	0.013	0.023	0.107	8.22	1.25	0.22	0.21	0.21
12-Month Momentum	R1000	0.024	0.029	0.179	7.31	0.91	0.16	0.16	0.16
	R2000	0.016	0.055	0.128	7.86	2.46	0.43	0.43	0.43
	R3000	0.013	0.049	0.137	10.59	2.68	0.36	0.36	0.36
Share Repurchase	R1000	0.024	0.015	0.089	3.63	0.46	0.17	0.16	0.16
	R2000	0.016	0.026	0.084	5.20	1.16	0.31	0.30	0.30
	R3000	0.013	0.024	0.083	6.38	1.30	0.29	0.28	0.29
Percent Short	R1000	0.024	0.022	0.118	4.81	0.67	0.18	0.18	0.18
	R2000	0.016	0.037	0.105	6.48	1.67	0.36	0.35	0.35
	R3000	0.013	0.029	0.101	7.80	1.56	0.28	0.28	0.28

Empirical findings here show that the theoretically calculated IR number from Grinold's fundamental law needs to be cut by much more than half to be realistic. For a typical investment universe of 1000 or 2000 stocks, the empirically calculated IR numbers from formulas derived by Qian and Hua (2004), Ye (2008) and this paper give a more realistic estimate of achievable IR. For investment universes less than 500, an IR using the formula derived in this paper will give a better estimate. The difference will become more significant for investment strategies with a much smaller selection universe, such as a global macro strategy, or a tactical asset allocation strategy. The idiosyncratic risk still plays a role when *N* is small. Table 4 shows theoretical examples when the investable universes have much less choices.

Table 4. Theoretical IR Comparison when N is Small

		GK	QH	YE	DING	GK	QH	YE	DING
IC	σ_{IC}	N=10				N=50			
0.10	0.10	0.32	1.00	0.30	0.41	0.32	1.00	0.58	0.71
	0.15	0.32	0.67	0.29	0.37	0.32	0.67	0.49	0.55
	0.20	0.32	0.50	0.27	0.33	0.32	0.50	0.41	0.45
0.15	0.10	0.47	1.50	0.45	0.61	0.47	1.50	0.87	1.06
	0.15	0.47	1.00	0.43	0.56	0.47	1.00	0.73	0.83
	0.20	0.47	0.75	0.40	0.50	0.47	0.75	0.61	0.67
		N=100				N=200			
0.05	0.10	0.50	0.50	0.35	0.41	0.50	0.50	0.41	0.45
	0.15	0.50	0.33	0.28	0.30	0.50	0.33	0.30	0.32
	0.20	0.50	0.25	0.22	0.24	0.50	0.25	0.24	0.24
0.10	0.10	1.00	1.00	0.71	0.82	1.41	1.00	0.82	0.89
	0.15	1.00	0.67	0.55	0.60	1.41	0.67	0.60	0.63
	0.20	1.00	0.50	0.45	0.47	1.41	0.50	0.47	0.49

Conclusion

I have derived a generalized version of the fundamental law of active management under some weak assumptions. The original fundamental law of Grinold (1989), the generalized fundamental laws of Clarke *et al*. (2002), Qian and Hua (2004), and Ye (2008) are all special cases of the fundamental law derived in this paper. I show that cross-sectional ICs are usually different from time series ICs, and they will be the same only under the strong assumption that either the residual return volatilities are the same across all the securities or the ICs are calculated using risk-adjusted residual returns with the forecast signal.

I also show that the form of the fundamental law derived in this paper is quite robust to forecast model specification. According to our generalized fundamental law, the variation in IC (IC volatility over time) has a much bigger impact to portfolio IR than the breadth *N* for a typical investment universe. The fundamental law by Qian and Hua (2004) sets a "Chinese Wall" as the upper limit for the portfolio IR a portfolio manager can reach when the cross-sectional IC varies over time. The fundamental law by Grinold (1989) is derived under some unrealistic assumptions and always overestimates by a large margin the IR a portfolio manager can actually reach. I extend the fundamental law to models with multiple factors and study the impact of missing one or more return or risk factors. It is shown that a mis-specified model, whether it is mis-specified in the return forecast part or risk forecast part, will almost always hurt performance. The exception occurs when a missing risk factor (IC=0) has a zero time series IC correlation with all the other factors. For the commonly used quantitative return and risk factors, I found that the impact of a missing risk factor is usually small.

Our results also show that the transfer coefficient as originally defined by Clarke *et al.* (2002) is not able to capture the impact of constraints to portfolio IR in the presence of IC variation. One will get the wrong conclusion that portfolio constraints do not have much impact on portfolio IR in the presence of IC variation when N is large. I redefine the concept of transfer coefficient using the cross-sectional correlation between the total conditional covariance adjusted weights and alphas. The modified transfer coefficient captures the impact of portfolio constraints on portfolio IR as desired.

One insight from this paper is that portfolio managers should try to play well (high IC) and play precisely (low σ_{IC}). Extra efforts should be made to process the information and to build models that can increase IC and reduce IC variation.

I thank Xiaohong Chen, Roger Clarke, Russell Fuller, Tom Fuller, John Kling, Doug Stone, Wei Su, Yixiao Sun, Yining Tung, Jia Ye, and two anonymous referees for helpful discussions and comments. Richard Grinold provided me with his original technical notes. Yining Tung helped with some empirical calculations in the paper.

Appendix A

Given the conditional forecasting error covariance matrix in Equation (39) and based on the Woodbury matrix identity, we have the inverse matrix of $\mathbf{\Omega}_t$ as

$$\mathbf{\Omega}_t^{-1} = \mathbf{\Sigma}_t^{-1} - \varphi \mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}, \tag{A1}$$

where

$$\varphi = \frac{\sigma_{\mathrm{IC}}^2\delta^2}{1+\sigma_{\mathrm{IC}}^2\delta^2\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}}. \tag{A2}$$

Substituting (A1) into Equation (15) we have

$$\begin{aligned}
\alpha_{Pt} &= \sigma_{Pt}\sqrt{\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}(\boldsymbol{\alpha}_t-\kappa\mathbf{1})} \\
&= \sigma_{Pt}\sqrt{\boldsymbol{\alpha}_t'(\mathbf{\Sigma}_t^{-1}-\varphi\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1})(\boldsymbol{\alpha}_t-\kappa\mathbf{1})} \\
&= \sigma_{Pt}\sqrt{\boldsymbol{\alpha}_t'(\mathbf{\Sigma}_t^{-1}-\varphi\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1})\boldsymbol{\alpha}_t-\kappa\boldsymbol{\alpha}_t'(\mathbf{\Sigma}_t^{-1}-\varphi\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1})\mathbf{1}} \\
&= \sigma_{Pt}\sqrt{\boldsymbol{\alpha}_t'\mathbf{\Sigma}_t^{-1}\boldsymbol{\alpha}_t-\varphi\boldsymbol{\alpha}_t'\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\boldsymbol{\alpha}_t-\kappa\boldsymbol{\alpha}_t'(\mathbf{\Sigma}_t^{-1}-\varphi\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1})\mathbf{1}} \\
&= \sigma_{Pt}\,\mathrm{IC}\delta\sqrt{(\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}-\kappa\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{1}/(\mathrm{IC}\delta))(1-\varphi\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1})} \\
&= \sigma_{Pt}\,\mathrm{IC}\delta\sqrt{(\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1}-\kappa\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{1}/(\mathrm{IC}\delta))/(1+\sigma_{\mathrm{IC}}^2\delta^2\mathbf{z}_{t-1}'\mathbf{\Sigma}_t^{-1}\mathbf{z}_{t-1})}\ .
\end{aligned} \tag{A3}$$

When σ_{ε_i}, z_{it-1} are cross-sectionally independent, then as N becomes large we have

$$
\begin{aligned}
\alpha_{Pt} &= \sigma_{Pt}\mathrm{IC}\delta\sqrt{\left(\sum_{i=1}^{N}(z_{it-1}/\sigma_{\varepsilon_i})^2 - \frac{\kappa}{\mathrm{IC}\delta}\sum_{i=1}^{N} z_{it-1}/\sigma_{\varepsilon_i}^2\right)\left(1+\sigma_{\mathrm{IC}}^2\delta^2\sum_{i=1}^{N}(z_{it-1}/\sigma_{\varepsilon_i})^2\right)^{-1}} \\
&= \sigma_{Pt}\mathrm{IC}\delta\sqrt{\sum_{i=1}^{N}(z_{it-1}/\sigma_{\varepsilon_i})^2 \Big/ \left(1+\sigma_{\mathrm{IC}}^2\delta^2\sum_{i=1}^{N}(z_{it-1}/\sigma_{\varepsilon_i})^2\right)} \\
&= \sigma_{Pt}\mathrm{IC}\sqrt{N}\delta\sqrt{E_{cs}(z_{it-1}/\sigma_{\varepsilon_i})^2 / (1+N\sigma_{\mathrm{IC}}^2\delta^2 E_{cs}(z_{it-1}/\sigma_{\varepsilon_i})^2)} \\
&= \sigma_{Pt}\mathrm{IC}\sqrt{N}\delta\sqrt{E_{cs}(z_{it-1}^2)E_{cs}(1/\sigma_{\varepsilon_i}^2) / (1+N\sigma_{\mathrm{IC}}^2\delta^2 E_{cs}(z_{it-1}^2)E_{cs}(1/\sigma_{\varepsilon_i}^2))} \\
&= \sigma_{Pt}\mathrm{IC}\sqrt{N}\sqrt{\delta^2 E_{cs}(1/\sigma_{\varepsilon_i}^2) / (1+N\sigma_{\mathrm{IC}}^2\delta^2 E_{cs}(1/\sigma_{\varepsilon_i}^2))} \\
&= \sigma_{Pt}\frac{\mathrm{IC}}{\sqrt{1/(\delta^2 E_{cs}(1/\sigma_{\varepsilon_i}^2)N) + \sigma_{\mathrm{IC}}^2}} \\
&= \sigma_{Pt}\frac{\mathrm{IC}}{\sqrt{1/(\phi N) + \sigma_{\mathrm{IC}}^2}},
\end{aligned}
\tag{A4}
$$

where

$$
\begin{aligned}
\phi &= \frac{1}{N}\sum_{i=1}^{N}\sigma_{r_i}^2 \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_{\varepsilon_i}^2} \\
&= \frac{1}{N}\sum_{i=1}^{N}\sigma_{r_i}^2 \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_{r_i}^2 - (\mathrm{IC}^2+\sigma_{\mathrm{IC}}^2)\delta^2} \\
&\geq \frac{1}{N}\sum_{i=1}^{N}\sigma_{r_i}^2 \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_{r_i}^2} \\
&\geq 1.
\end{aligned}
\tag{A5}
$$

The last line in (A5) is based on Jensen's inequality. In the derivation we used the fact that $\frac{1}{N}\sum_{i=1}^{N} z_{it-1}/\sigma_{\varepsilon_i}^2 = 0$ when $N \to \infty$ since z_{it-1} and σ_{ε_i} are cross-sectionally independent by assumption.

Appendix B

Assume residual security returns $\mathbf{r}_t$ and security factor exposures $\mathbf{Z}_{t-1}$ are related through a linear factor model as follows

$$
\mathbf{r}_t = \mathbf{Z}_{t-1}\mathbf{F}_t + \boldsymbol{\varepsilon}_t, \tag{B1}
$$

where $\mathbf{r}_t$ is an $N \times 1$ vector of residual returns, $\mathbf{Z}_{t-1}$ is an $N \times K$ matrix of factor exposures that become known at the end of time t-1, $\mathbf{F}_t$ is a $K \times 1$ vector of factor returns, and $\boldsymbol{\varepsilon}_t \mid \mathrm{I}_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ is an $N \times 1$ vector of idiosyncratic noise with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_\varepsilon = diag(\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2, \cdots, \sigma_{\varepsilon_N}^2)$. The factor exposures are normalized to have both time series and cross-sectional mean 0 and standard deviation 1**,** and are cross-sectionally orthogonal to each other so that $\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}/N=\mathbf{I}$, Other regularity assumptions like those in C1) and C2) also apply. We further assume that factor returns follow a multivariate normal distribution

$$\mathbf{F}_t \mid \mathrm{I}_{t-1} \sim N(\mathbf{F},\ \mathbf{\Sigma}_{\mathrm{F}})\,. \tag{B2}$$

Based on the above assumptions, we have

$$\boldsymbol{\alpha}_t=\mathbf{Z}_{t-1}\mathbf{F}\,, \tag{B3}$$

and

$$\mathbf{\Omega}_t=\mathbf{Z}_{t-1}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{Z}_{t-1}'+\Sigma_\varepsilon\,. \tag{B4}$$

Applying Woodbury matrix identity, we get the inverse of the conditional covariance matrix as

$$\mathbf{\Omega}_t^{-1}=\Sigma_\varepsilon^{-1}-\Sigma_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\Sigma_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\Sigma_\varepsilon^{-1}\,. \tag{B5}$$

Substituting Equations (B3) and (B5) into the two components of the IR formula in Equation (16) we get

$$\begin{aligned}
\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\alpha}_t &= \mathbf{F}'\mathbf{Z}_{t-1}'(\mathbf{\Sigma}_\varepsilon^{-1}-\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1})\mathbf{Z}_{t-1}\mathbf{F}\\
&= \mathbf{F}'\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{I}-(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})\mathbf{F}\\
&= \mathbf{F}'\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{\Sigma}_{\mathrm{F}}^{-1}\mathbf{F}\\
&= \mathbf{F}'(\mathbf{\Sigma}_{\mathrm{F}}(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})(\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1})^{-1}\mathbf{F}\\
&= \mathbf{F}'((\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}+\mathbf{\Sigma}_{\mathrm{F}})^{-1}\mathbf{F}\\
&= \mathbf{F}'((\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}/N)^{-1}/N+\mathbf{\Sigma}_{\mathrm{F}})^{-1}\mathbf{F}\\
&= \mathbf{F}'(1/(\tau N)\,\mathbf{I}+\mathbf{\Sigma}_{\mathrm{F}})^{-1}\mathbf{F}
\end{aligned} \tag{B6}$$

and

$$\begin{aligned}
\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\mathbf{1} &= \mathbf{F}'\mathbf{Z}_{t-1}'(\mathbf{\Sigma}_\varepsilon^{-1}-\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1})\mathbf{1}\\
&= \mathbf{F}'(\mathbf{I}-\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathrm{F}}^{-1}+\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1})\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{1}\\
&= \mathbf{F}'(1/(\tau N)\,\mathbf{I}+\mathbf{\Sigma}_{\mathrm{F}})^{-1}\mathbf{Z}_{t-1}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{1}/(\tau N)\\
&= \mathbf{0}\,,
\end{aligned} \tag{B7}$$

where we assumed z_{ki} and σ_{ε_i} to be cross-sectionally independent and used the facts that for $k,l=1,2,\cdots,K$,

$$
\begin{aligned}
\mathbf{Z}_{k,t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{l,t-1}/N &= \frac{1}{N}\sum_{i=1}^{N}\frac{z_{k,it-1}z_{l,it-1}}{\sigma_{\varepsilon_i}^2} \\
&= E_{cs}\left(\frac{z_{k,it-1}z_{l,it-1}}{\sigma_{\varepsilon_i}^2}\right) \\
&= E_{cs}\left(z_{k,it-1}z_{l,it-1}\right)E_{cs}(1/\sigma_{\varepsilon_i}^2) \\
&= \begin{cases} E_{cs}(1/\sigma_{\varepsilon_i}^2) = \dfrac{1}{N}\sum_{i=1}^{N}(1/\sigma_{\varepsilon_i}^2) = \tau & \text{when } k = l \\ 0 & \text{when } k \neq l \end{cases}
\end{aligned} \tag{B8}
$$

and

$$
\begin{aligned}
\mathbf{Z}_{k,t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{1}/N &= \frac{1}{N}\sum_{i=1}^{N}(z_{k,it-1}/\sigma_{\varepsilon_i}^2) \\
&= E_{cs}(z_{k,it-1}/\sigma_{\varepsilon_i}^2) \\
&= E_{cs}(z_{k,it-1})E_{cs}(1/\sigma_{\varepsilon_i}^2) \\
&= 0\,.
\end{aligned} \tag{B9}
$$

So the *ex ante* expected portfolio IR is

$$
\begin{aligned}
\mathrm{IR} &= E\left(\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \mathbf{1})}\right) \\
&= E\left(\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t}\right) \\
&= \sqrt{\mathbf{F}'(1/(\tau N)\,\mathbf{I} + \boldsymbol{\Sigma}_{\mathrm{F}})^{-1}\mathbf{F}} \\
&\approx \sqrt{\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\mathbf{F}}\,.
\end{aligned} \tag{B10}
$$

For a one factor model, Equation (B10) simplifies to

$$
\begin{aligned}
\mathrm{IR} &= \frac{f}{\sqrt{\left(E_{cs}(1/\sigma_{\varepsilon_i}^2)\right)^{-1}/N + \sigma_f^2}} \\
&\approx \frac{f}{\sigma_f}\,,
\end{aligned} \tag{B11}
$$

i.e., the expected portfolio IR is just the IR of the factor-mimicking portfolio. When the cross-sectional residual return dispersion is a constant, i.e., $d(\mathbf{r}_t) = \delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sigma_{r_i}^2}$, then Equation (B11) becomes

$$
\begin{aligned}
\mathrm{IR} &= \frac{\mathrm{IC}\,\delta}{\sqrt{\left(E_{cs}(1/\sigma_{\varepsilon_i}^2)\right)^{-1}/N+\delta^2\sigma_{\mathrm{IC}}^2}} \\
&= \frac{\mathrm{IC}}{\sqrt{\left(\delta^2 E_{cs}(1/\sigma_{\varepsilon_i}^2)\right)^{-1}/N+\sigma_{\mathrm{IC}}^2}} \\
&= \frac{\mathrm{IC}}{\sqrt{1/(N\,\phi)+\sigma_{\mathrm{IC}}^2}}
\end{aligned}
\tag{B12}
$$

where ϕ is the same as defined in (A5). The formula above is exactly the same as Equation (42) which is what should be expected.

By applying the same assumptions for deriving Equation (B12) to Equation (B10), we get the multifactor fundamental law in terms of IC as follows:

$$
\begin{aligned}
\mathrm{IR} &= \sqrt{\mathbf{F}'(1/(\tau N)\,\mathbf{I}+\mathbf{\Sigma}_{\mathrm{F}})^{-1}\mathbf{F}} \\
&= \sqrt{\mathbf{IC}'\left(1/(\phi\,N)\,\mathbf{I}+\mathbf{\Sigma}_{\mathrm{IC}}\right)^{-1}\mathbf{IC}} \\
&\approx \sqrt{\mathbf{IC}'\,\mathbf{\Sigma}_{\mathrm{IC}}^{-1}\mathbf{IC}}
\end{aligned}
\tag{B13}
$$

where $\mathbf{IC}=\mathbf{F}/\delta$ is the cross-section correlation vector between factor exposures and residual security returns, and $\mathbf{\Sigma}_{\mathrm{IC}}=\mathbf{\Sigma}_{\mathrm{F}}/\delta^2$ is the factor IC covariance matrix. It should be emphasized that the results in Equations (B12) and (B13) are only valid when the cross-sectional residual return dispersion is a constant. When this assumption is violated, then the IR calculated from Equations (B10) and (B11) will usually be smaller than that from (B12) and (B13).

To avoid the problem of cross-sectional heteroskedasticity in cross-sectional regression, one can use the risk-adjusted residual security returns as the dependant variable, i.e.,

$$
\tilde{\mathbf{r}}_t=\mathbf{\Lambda}_t^{-1/2}\mathbf{r}_t=\mathbf{Z}_{t-1}\mathbf{IC}_t+\boldsymbol{\varepsilon}_t
$$

where $\mathbf{\Lambda}_t=diag(\sigma_{r_1}^2,\sigma_{r_2}^2,\cdots,\sigma_{r_N}^2)$, and $\sigma_{r_i}^2$ is the residual return variance for security *i*. By using the same algebra one can get

$$
\begin{aligned}
\mathrm{IR} &= \sqrt{\mathbf{IC}'\left(\sigma_{\varepsilon}^2/N\,\mathbf{I}+\mathbf{\Sigma}_{\mathrm{IC}}\right)^{-1}\mathbf{IC}} \\
&\approx \sqrt{\mathbf{IC}'\,\mathbf{\Sigma}_{\mathrm{IC}}^{-1}\mathbf{IC}}
\end{aligned}
\tag{B14}
$$

where $\sigma_{\varepsilon}^2=1-\sum_{k=1}^{K}(\sigma_{\mathrm{IC},k}^2+\mathrm{IC}_k^2)$. It should be emphasized again that the ICs in Equation (B14) are the cross-sectional correlation between risk-adjusted residual security returns and factor exposures, while the ICs in Equation (B13) are the correlation between the raw residual security returns and factor exposures, hence they will usually be different.

With the fundamental law in multifactor format, we can easily study the impact of missing one or more return or risk factors. In the analysis below, I will study the impact of missing factors based on factor ICs, the analysis based on factor returns is almost identical. I will not purposely distinguish risk factors from alpha factors. Statistically, the only difference should be that the expected IC (or factor return) for risk factor is zero while that for alpha factor is different from zero. I will separate the factors into two groups with $\mathbf{IC}_i$ and $\mathbf{\Sigma}_{ii}$ (i=1,2) as their factor IC and IC covariance respectively. I will also assume that the inter-group factor IC covariance to be $\mathbf{\Sigma}_{12}$. Under these assumptions, we can write Equation (B13) as follows [7]

$$
\begin{aligned}
\mathrm{IR} &\approx \sqrt{\mathbf{IC}'\mathbf{\Sigma}_{\mathrm{IC}}^{-1}\,\mathbf{IC}} \\
&= \sqrt{(\mathbf{IC}_1{}'\ \mathbf{IC}_2{}')\begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{12}{}' & \mathbf{\Sigma}_{22} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{IC}_1 \\ \mathbf{IC}_2 \end{pmatrix}} \\
&= \sqrt{\mathbf{IC}_1{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1 + (\mathbf{IC}_2 - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1)'\mathbf{E}^{-1}(\mathbf{IC}_2 - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1)} \\
&\geq \sqrt{\mathbf{IC}_1{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1}
\end{aligned} \qquad \text{(B15)}
$$

where $\mathbf{E} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$.

So IR^2 will be reduced by a amount of

$$
(\mathbf{IC}_2 - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1)'(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}(\mathbf{IC}_2 - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1) \geq 0 \qquad \text{(B16)}
$$

when the second group of k_2 factors are missing. The impacts come from both alpha model mis-specification (when $\mathbf{IC}_2 \neq \mathbf{0}$) and risk model mis-specification (when $\mathbf{IC}_2 = \mathbf{0}$ but $\mathbf{IC}_1{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1) > 0$).

Alternatively the IR can be expressed as

$$
\mathrm{IR} \approx \sqrt{\mathbf{IC}_2{}'\mathbf{\Sigma}_{22}^{-1}\mathbf{IC}_2 + (\mathbf{IC}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{IC}_2)'\mathbf{D}^{-1}(\mathbf{IC}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{IC}_2)} \qquad \text{(B17)}
$$

where $\mathbf{D} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}{}'$. When $\mathbf{IC}_2 = \mathbf{0}$, then the missing group is purely risk factors,

$$
\begin{aligned}
\mathrm{IR} &= \sqrt{\mathbf{IC}_1{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1 + \mathbf{IC}_1{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{12}{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1} \\
&= \sqrt{\mathbf{IC}_1{}'(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}{}')^{-1}\mathbf{IC}_1} \\
&\geq \sqrt{\mathbf{IC}_1{}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1},
\end{aligned} \qquad \text{(B18)}
$$

so the reduction in IR comes only from missed risk allocation. When $\mathbf{\Sigma}_{12} = \mathbf{0}$, i.e., the alpha group factor ICs and risk group factor ICs are not correlated, then missing risk factors will not impact the final portfolio performance.

Notes

[1] We used the fact that the benchmark residual return is zero in deriving Equation (7), i.e.,

$$
\sum_{i=1}^{N} w_{B,it} r_{it} = 0 .
$$

This is true because

$$R_{B,t} = \sum_{i=1}^{N} w_{B,it} r_{it}^{Total} = \sum_{i=1}^{N} w_{B,it} \beta_{it} R_{B,t} + \sum_{i=1}^{N} w_{B,it} r_{it} = R_{B,t} + \sum_{i=1}^{N} w_{B,it} r_{it} .$$

[2] We define the realized cross-sectional residual return dispersion at time *t* as

$$d(\mathbf{r}_t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_{it} - \bar{r}_t)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_{it}^2 - \bar{r}_t^2)} = \sqrt{E_{cs}(r_{it}^2 - \bar{r}_t^2)} ,$$

where $\bar{r}_t = \frac{1}{N}\sum_{i=1}^{N} r_{it}$ is the average cross-sectional residual return which we will assume to be zero in this article. The expected cross-sectional residual return dispersion is then

$$\delta = E(d(\mathbf{r}_t)) = E\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_{it}^2 - \bar{r}_t^2)}\right) = E\left(\sqrt{E_{cs}(r_{it}^2 - \bar{r}_t^2)}\right).$$

[3] We can decompose r_{it} as $r_{it} = \sigma_r e_{it}$ where $e_{it} \sim N(0,\ 1)$. So

$$\delta = E(d(\mathbf{r}_t)) = E\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} r_{it}^2}\right) = \sigma_r E\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} e_{it}^2}\right) = \sigma_r$$

as $N \to \infty$ by law of large numbers.

[4] When we assume the cross-sectional residual return dispersion is a constant, i.e.,

$$d(\mathbf{r}_t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} r_{it}^2} = d ,$$

then

$$\delta = E(d(\mathbf{r}_t)) = d .$$

On the other hand,

$$E(d^2(\mathbf{r}_t)) = \frac{1}{N}\sum_{i=1}^{N} E(r_{it}^2) = \frac{1}{N}\sum_{i=1}^{N} \sigma_{r_i}^2 = d^2 .$$

So we have

$$\delta = E(d(\mathbf{r}_t)) = d = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \sigma_{r_i}^2} .$$

[5] The assumption of normality in the information coefficient is approximate because IC is bounded by ± 1.

[6] The unconditional covariance of $\mathbf{r}_t$ is $E(\mathbf{r}_t \mathbf{r}_t{}') = (\mathrm{IC}^2 + \sigma_{\mathrm{IC}}^2)\delta^2 \mathbf{\Sigma}_z + \mathbf{\Sigma}$, where $\mathbf{\Sigma}_z$ is the covariance matrix of $\mathbf{z}_{t-1}$ with 1 in the diagonal.

[7] The inverse of a partitioned matrix is repeatedly used in the derivation, see Magnus and Neudecker (2002, p11) .

References

Clarke, Roger, Harindra de Silva, and Steven Thorley. 2002. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal*, vol. 58, no. 5 (September/October):48–66.

Grinold, Richard C. 1989. "The Fundamental Law of Active Management." *The Journal of Portfolio Management*, vol. 15, no. 3 (Spring): 30–38.

Grinold, Richard C. 1994. "Alpha is Volatility Times IC times Score." *The Journal of Portfolio Management*, vol. 20, no. 4 (Summer): 9–16.

Grinold, Richard C. 2007. "Dynamic Portfolio Analysis." *The Journal of Portfolio Management*, vol. 34, no. 1 (Fall): 12–26.

Grinold, Richard C., and Ronald N. Kahn. 2000. *Active Portfolio Management*. 2nd ed. New York: McGraw-Hill.

Ibragimov, R. and Müller, U. (2009). "t-statistic Based Correlation and Heterogeneity Robust Inference," forthcoming in the *Journal of Business & Economic Statistics*.

Kahn, Ronald, 1997. "Seven Quantitative Insights into Active Management Part 3: The Fundamental Law of Active Management," BARRA Newsletter, Winter.

Lee, Jyh-Huei and Dan Stefek. "Do Risk Factors Eat Alphas." *The Journal of Portfolio Management*, vol. 34, no. 4 (Summer 2008), pp. 12-25.

Magnus, Jan R. and Heinz Neudecker. 2002. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Revised Edition. New York: John Wiley & Sons.

Petersen, M. A. (2009), "Estimating standard errors in finance panel data sets: Comparing approaches," The Review of Financial Studies, 22, 435-480.

Qian, Edward, and Ronald Hua. "Active Risk and Information Ratio." *The Journal of Investment Management*, vol 2, no. 3 (2004), pp. 20-34.

Qian, E., Hua, R., and Sorensen, E.H. (2007). *Quantitative Equity Portfolio Management: Modern Techniques and Applications,* London: CRC Press.

Sorensen, Eric H., Ronald Hua, Edward Qian, and Robert Schoen. "Multiple Alpha Sources and Active Management." *The Journal of Portfolio Management*, vol. 30, no. 2 (Winter 2004), pp. 39-45.

Sorensen, Eric H., Ronald Hua, and Edward Qian. 2007. "Aspects of Constrained Long–Short Equity Portfolios." *The Journal of Portfolio Management*, vol. 33, no. 2 (Winter):12–22.

Ye, Jia. "How Variation in Signal Quality Affects Performance." *Financial Analysts Journal*, vol. 64, no. 4 (2008), 48-61.