Journal of Empirical Finance 43 (2017) 91–114

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

The fundamental law of active management: Redux

Zhuanxin Ding[a,*], R. Douglas Martin[b]

[a] Analytic Investors, 555 W 5th St, 50th Floor, Los Angeles, CA 90013, USA
[b] University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Keywords:
The fundamental law of active management
Information coefficient
Information ratio
Factor model
Time series
Cross section

ABSTRACT

We develop a fundamental law of active management based on cross-section factor models for residual returns where the latter have unconditional mean zero and the factor exposures have zero mean and unit variance. Under our model framework the factor returns are cross-sectional information coefficients IC_t that vary randomly over time with constant mean and variance. The fundamental law holds for portfolio managers who use conditional expectation of the residual returns and the associated conditional covariance matrix as inputs to active quadratic utility portfolio optimization. The fundamental law formula shows that the optimal portfolio's information ratio (IR) is positively related to the mean of IC_t and the number of assets N in the portfolio selection universe, inversely related to the volatility of IC_t, and is an absolute upper bound on IR as N tends to infinity. Support for our choice of factor model and our fundamental law is provided by an empirical study showing that significantly higher IR values are obtained using our choice of factor model as compared with IR values using an industry standard factor model.

1. Introduction

In his seminal work Grinold (1989) introduced the concept of information ratio and stated without proof a "fundamental law of active management" in the form of a simple formula for the information ratio. We recall that the information ratio (IR) of an actively managed portfolio is defined as the portfolio's mean return in excess of a benchmark portfolio return divided by the tracking error (standard deviation of the excess return). Since its initial introduction the IR has played a fundamental and ubiquitous role in gauging the performance of an active portfolio manager who manages a portfolio relative to a benchmark. Goodwin (1998) provides a useful reference on IR basics. A more well-developed version of the ideas in Grinold (1989) play a central role in the Grinold and Kahn (2000) book *Active Portfolio Management* (G & K) that became the "bible" on the topic for many active portfolio managers for many years.

The beginning of Chapter 6 in G & K introduces the fundamental law of active management (FLAM), and states that the information ratio is approximated by the formula $IR = IC \cdot \sqrt{BR}$ where IC is the "information coefficient", and BR is "breadth" defined to be the number of independent forecasts of exceptional returns made each year. Here "exceptional" means the returns obtained after removing beta-adjusted benchmark returns. The term information coefficient is simply a colorful way of referring to a correlation coefficient, namely the correlation between asset returns and the predictor signal used by the manager to forecast the returns. Here asset returns refers to "residual" returns as defined subsequently, and also referred to as "exceptional" returns in G & K. Once one realizes that an IC is a correlation coefficient between returns and a predictor of returns it is not surprising that Grinold identified the IC as a measure of skill. More skilled active managers find better signals that have larger correlation with the returns to be predicted and smaller prediction errors, leading to more profitable portfolios.

* Corresponding author.
E-mail addresses: zding@aninvestor.com (Z. Ding), doug@amath.washington.edu (R.D. Martin).

http://dx.doi.org/10.1016/j.jempfin.2017.05.005
Received 25 October 2011; Accepted 24 May 2017
Available online 30 May 2017

The Chapter 6 Appendix of G & K combines the quadratic utility method of finding optimal active portfolio weights with best linear predictor theory to derive an optimal information ratio. But then the fundamental law as stated was taken without justification to be for the very special case that “all signals have equal value”, i.e., there is a common correlation for all predictors. Taken as a totality there are several problems with the above results. First of all, being based on generic best linear predictor theory the law is not based on any econometric data-generating model as a foundation. Second, although the derivation makes use of the conditional mean predictors it does not make appropriate use of the conditional forecast error covariance structure. Finally, the law as stated above does not provide any guidance on how one should calculate the number of independent forecasts. As a consequence, practitioners have sometimes taken breadth to be the number of assets in the portfolio selection universe, and this leads to IR values that are far too optimistic relative to what active managers can achieve in practice.

Subsequent to G & K a number of authors have tried to improve the formulation of the fundamental law. Buckle (2004) again used a best linear predictor framework along with some of the G & K assumptions to derive a more general form of the law involving predictor correlations that reduces to the G & K simple form when the predictors are independent. Grinold (2007) also made some proposals on measuring breadth. Clarke et al. (2002) extended the law to take into account portfolio optimization constraints, but under an unrealistic assumption that residual returns are uncorrelated. Clarke et al. (2006) extended the law to the case of correlated assets with constraints. The results show not unexpectedly that the impact of constraints is to reduce the information ratio. A completely new and important fundamental law ingredient was introduced by Qian and Hua (2004). They derived a fundamental law that reflects the randomness of the IC over time by virtue of the appearance of the IC mean and standard deviation in their fundamental law, and referred to this randomness as “strategy risk”. Subsequently Ye (2008) derived a version of the law that took into account randomness of the IC.

In spite of the considerable literature on the fundamental law and its extensions, there exist basic limitations of the existing derivations of the law. The first and most basic is that none of the derivations use an econometric data-generating model as a foundation. Second, most of the approaches lack a well-defined model for IC randomness. Last but not least, the lack of adequate treatment of conditional versus unconditional variance and covariance in the derivations results in overly optimistic information ratios for the resulting optimized portfolios. Our work removes these limitations by: (1) using an econometric factor model motivated abstractly by the arbitrage pricing theory and concretely by the well–established industry use of cross-section fundamental factor models in asset, but with a significant difference from the latter, (2) endowing the econometric model with a key randomness component that results in IC randomness, and (3) proper handling of conditional and unconditional predictions and their covariance. As a result we obtain a new general form of fundamental law of active management based on a solid foundation that contains prior versions as special cases.

In Section 2.1 we discuss residual returns from a benchmark single index model and a CAPM motivation for a key property of residual returns. Section 2.2 defines residual returns conditional mean forecast and related conditional forecast error covariance matrix and Section 2.3 describes the quadratic utility active portfolio optimization method used in developing the fundamental law. Section 2.4 defines and discusses the multi-factor model assumptions we use, and proves two results used in subsequent development of the fundamental law. Section 3 focuses on single-factor models, defines factor returns and identifies them as time series of random information coefficients and derives the single factor fundamental law. A key result is that there is an absolute upper bound to the information ratio no matter how many assets the portfolio includes, and it is shown that the G & K formula is an unrealistic limiting case of our single factor fundamental law. Section 4 extends the single factor model results to the more general case of a multi-factor model, and briefly discusses the negative impact of using a factor model that is misspecified by virtue of leaving out some needed factors. Section 5 provides discussion and results of a detailed empirical study in support of our new fundamental law for the case of a single factor model. Section 6 provides a summary and concluding comments on future research needed.

2. Assumptions and discussion

The four subsections to follow discuss four important sets of framework assumptions we use in deriving the fundamental law of active management. The first has to do with the use of residual returns relative to a benchmark as the input to portfolio optimization. The second is concerned with the use of conditional mean forecasts and the associated conditional covariance matrix. The third concerns our use of portfolio optimization based on quadratic utility optimization. The fourth subsection provides our assumptions concerning the multi-factor model used along with related discussion of the assumptions.

2.1. Benchmark residual returns

The following single index model relative to the benchmark is a standard tool in active portfolios management:

$$r_{it}^{Total} = \beta_i r_{B,t} + r_{it} \ , \ i = 1, \cdots, N, \ t = 1, \cdots, T, \tag{1}$$

where

r_{it}^{Total} = total return of security i at time t in excess of the risk-free rate
β_i = beta of security i with respect to the benchmark
$r_{B,t}$ = benchmark return at time t in excess of the risk-free rate
r_{it} = residual return of security i at time t

and the residual returns are uncorrelated with the benchmark returns. The benchmark is assumed to be a broad market proxy index, e.g., the Russell 3000 or Wilshire 5000 for the U.S. market and the MSCI All-World Index for the global market.

Use of the above residual returns model without a time index goes back to the important but rather overlooked work of Treynor and Black (1973), who introduced the concept of "active management". Their normative flavored paper "How to Use Security Analysis to Improve Portfolio Selection" used the above model with two assumptions. The first is that the vector residual returns, which they called "independent return(s) on the *i*-th security" has a diagonal covariance matrix. The second and quite innovative assumption was that mean values $\mu_i = E(r_i)$ of the residual returns, which Treynor and Black called the "appraisal" values, can occasionally be non-zero and can be exploited for a profit by the portfolio manager who has "special information".

Subsequently G & K used the residual returns model in their mathematical development of the fundamental law of active management in the Technical Appendix to Chapter 6 of Grinold and Kahn (2000). As in Treynor and Black (1973), the G & K residual returns model lacks time subscripts, and their additional assumptions include: (1) the vector residual returns have an arbitrary covariance matrix, (2) the mean of the residual returns is zero, (3) there exists a vector of "signals" that have positive predictive power with respect to the residual returns by virtue of non-zero correlations between the residual returns and the signals. As a consequence of the third assumption the conditional mean of the residual returns, conditioned on the signal values is generally non-zero. These non-zero conditional means in effect replace the vagueness of the Treynor and Black assumption that the residual means are occasionally non-zero.

In some of the later literature on active management such as in Clarke et al. (2002, 2006) it is assumed that the residual returns have a non-zero mean value "alpha" that can be exploited to obtain a profit over and above that obtainable from passive investing in the benchmark. However the existence of such a model implies that all investors with the same information would be able to make such a profit and one would expect such profit to quickly disappear as in Treynor and Black (1973) as the market adjusts to this possibility. Thus the assumption of a non-zero mean of residual returns is not a very reasonable one.

The Capital Asset Pricing Model (CAPM) provides another rationale in favor of zero mean residual returns. Although considerable empirical research evidence has accumulated over the years indicating that the CAPM is not exactly correct, the model has none-the-less served as a useful approximation in many quantitative applications in finance. So for the cases of large market index benchmarks which serve as market proxies, we assume the CAPM is a useful approximation for developing a fundamental law of active management.

Thus with the above observations in mind we make the assumption throughout that

$$E(r_{it}) = 0,\ i = 1,\ \cdots,\ N. \tag{2}$$

At the portfolio level the beta-adjusted active return is

$$r_{A,t} = r_{P,t} - \beta_{P,t} r_{B,t} \tag{3}$$

where $\beta_{P,t} = \sum_{i=1}^{N} w_{P,it}\beta_i$ is the portfolio beta relative to the benchmark. Throughout the paper we will for the sake of brevity refer to $r_{A,t}$ as simply the *active return*.[1] We define *active weights* in the usual manner as benchmark relative portfolio weights:

$$\Delta w_{it} = w_{P,it} - w_{B,it}. \tag{4}$$

It then follows that[2]

$$r_{A,t} = \sum_{i=1}^{N} \Delta w_{it} r_{it} = \Delta \mathbf{w}_t' \mathbf{r}_t\ . \tag{5}$$

We note that under the standard assumption that the portfolio and the benchmark are fully-invested, the sum of the benchmark relative weights is zero. In the parlance of practitioners the active weights are dollar neutral.

2.2. The conditional mean forecast and error covariance matrix

We define the *ex ante* alpha of security i in period t as the *expected* residual return conditional on random information I_{t-1} available at time $t-1$:

$$\boldsymbol{\alpha}_t = E(\mathbf{r}_t \mid \mathrm{I}_{t-1}) \tag{6}$$

where $\boldsymbol{\alpha}_t$ and $\mathbf{r}_t$ are $N \times 1$ vectors with α_{it} and r_{it} as their elements respectively. Note that since the conditioning information is random in nature $\boldsymbol{\alpha}_t$ is a random variable. Furthermore the assumption of zero unconditional expected residual return implies that $\boldsymbol{\alpha}_t$ has an unconditional expected value of zero:

$$E(\boldsymbol{\alpha}_t) = E(E(\mathbf{r}_t \mid \mathrm{I}_{t-1})) = 0. \tag{7}$$

We assume that while all investors have access to all publicly available information not all investors exploit this information equally well. Some investors will use publicly available information to make better forecasts and hence better bets on individual securities and thereby construct portfolios with higher risk-adjusted returns than other investors. This is very much in the spirit of the early work of Treynor and Black (1973), who assumed that better informed investors could exploit temporary non-zero values of

[1] Our definition is in contrast to the common literature definition of active return as $r_{A,t} = r_{P,t} - r_{B,t}$. The latter is a special case of (3) when $\beta_{P,t} = 1$.

[2] This fact was pointed out by Clarke et al. (2002) without stating the following simple algebra needed:

$$\sum_{i=1}^{N} \Delta w_{it} r_{it} = \sum_{i=1}^{N} (w_{P,it} - w_{B,it})(r_{it}^{Total} - \beta_{it} r_{B,\,t}) = \sum_{i=1}^{N} w_{P,it} r_{it}^{Total} - \sum_{i=1}^{N} w_{B,it} r_{it}^{Total} - \sum_{i=1}^{N} w_{P,it}\beta_{it} r_{B,\,t} + \sum_{i=1}^{N} w_{B,it}\beta_{it} r_{B,\,t}$$

and since the benchmark beta is one the right-hand side is $r_{P,t} - r_{B,t} - \beta_P r_{B,t} + \beta_B r_{B,t} = r_{P,t} - \beta_P r_{B,t} = r_{A,t}$.

expected residual returns. While their work did not provide a theory or formal implementation model it provides considerable motivation for our approach.

Under our residual return model, the following covariance matrix of errors in our forecast $\boldsymbol{\alpha}_t = E(\mathbf{r}_t \mid \mathrm{I}_{t-1})$ is the relevant measure of risk:

$$\boldsymbol{\Omega}_t = E[(\mathbf{r}_t - \boldsymbol{\alpha}_t)(\mathbf{r}_t - \boldsymbol{\alpha}_t)' \mid \mathrm{I}_{t-1}]. \tag{8}$$

The average value of the above conditional covariance matrix associated with our alpha estimator is bounded above by the total risk around the unconditional alpha expectation value of zero, and the difference between the two reflects the quality of the forecast.[3]

It follows from the above that the conditional mean and conditional variance of the active return $r_{A,t}$ are given by:

$$E(r_{A,t} \mid \mathrm{I}_{t-1}) = E(\Delta \boldsymbol{w}_t' \mathbf{r}_t \mid \mathrm{I}_{t-1}) \quad = \quad \Delta \boldsymbol{w}_t' \boldsymbol{\alpha}_t \tag{9}$$

and

$$\mathrm{var}(r_{A,t} \mid \mathrm{I}_{t-1}) = \mathrm{var}(\Delta \boldsymbol{w}_t' \mathbf{r}_t \mid \mathrm{I}_{t-1}) = \Delta \boldsymbol{w}_t' \boldsymbol{\Omega}_t \Delta \mathbf{w}_t. \tag{10}$$

Our use of conditional forecast error variance is a 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, 2006), Qian and Hua (2004), Qian et al. (2007), and Ye (2008).

2.3. Quadratic utility optimization

The original G & K derivation as well as all subsequent treatments are based on the use of optimal active portfolio weights obtained using the expected quadratic utility variant of the Markowitz (1952) mean-variance portfolio optimization theory. Quadratic utility decreases for sufficiently large values of wealth and has increasing absolute and relative risk aversion, all of which are inconsistent with typical investor behavior. However justification for using the expected quadratic utility variant of mean-variance optimization was provided by Levy and Markowitz (1979), who showed that the mean-variance approach provides a good approximation to expected utility maximization for increasing concave utility functions. Furthermore quadratic utility is the approach used almost universally by practitioners, the more so if they use a commercial portfolio optimization and risk management software product. Thus it is worthwhile to develop a fundamental law of active management in the expected quadratic utility framework.

We apply the expected quadratic utility (henceforth "quadratic utility" for short) to the conditional mean forecast of residual returns and the associated conditional covariance matrix for the following reasons. First, the conditional mean is used because one wants to optimize the end of period portfolio performance based on beginning of period information. We use the residual returns because they are uncorrelated with the benchmark returns and as such do not contain benchmark returns information. It follows that the resulting portfolio is benchmark neutral, i.e., has a zero beta on the benchmark. Furthermore, once the optimal active portfolio is constructed based on residual returns, the portfolio manager may invest in a weighted combination of the benchmark and the residual returns optimal portfolio to achieve a desired degree of exposure to the benchmark. We note that constructing an optimal portfolio based on residual returns is the approach also taken by G & K and by Clarke et al. (2002, 2006) among others.

Using $\boldsymbol{\alpha}_t$ and $\boldsymbol{\Omega}_t$ as defined above and recalling that active weights are dollar neutral, the quadratic utility active portfolio optimization problem is:

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

where

$\alpha_{A,\,t}$ = portfolio active return conditional mean
$\sigma_{A,\,t}^2$ = portfolio active return conditional variance
λ = a risk-aversion parameter
$\mathbf{1}$ = an N-vector of 1's.

The solution for this optimization problem is easily obtained using the method of Lagrange multipliers to handle the constraint, and the result is:

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

where $\kappa = (\boldsymbol{\alpha}_t' \boldsymbol{\Omega}_t^{-1} \mathbf{1})/(\mathbf{1}' \boldsymbol{\Omega}_t^{-1} \mathbf{1})$.

In the active portfolio management literature the quantity

$$\sigma_{A,t} = \sqrt{\Delta \mathbf{w}_t' \boldsymbol{\Omega}_t \Delta \mathbf{w}_t} \tag{12}$$

[3] This follows from application of $\mathbf{cov}(\mathbf{y}) = E[\mathbf{cov}(\mathbf{y} \mid \mathrm{X})] + \mathbf{cov}[E(\mathbf{y} \mid \mathrm{X})]$ with $\mathbf{y} = \mathbf{r}_t$, $\mathrm{X} = \mathrm{I}_{t-1}$, $E(\mathbf{y} \mid \mathrm{X}) = E(\mathbf{r}_t \mid \mathrm{I}_{t-1}) = \boldsymbol{\alpha}_t$, $\mathbf{cov}(\mathbf{y} \mid \mathrm{X}) = \boldsymbol{\Omega}_t$ and noting that $E(\boldsymbol{\alpha}_t) = E[E(\mathbf{r}_t \mid \mathrm{I}_{t-1})] = \mathbf{0}$, thereby obtaining $\mathbf{cov}(\mathbf{r}_t) = E(\boldsymbol{\Omega}_t) + E(\boldsymbol{\alpha}_t \boldsymbol{\alpha}'_t)$.

is referred to as the portfolio's tracking error (TE). In practice active portfolio managers set a target TE, i.e., a target value for $\sigma_{A,t}$ as part of their mandate. Substitution of (11) into the expression for $\sigma_{A,t}$ in (12) results in the following expression for the risk aversion parameter:

$$\lambda = \sigma_{A,t}^{-1}\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t - \kappa\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t}\,. \tag{13}$$

Using the above in (11) gives the optimal portfolio active weight

$$\Delta\mathbf{w}_t = \sigma_{A,t}\frac{\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}{\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}}. \tag{14}$$

The corresponding portfolio active return conditional mean is

$$\alpha_{A,t} \triangleq \Delta\mathbf{w}_t'\boldsymbol{\alpha}_t = \sigma_{A,t}\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}\,. \tag{15}$$

Thus we define the conditional information ratio as

$$\mathrm{IR}_t \triangleq \frac{\alpha_{A,t}}{\sigma_{A,t}} = \sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})} \tag{16}$$

and the unconditional information ratio of the portfolio is

$$\mathrm{IR} \triangleq E(\mathrm{IR}_t) = E(\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}\,). \tag{17}$$

This is a very general result that holds under quadratic utility optimization using a conditional mean predictor $\boldsymbol{\alpha}_t$ of residual return and corresponding conditional covariance matrix $\boldsymbol{\Omega}_t$. It will be seen in what follows that an interesting part of the fundamental law we derive is associated with the specific structure of $\boldsymbol{\Omega}_t$ and $\boldsymbol{\alpha}_t$ under the cross-section factor models we use, and this structure provides analysts with more insight concerning the ex-ante expected information ratio.

2.4. The multi-factor model

Multi-factor models have played a central role in both empirical asset pricing research and especially in the practice of portfolio optimization and risk management.[4] The three main types of factor models are fundamental factor models, statistical factor models and time series factor models. Among these, fundamental factor models are by far the most frequently used in practice, and this is no small way due to basic studies involving their use by Fama and MacBeth (1973), Rosenberg (1974) and Fama and French (1992), among others.

Fundamental factor models (also called characteristics based models) are cross-section regression models where the explanatory variables are "exposures" to risk factors that include continuous variables such as book-to-market, earnings-to-price, size, momentum, etc., and categorical variables reflecting sector, industry and country exposures. In such multi-factor models a cross-section regression model fit is computed at each time period in an overall time interval of interest. The resulting coefficient estimates form a time series of "factor returns" over the period of interest, and the model is used in practice for both portfolio risk analysis and asset returns forecasting. For discussion on the importance of using the same factor model for risk analysis and forecasting see Lee and Stefek (2008).

The multi-factor model we use in developing our fundamental law of active management is an important variant of a standard fundamental multi-factor model in which the residual returns are standardized. Specifically we assume that an $N \times 1$ vector $\tilde{\mathbf{r}}_t$ of standardized residual returns is obtained from residual returns $\mathbf{r}_t$ by the transformation

$$\tilde{\mathbf{r}}_t = \boldsymbol{\Lambda}_t^{-1/2}\mathbf{r}_t \tag{18}$$

where $\boldsymbol{\Lambda}_t = diag(\sigma_{r_{1t}}^2, \sigma_{r_{2t}}^2, \cdots, \sigma_{r_{Nt}}^2)$ with $\sigma_{r_{it}}^2$ a conditional variance of r_{it} for $i = 1, 2, \cdots, N$ based on past information described in assumption (A1) below. The standardized residual returns have variance one and according to (7) that they have unconditional mean zero.

Let $\mathbf{z}_{i,t-1} = (z_{i1,t-1},\ z_{i2,t-1},\ \cdots,\ z_{iK,t-1})$ be a $1 \times K$ vector of random factor exposures for asset i whose values are known at time $t-1$. The exposures are typically chosen to be properly standardized firm specific quantities such as book-to-market, earnings-to-price, firm size (log of market capitalization in millions of dollars), momentum, etc. Let $\mathbf{f}_t = (f_{1t}, f_{2t}, \cdots, f_{Kt})'$ be a $K \times 1$ random vector of factor returns at time t whose values are not known at time $t-1$. Then our multi-factor model is

$$\tilde{r}_{it} = \mathbf{z}_{i,t-1}\mathbf{f}_t + \varepsilon_{it},\ i = 1, \cdots, N \tag{19}$$

equivalently

$$\tilde{\mathbf{r}}_t = \mathbf{Z}_{t-1}\mathbf{f}_t + \boldsymbol{\varepsilon}_t \tag{20}$$

where $\mathbf{Z}_{t-1}$ is an $N \times K$ matrix of factor exposures $z_{ik,t-1}$, $i = 1, \cdots, N$, $k = 1, \cdots, K$ and $\boldsymbol{\varepsilon}_t$ is the $N \times 1$ error vector.

We make the following assumptions concerning the multi-factor model.

[4] See for example the literature on multi-factor model portfolio optimization and risk analysis systems provided by commercial providers such as MSCI-Barra (https://www.msci.com/), Axioma (http://axioma.com), Sungard APT (http://www.sungard.com/apt) and Northfield (http://www.northinfo.com).

(A1). The $\sigma_{r_{it}}$ depend on the lagged returns r_{iu}, $u \le t-1$ and these standard deviations have a lower bound $b_{lo} > 0$ and a finite upper bound b_{up} uniformly across assets.

(A2). The exposures $z_{ik,t-1}$, $i = 1, \cdots, N$, $k = 1, \cdots, K$ have unconditional mean zero, variance one and finite fourth moment, and for each k the exposures $z_{ik,t-1}$ and $z_{jk,t-1}$ are independent for $i \ne j$.

(A3). The exposures $z_{ik,t-1}$ and $z_{im,t-1}$ are uncorrelated for $k \ne m$, $k, m = 1, \cdots, K$, $i = 1, \cdots, N$.

(A4). The random factor return $\mathbf{f}_t$ is a stationary random process with known mean $E(\mathbf{f}_t) = \mathbf{f}$ and known covariance matrix $\boldsymbol{\Sigma}_f$.

(A5). The error terms satisfy $E(\varepsilon_{it} \mid \mathbf{f}_t, \mathrm{I}_{t-1}) = 0$ and $E(\varepsilon_{it}^2 \mid \mathbf{f}_t, \mathrm{I}_{t-1}) = \sigma_{\varepsilon_i}^2$ for all i, and $E(\varepsilon_{it}\varepsilon_{jt} \mid \mathbf{f}_t, \mathrm{I}_{t-1}) = 0$ when $i \ne j$.

2.4.1. Discussion of assumptions

Our use of $t-1$ as the time subscript on exposures and time subscript t on asset returns is not common in the academic literature. For example the arbitrage pricing literature typically uses no time subscript at all, e.g., as in Ross (1976) and Huberman and Wang (2008). Some academic literature uses a common time subscript t on both asset returns and exposures, e.g., as in Fama and MacBeth (1973) and Fama and French (1992), and other academic literature uses a time subscript t on both asset returns and factor returns but no time subscript on exposures, for example as in Menchero and Mitra (2008) and Connor and Korajczyk (2010).

However, among practitioners the convention we use with respect to time subscripts is quite common; see for example Bloch et al. (1993), Guerard et al. (2013), and Zangari (2003). The reason for this is that in practice one constructs a portfolio optimized with respect to performance at time t given information up to and including the previous time $t-1$. This is consistent with our Section 2.2 introduction of the conditional mean alpha and associated conditional covariance matrix. We also note that our approach formalizes the G & K use of the vague term "signals" by identifying them with past exposures.

Now we turn to discussing the assumptions (A1), (A2), (A3), (A4) and (A5) and label our comments accordingly as (D1), (D2), (D3), (D4) and (D5).

(D1). In practice $\sigma_{r_{it}}$ would typically and easily be estimated with a GARCH model based on past r_{it} to account for the time-varying volatility of the residual returns. The lower bound and upper bound reflect that in practice a portfolio contains no assets with arbitrarily small or arbitrarily large risk.

(D2). The zero mean and unit variance exposures of our model are standardizations of raw exposures $x_{ik,t-1}$, $i = 1, \cdots, N$, $k = 1, \cdots, K$. If for each exposure type (i.e., for each k) the raw exposures $x_{ik,t-1}$ for all assets (i.e., for all i) all had the same mean and variance, then a simple cross-sectional data-based standardization used by most practitioners as described in Section 5.1 would suffice to obtain good approximations to $z_{ik,t-1}$. However, raw exposures $x_{ik,t-1}$ arising in practice have different means and variances across assets for each fixed exposure type, and in this case the simple cross-section data-based standardization does not work, as is shown in Section 5.1. Thus our assumption concerning the raw exposures is that for each fixed k the means $\mu_{ik,t-1} = E(x_{ik,t-1})$, $i = 1, 2, \cdots, N$ and variances $\sigma_{ik,t-1}^2 = \mathrm{var}(x_{ik,t-1})$, $i = 1, 2, \cdots, N$ are different and unknown. In this case the exact standardization of the raw exposures for each k is $z_{ik,t-1} = (x_{ik,t-1} - \mu_{ik,t-1})/\sigma_{ik,t-1}$, $i = 1, \cdots, N$. In order to implement this standardization one needs to obtain estimates $\hat{\mu}_{ik,t-1}$ and $\hat{\sigma}_{ik,t-1}^2$ on an asset-by-asset basis (varying i) as well as on an exposure by exposure basis (varying k). In practice it turns out that the cross-section variation in exposure variances $\sigma_{ik,t-1}^2$ are typically substantially larger than the cross-section variation in expected values $\mu_{ik,t-1}$, and it is correspondingly more important to get a good estimate of the $\sigma_{ik,t-1}^2$ on an asset-by-asset basis than to get a good estimate of the $\mu_{ik,t-1}$ on an asset-by-asset basis. It is shown in Section 5 that good estimators of the $\sigma_{ik,t-1}^2$ on an asset-by-asset basis can be obtained using an EWMA estimator, and that a simple cross-section sample mean estimator suffices as a surrogate for an asset-by-asset estimator of the $\mu_{ik,t-1}$.

(D3). In case the original vectors $\mathbf{z}_{i,t-1} = (z_{i1,t-1}, z_{i2,t-1}, \cdots, z_{iK,t-1})$ have a non-diagonal positive definite covariance matrix, a non-singular transformation can be used to achieve a diagonal structure and the inverse of that transformation used to transform the factor returns $\mathbf{f}_t$ without changing the model, and the original exposures and factor returns can be recovered by reversing the transformations. Note (D2) and (D3) imply that $\mathbf{z}_{i,t-1}$ has covariance matrix $\boldsymbol{\Sigma}_{\mathbf{z}} = \mathbf{I}$.

(D4). We comment on this further in the discussion on factor returns and information coefficients.

(D5). These are standard assumptions for linear regression models extended to conditioning on past information.

2.4.2. Preliminary results

For future reference we have the following results, the first of which extends a standard constant coefficient linear regression model assumption to our case of random factor returns, and the second of which is consequence of the residual returns standardization (18).

Lemma 1. The error term satisfies $E(\varepsilon_{it} \mid \mathbf{Z}_{t-1}) = E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2 \mid \mathbf{Z}_{t-1}) = E(\varepsilon_{it}^2) = \sigma_{\varepsilon_i}^2$ for $i = 1, 2, \cdots, N$, and ε_{it} and $f_{kt} z_{jk,t-1}$ are uncorrelated for $i, j = 1, \cdots N$, $k = 1, \cdots, K$.

Proof. The conditional mean and conditional variance results for ε_{it} follow from (A5) by noting that $\mathbf{Z}_{t-1}$ is contained in I_{t-1} and taking expectation with respect to $\mathbf{f}_t$ and the part of I_{t-1} (if any) that does not contain $\mathbf{Z}_{t-1}$. Since $E(\varepsilon_{it} f_{kt} z_{j,t-1} \mid f_{kt} z_{j,t-1}) = f_{kt} z_{j,t-1} E(\varepsilon_{it} \mid f_{kt} z_{j,t-1}) = 0$ by similar reasoning, it follows that $E(\varepsilon_{it} f_{kt} z_{j,t-1}) = E(E(\varepsilon_{it} f_{kt} z_{j,t-1} \mid f_{kt} z_{j,t-1})) = 0$ and so ε_{it} and $f_{kt} z_{jk,t-1}$ are uncorrelated.

Lemma 2. Conditioned on the information set I_{t-1} the errors have constant variances and the error covariance matrix is $\mathbf{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \cdot \mathbf{I}$ where $\sigma_\varepsilon^2 = 1 - \sum_{k=1}^{K} (f_k^2 + \sigma_{f,k}^2)$.

Proof. Since the two terms on the right-hand side of (19) are uncorrelated and the left-hand side has variance one, we have

$$1 = \mathrm{var}(\mathbf{z}'_{i,t-1}\mathbf{f}_t) + \sigma_{\varepsilon_i}^2,\ i = 1, \cdots, N.$$

Applying the law of total variance to the first term on the right-hand side we have

$$\begin{aligned} \mathrm{var}(\mathbf{z}'_{i,t-1}\mathbf{f}_t) &= \mathrm{var}[E(\mathbf{z}'_{i,t-1}\mathbf{f}_t \mid \mathbf{z}'_{i,t-1})] + E[\mathrm{var}(\mathbf{z}'_{i,t-1}\mathbf{f}_t \mid \mathbf{z}_{i,t-1})] \\ &= \mathrm{var}(\mathbf{z}'_{i,t-1}\mathbf{f}) + E(\mathbf{z}'_{i,t-1}\mathbf{\Sigma}_f \mathbf{z}_{i,t-1}) \\ &= \mathrm{var}\left(\textstyle\sum_{k=1}^{K} z_{ik,t-1} f_k\right) + E\left(\textstyle\sum_{k,l=1}^{K} z_{ik,t-1} z_{il,t-1} \mathbf{\Sigma}_{f,kl}\right) \\ &= \textstyle\sum_{k=1}^{K} f_k^2 + \sum_{k=1}^{K} \mathbf{\Sigma}_{f,kk} \end{aligned} \tag{21}$$

where $\mathbf{\Sigma}_{f,kl}$ is element of row k and column l of $\mathbf{\Sigma}_f$, and we have used the fact that $z_{ik,t-1}$ and $z_{il,t-1}$ are uncorrelated for $k \neq l$. Thus we have the constant error variances

$$\sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2 = 1 - \sum_{k=1}^{K} (f_k^2 + \sigma_{f,k}^2),\ i = 1, \cdots, N. \tag{22}$$

It follows from (A5) that $E(\varepsilon_{it}\varepsilon_{jt}) = 0$ when $i \neq j$ and so the error covariance matrix is:

$$\mathbf{\Sigma}_\varepsilon = \sigma_\varepsilon^2\ \mathbf{I} = \left(1 - \sum_{k=1}^{K} (f_k^2 + \sigma_{f,k}^2)\right) \mathbf{I}. \tag{23}$$

3. Fundamental law for single factor models

For a single factor model Eq. (19) becomes

$$\tilde{r}_{it} \triangleq r_{it}/\sigma_{r_{it}} = f_t z_{i,t-1} + \varepsilon_{it},\ i = 1, \cdots, N,\ t = 1, \cdots, T \tag{24}$$

where $\sigma_{r_{it}}$ is the conditional volatility for security i based on lagged returns r_{iu}, $u \leq t-1$. As was mentioned earlier our scaling of the residual returns such that $\mathrm{var}(\tilde{r}_{it}) = 1$ is not standard and it is an important feature of our model that leads to the following two important consequences.[5]

(P1). The model has the economic intuition that *one unit of risk-adjusted exposure is expected to get rewarded (paid off) with the same amount of risk-adjusted residual return across securities.*

(P2). The standardization prevents high volatility assets from having an undue influence on a cross-section regression estimator of the factor return f_t.

3.1. Factor returns are information coefficients

Motivated by the common practice of estimating the factor returns by a cross-section regression, we define the random factor returns in our model (24) as

$$f_t \triangleq \frac{\mathrm{cov}(\tilde{r}_{it}, z_{i,t-1})}{\mathrm{var}(z_{i,t-1})} \tag{25}$$

which does not depend on the value of the assets index i.

Since both the risk-adjusted return and the factor exposure have unit variance our factor return is the correlation between the exposures and the standardized residuals, and as such is commonly called the information coefficient

$$\mathrm{IC}_t \triangleq f_t = corr(\tilde{r}_{it}, z_{i,t-1}). \tag{26}$$

The random character of f_t and hence IC_t is due to the random variation in the covariance between $\tilde{r}_{it}$ and $z_{i,t-1}$, which at each time has a different realized value. This random variation is intuitively appealing and corresponds to observed empirical time series behavior of cross-section regression estimates of f_t, examples of which are provided in Section 5. We make only the weak assumption that $\tilde{r}_{it}$, $z_{i,t-1}$ and their covariance are stationary random processes.

Since we have identified that under our model the random factor return f_t is equal to the information coefficient IC_t, we use IC_t as the notation for factor return instead of f_t. In keeping with the convention of using f as the notation for the expected value of f_t we use IC as the notation for the expected value of IC_t.

[5] It is common practice in fundamental factor model fitting to use a GARCH conditional volatility model for the error term ε_{it} to account for the time varying specific volatility risk. But this is not typically done for the security returns or single factor benchmark model residual returns r_{it}.

3.2. Single factor model fundamental law

In order to derive the fundamental law for the single factor model we need a vector version of the alpha conditional forecast model and the matrix version of the conditional covariance matrix. Note that under our factor model assumptions of Section 2.4 it follows from (24) that the conditional expected value predictor of r_{it} conditioned on known $z_{i,t-1}$ is

$$\alpha_{it} = E(r_{it}\,|\,\mathrm{I}_{t-1}) = E(\sigma_{r_{it}}\mathrm{IC}_t z_{i,t-1} + \sigma_{r_{it}}\varepsilon_{it}\,|\,\mathrm{I}_{t-1}) = \sigma_{r_{it}}\cdot\mathrm{IC}\cdot z_{i,t-1}. \tag{27}$$

Thus with $\boldsymbol{\Lambda}_t = diag(\sigma^2_{r_{it}})$ the vector of conditional mean forecasts is:

$$\boldsymbol{\alpha}_t = \boldsymbol{\Lambda}_t^{1/2}\cdot\mathrm{IC}\cdot\mathbf{z}_{t-1}. \tag{28}$$

The above formula is the vector version of the formula $Alpha = Volatility \times \mathrm{IC} \times Score$ in Grinold (1994).

Furthermore the conditional forecast error covariance for r_{it} and r_{jt} is given by

$$\begin{aligned} E((r_{it}-\alpha_{it})(r_{jt}-\alpha_{jt})\,|\,\mathrm{I}_{t-1}) &= E\left(\sigma_{r_{it}}((\mathrm{IC}_t-\mathrm{IC})z_{i,t-1}+\varepsilon_{it})\sigma_{r_{jt}}((\mathrm{IC}_t-\mathrm{IC})z_{j,t-1}+\varepsilon_{jt})\,|\,\mathrm{I}_{t-1}\right) \\ &= \sigma_{r_{it}}(E(\mathrm{IC}_t-\mathrm{IC})^2 z_{i,t-1}z_{j,t-1} + E(\varepsilon_{it}\varepsilon_{jt}\,|\,\mathrm{I}_{t-1}))\sigma_{r_{jt}}. \end{aligned}$$

Thus with $\sigma^2_{\mathrm{IC}} = E(\mathrm{IC}_t - \mathrm{IC})^2$ and $\boldsymbol{\Sigma}_\varepsilon = \sigma^2_\varepsilon\,\mathbf{I}$ the conditional forecast error covariance matrix is:

$$\boldsymbol{\Omega}_t = E((\mathbf{r}_t-\boldsymbol{\alpha}_t)(\mathbf{r}_t-\boldsymbol{\alpha}_t)'\,|\,\mathrm{I}_{t-1}) = \boldsymbol{\Lambda}_t^{1/2}(\sigma^2_{\mathrm{IC}}\mathbf{z}_{t-1}\mathbf{z}'_{t-1} + \sigma^2_\varepsilon\,\mathrm{I})\boldsymbol{\Lambda}_t^{1/2}. \tag{29}$$

Using a matrix inversion lemma to obtain a formula for $\boldsymbol{\Omega}_t^{-1}$ and substituting it along with above expression for $\boldsymbol{\alpha}_t$ into the expected information ratio formula (17) we have the following result.

Proposition 1. Given assumptions (A1), (A2), (A3), (A4) and (A5) in Section 2.4, the fundamental law of active management for a single factor model is given by the information ratio formula

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sqrt{\sigma^2_\varepsilon/N + \sigma^2_{\mathrm{IC}}}} = \frac{\mathrm{IC}}{\sqrt{(1-\mathrm{IC}^2-\sigma^2_{\mathrm{IC}})/N + \sigma^2_{\mathrm{IC}}}}. \tag{30}$$

A derivation of the first equality above is provided in Appendix A. The second equality follows from (22) for the case of a single factor model. This formula is in fact a highly accurate approximation to a formula based on the expected value of a complicated non-linear function of random exposures. This is discussed further in Appendix A.

The portfolio IR is positively related to the IC_t average IC (skill) and N, but inversely related to the IC_t volatility σ_{IC}. This result should not be surprising to students of modern portfolio theory. If a portfolio is based on a sufficiently large asset universe size N, the main risk to the portfolio comes from a bet on the alpha factor that generates an uncertain payoff stream (strategy risk) but none-the-less has a positive average payoff.

Corollary 1. When the number of assets N in the selection universe goes to infinity our fundamental law for a single factor model reduces exactly to.

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sigma_{\mathrm{IC}}}. \tag{31}$$

In this case the expected information ratio is proportional to IC as a measure of skill, but inversely proportional to the volatility σ_{IC} of the random IC_t values. Thus σ_{IC} is a basic active management risk measure and the larger it is for a given IC the lower the IR.

Note that as N becomes increasingly large one expects to get increasingly accurate estimates of IC, but referring to the general form (30) of our fundamental law one sees that at the same time the risk associated with estimation error due to single factor model error variance $\sigma^2_\varepsilon = 1 - \mathrm{IC}^2 - \sigma^2_{\mathrm{IC}}$ in the denominator of (30) disappears. However, the information coefficient volatility remains as an intrinsic active risk measure.

Qian and Hua (2004) referred to σ_{IC} as "strategy risk". In the context of our cross-section based single factor model there is a very natural way to think of strategy risk and associate it with IC volatility. Suppose a portfolio manager chooses a single factor such as book-to-market, earnings-to-price, or momentum, resulting in a set of random variables $z_{i,t-1}$ that have correlation coefficient IC_t with r_{it}. Some choices of the factor lead to relatively small IC_t volatilities σ_{IC} and hence relatively large IR, and other choices lead to larger volatilities and hence smaller IR's. In this context strategy risk is the risk of choosing a factor that has a large IC_t volatility σ_{IC}.

3.3. Constant information coefficient and the G & K formula

If we make the unreasonable assumption that the cross-sectional IC_t is constant over time then $\sigma_{\mathrm{IC}} = 0$ and we have

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sqrt{1-\mathrm{IC}^2}}\sqrt{N}. \tag{32}$$

This constant information coefficient version of the fundamental law states that an arbitrarily large expected information ratio IR can be obtained by increasing N arbitrarily, which is not intuitively reasonable. In this case the strategy becomes a money machine

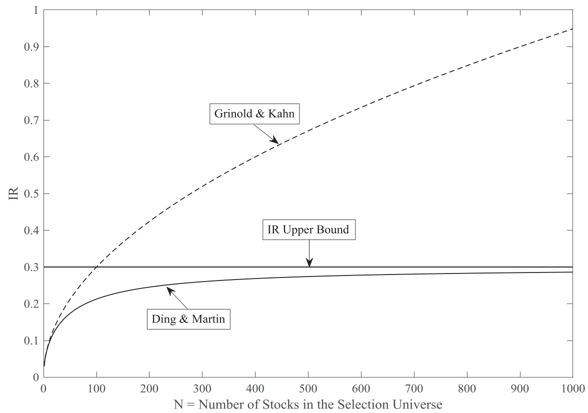

Fig. 1. Fundamental law of active management: comparisons (IC = 0.03, $\sigma_{\mathrm{IC}} = 0.1$).

that generates arbitrarily large expected active risk adjusted excess return every single period with arbitrarily small active risk.

Typical values of IC encountered in practice are less than one tenth. In that case (32) will be well approximated by the Grinold (1989) and Grinold and Kahn (2000) formula

$$\mathrm{IR}_{GK} = \mathrm{IC}\sqrt{N}. \tag{33}$$

3.4. Comparison of new fundamental law with G & K

Now we compare the behavior of the fundamental law IR given by (30) as a function of N with that of the special cases (31) for an infinite cross-section of assets, and with the G & K formula (33) for realistic values of IC and σ_{IC}. Fig. 1 plots the relationship between portfolio IR and N for these three formulas for the case where IC = 0.03 and $\sigma_{\mathrm{IC}} = 0.1$. For these values our zero IC volatility formula (32) is almost identical to the G & K formula (33), both of which increase without bound at the rate of the square root of N. But the assumption of zero volatility of the information coefficient is clearly unreasonable as is the resulting IR behavior as a function of N. A forecast signal's IC_t changes randomly over time and under this more realistic assumption one has $\sigma_{\mathrm{IC}} > 0$ and the IC volatility adjusted ratio $\mathrm{IC}/\sigma_{\mathrm{IC}}$ sets an upper bound to the IR one can achieve.

4. Multi-factor fundamental law and the impact of missing factors

In this section we extend the fundamental law to the case of a multivariate factor model as specified in Section 2.4. The result specifies the information ratio in terms of the mean value of a vector of randomly time-varying information coefficients $\mathbf{IC}_t$ and the associated covariance matrix $\boldsymbol{\Sigma}_{\mathrm{IC}}$ of the $\mathbf{IC}_t$. As in the case of the single factor model we provide a fundamental law formula for both a finite cross-section of assets and the limiting case as the number of assets tends to infinity. We also analyze the impact of using a factor model that is misspecified by virtue of not including all factors that are part of a true model. In this case we show that, not surprisingly, the achievable information ratio will typically be smaller than the information ratio that can be obtained with the full model.

4.1. The multi-factor fundamental law

Motivated by the common practice of fitting the model (19) via a cross-section regression, we assume that the random factor returns $\mathbf{f}_t$ are given by

$$\begin{aligned} \mathbf{f}_t &= \boldsymbol{\Sigma}_{\mathbf{z}}^{-1}\mathrm{cov}(\tilde{r}_{it}, \mathbf{z}'_{i,t-1}) \\ &= (\mathrm{cov}(\tilde{r}_{it}, z_{i1,t-1}), \mathrm{cov}(\tilde{r}_{it}, z_{i2,t-1}), \cdots, \mathrm{cov}(\tilde{r}_{it}, z_{iK,t-1}))' \\ &= (f_{1t}, f_{2t}, \cdots, f_{Kt})' \end{aligned} \tag{34}$$

which by definition is the same for all assets. Furthermore, since both $\tilde{r}_{it}$ and $\mathbf{z}_{i,t-1}$ have unit variances, the components of the factor returns are just the correlations of exposures components with the standardized residual returns. Thus they are information coefficients and so we use the notation

$$\mathbf{IC}_t \triangleq \mathbf{f}_t = (\mathrm{IC}_{1t}, \mathrm{IC}_{2t}, \ldots, \mathrm{IC}_{Kt})'. \tag{35}$$

Likewise, we replace the expected value $\mathbf{f} = E(\mathbf{f}_t)$ by $\mathbf{IC} = E(\mathbf{IC}_t)$, and use the notation $\boldsymbol{\Sigma}_{\mathrm{IC}}$ in place of $\boldsymbol{\Sigma}_f$.Thus we write the model in Eq. (20) as

$$\tilde{\mathbf{r}}_t = \mathbf{Z}_{t-1}\mathbf{IC}_t + \boldsymbol{\varepsilon}_t \tag{36}$$

and write the multi-factor model conditional mean forecast of $\mathbf{r}_t$ and conditional forecast error covariance respectively as

$$\boldsymbol{\alpha}_t = \boldsymbol{\Lambda}_t^{1/2}\mathbf{Z}_{t-1}\mathbf{IC} \tag{37}$$

and

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda}_t^{1/2}(\mathbf{Z}_{t-1}\boldsymbol{\Sigma}_{\mathbf{IC}}\mathbf{Z}_{t-1}' + \sigma_\varepsilon^2\mathbf{I})\boldsymbol{\Lambda}_t^{1/2}. \tag{38}$$

Appendix B shows that using an appropriate matrix inversion lemma to obtain a formula for $\boldsymbol{\Omega}_t^{-1}$ and substituting it along with above expression for $\boldsymbol{\alpha}_t$ into the expected information ratio formula (17) gives the following result.

Proposition 2. Given assumptions (A1), (A2), (A3), (A4) and (A5) in Section 2.4 the fundamental law of active management for a multi-factor model is given by the information ratio formula.

$$\mathrm{IR} = \sqrt{\mathbf{IC}'(\,\mathbf{I}\sigma_\varepsilon^2/N\ + \boldsymbol{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}}\,. \tag{39}$$

where the idiosyncratic risk variance is

$$\sigma_\varepsilon^2 = (1 - \sum\nolimits_{k=1}^{K}(\sigma_{\mathrm{IC},k}^2 + \mathrm{IC}_k^2)) \tag{40}$$

with $\sigma_{\mathrm{IC},k}^2$ the diagonal elements of $\boldsymbol{\Sigma}_{\mathbf{IC}}$ and IC_k the elements of $\mathbf{IC}$.

Corollary 2. When the number of assets N in the selection universe goes to infinity our fundamental law for a multi-factor model is given exactly by.

$$\mathrm{IR} = \sqrt{\mathbf{IC}'\boldsymbol{\Sigma}_{\mathrm{IC}}^{-1}\mathbf{IC}}\,. \tag{41}$$

As in the univariate case, the idiosyncratic part of the risk in (40) will be diversified away as N gets larger, and the remaining risk is the "strategy risk" represented by the factor return covariance matrix $\boldsymbol{\Sigma}_f = \boldsymbol{\Sigma}_{\mathrm{IC}}$ in (41) that cannot be diversified away. We note that the IR is a metric similar to the Mahalanobis distance that for a given covariance structure depends on the expected values of the random information coefficients. It measures the dissimilarity between the mean information coefficients vector $\mathbf{IC}$ and the vector of $\mathbf{0}$. The larger the $\mathbf{IC}$ in the coordinates implied by $\boldsymbol{\Sigma}_{\mathbf{IC}}$ the larger the IR.

4.2. The impact of missing factors

Consider the multi-factor model with two sub-groups of factors having randomly time-varying information coefficients $\mathbf{IC}_{1t}$ and $\mathbf{IC}_{2t}$. The corresponding mean information coefficient vectors are $\mathbf{IC}_1$, $\mathbf{IC}_2$ and the partitioned covariance matrix is

$$\boldsymbol{\Sigma}_{\mathrm{IC}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{42}$$

The asymptotic information ratio is

$$\mathrm{IR} = \sqrt{\mathbf{IC}'\boldsymbol{\Sigma}_{\mathrm{IC}}^{-1}\,\mathbf{IC}}\ = \sqrt{(\mathbf{IC}_1'\ \mathbf{IC}_2')\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{IC}_1 \\ \mathbf{IC}_2 \end{pmatrix}}. \tag{43}$$

Then with $\mathbf{E} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ positive definite, a matrix inversion formula for partitioned matrices[6] gives

$$\mathrm{IR} = \sqrt{\mathbf{IC}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{IC}_1 + (\mathbf{IC}_2 - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{IC}_1)'\mathbf{E}^{-1}(\mathbf{IC}_2 - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{IC}_1)}\,. \tag{44}$$

Note that

$$\mathrm{IR} \geq \sqrt{\mathbf{IC}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{IC}_1} \tag{45}$$

with strict inequality unless $(\mathbf{IC}_2 - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{IC}_1)'\mathbf{E}^{-1}(\mathbf{IC}_2 - \mathbf{E}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{IC}_1)$ is zero, which occurs when both $\mathbf{IC}_2 = \mathbf{0}$ and $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. Suppose the latter two conditions do not hold but the portfolio manager uses a factor model that contains only the factors in the first of the two groups when in fact the true model contains both groups. In that case the optimized portfolio will have an information ratio smaller than that achievable according to (44) by using the correct factor model.

Notice that when $\mathbf{IC}_2 = \mathbf{0}$ the second group of factors make no contribution to the alpha forecast given by (37). Nonetheless the second group of factors can make a positive contribution to the IR. This follows from the development below of expression (44) with $\mathbf{IC}_2 = \mathbf{0}$ [7]

[6] The inverse of a partitioned matrix is repeatedly used in the derivation, see Magnus and Neudecker (2002, p. 11).

[7] We get the second line of (46) using the following alternative expression for formula in (44) and set $\mathbf{IC}_2 = \mathbf{0}$:

$$\mathrm{IR} = \sqrt{\mathbf{IC}_2'\boldsymbol{\Sigma}_{22}^{-1}\mathbf{IC}_2 +\ (\mathbf{IC}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{IC}_2)'\mathbf{D}^{-1}(\mathbf{IC}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{IC}_2)}\ \text{with}\ \mathbf{D} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'.$$

$$\begin{aligned}\mathrm{IR} &= \sqrt{\mathbf{IC}_1'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1 + \mathbf{IC}_1'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{12}'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1} \\ &= \sqrt{\mathbf{IC}_1'(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}')^{-1}\mathbf{IC}_1} \\ &\geq \sqrt{\mathbf{IC}_1'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1}\ . \end{aligned} \tag{46}$$

From the second line of the above equation, it can be seen that the proper covariance matrix to use is $(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}')$ instead of just $\mathbf{\Sigma}_{11}$. When portfolio manager fails to include the second group of risk factors in the true model, it can result in a lower IR than is achievable using the second group of factors. The intuitive reason for this is that even when $\mathbf{IC}_2 = \mathbf{0}$ correlation between $\mathbf{IC}_{1t}$ and $\mathbf{IC}_{2t}$ can result in lower effective volatility of the first group of information coefficients.

It should also be noted that in the highly idealized case where $\mathbf{IC}_{1t}$ and $\mathbf{IC}_{2t}$ are uncorrelated $\mathbf{\Sigma}_{12} = \mathbf{0}$ the IR expression in (44) reduces to

$$\mathrm{IR} = \sqrt{\mathbf{IC}_1'\mathbf{\Sigma}_{11}^{-1}\mathbf{IC}_1 + \mathbf{IC}_2'\mathbf{\Sigma}_{22}^{-1}\mathbf{IC}_2} = \sqrt{\mathrm{IR}_1^2 + \mathrm{IR}_2^2}. \tag{47}$$

5. Empirical factor IR comparison

In this section we present an empirical study of the validity of our fundamental law IR finite cross-section formula (30) and infinite cross-section formula (31), as well as the G & K formula (33), for single factor models given by (24) for a large cross-section of Russell 3000 stock returns.

For each stock the residual returns are obtained as the residuals from a rolling OLS regression of stock excess return on market excess return using 5-years of monthly data. If a company has less than 5-years of data at the time of calculation, we assume a beta of one and compute the residual returns as the difference of excess returns.

For each of the eight single factor models studied we use as exposures one of the eight quantitative factors displayed in Table 1. These factors are some of the most commonly used ones in quantitative portfolio management practices.

Table 1
Factor descriptions.

Factor	Description	Sample period studied
B/P	Book to price ratio	1979–01 to 2010–06
C/P	Cash flow to price ratio	1990–02 to 2010–06
D/P	Dividend yield	1979–01 to 2010–06
E/P	Earnings to price ratio	1979–01 to 2010–06
FE/P	Forward earnings to price ratio	1979–01 to 2010–06
S/P	Sales to price ratio	1979–01 to 2010–06
MOM	Cumulative 11-month return from *t*-12 to *t*-2	1979–01 to 2010–06
SHORT	Short as a percent of total shares floating	1988–01 to 2010–06

There are two main purposes of our empirical study. The first is to show that our fundamental law asymptotic and finite sample formulas both give a much more accurate estimate of the portfolio IR than the classic G & K formula $\mathrm{IR}_{GK} = \mathrm{IC}\sqrt{N}$. In order to establish these results we use as a reference the IR obtained via direct use of the active portfolio weights to obtain the expected IR by actual portfolio simulation. The second purpose is to show that our new model (24) results in substantially higher information ratios than that obtained with a commonly used industry method that does not standardize residual returns and does not properly standardize exposures.

For each single factor model whose factor names are given in Table 1 we use as investment universe the Russell 3000 over the time intervals shown. The Russell 3000 measures the performance of the largest 3000 U.S. companies representing approximately 98% of the investable U.S. equity market. To reduce the impact of exposures data errors we apply 1% Winsorizing of the raw factor exposures.[8]

5.1. Standardization of residual returns and exposures

One of the most important modeling assumptions in Section 3 leading to our fundamental law is that *one unit of risk-adjusted exposure is expected to get rewarded (paid off) with the same amount of risk-adjusted residual return across securities*, and correspondingly the single factor model has the form

$$\tilde{r}_{it} = f_t z_{i,t-1} + \varepsilon_{it},\ i = 1, \cdots, N,\ t = 1, \cdots, T.$$

In this model the dependent variable $\tilde{r}_{it}$ is a random variable with mean zero and variance one conditioned on past information that excludes the exposure variable $z_{i,t-1}$, and the latter is a random variable with zero mean and unit variance. It is relatively easy to obtain the standardized residual returns $\tilde{r}_{it}$ using a GARCH model to compute $\sigma_{r_{it}}$ based on past values r_{iu}, $u \leq t-1$ for each $i = 1, \cdots, N$. For our

[8] An $\alpha \times 100$ percent Winsorized mean $\hat{\mu}_{win,\alpha}$ with $0 \leq \alpha < 0.5$ is defined as follows for a sample of size n of ordered data $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Let $k = \lfloor \alpha{\cdot}n \rfloor$ be the smallest integer greater or equal to $\alpha{\cdot}n$. Then $\hat{\mu}_{win,\alpha} = \frac{1}{n}(k{\cdot}(x_{(k)} + x_{(n-k+1)}) + \sum_{i=k+1}^{n-k} x_{(i)})$.

empirical results below we used a GARCH(1,1) model with fixed parameter values that are similar to those typically obtained for S & P 500 index returns.[9] However, obtaining the standardized $z_{i,t-1}$ from the raw exposures $x_{i,t-1}$ is more challenging, as we now discuss.

5.1.1. An inappropriate standardization used by practitioners

Before describing our method of standardizing the raw x_{it} exposures we first describe the cross-section method of standardizing exposures such as book-to-price ratio and earnings-to-price ratio used by most practitioners and by commercial software portfolio optimization products for many years. The cross-section standardization of the raw exposure x_{it} for asset i at time t computes

$$\tilde{x}_{it} = \frac{x_{it} - \bar{x}_t}{\delta_{x_t}} \tag{48}$$

where

$$\bar{x}_t = \frac{1}{N}\sum_{i=1}^{N} x_{it} \tag{49}$$

and

$$\delta_{x_t} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{it} - \bar{x}_t)^2}\,. \tag{50}$$

It is then commonly assumed that $\tilde{x}_{it}$ is a random variable with mean zero and variance one. This assumption would be reasonable if each raw exposure had the same expected value and the same standard deviation at time t, i.e., if $x_{it}\sim(x_t,\ \sigma_{x_t}^2)$ for all $i = 1, \cdots, N$. In that case it follows from the law of large numbers that the cross-section sample mean $\bar{x}_t$ converges in probability to the true mean x_t, and the cross-section sample dispersion δ_{x_t} converges in probability to the common exposure standard deviation σ_{x_t}. Consequently the above cross-section standardization results in $\tilde{x}_{it}$ being approximately a (0,1) random variable for a cross-section of size N and exactly so as $N \to \infty$.

However, when the raw exposures do not have the same mean and variance for each asset at time t the above standardization will not result in $\tilde{x}_{it}$ being a (0, 1) random variable. For example if the raw exposures have the same mean but different variances, i.e., $x_{it}\sim(x_t,\ \sigma_{x_{it}}^2)$, then the above cross-section standardization will result in $\tilde{x}_{it}$ being approximately a $(0,\ \sigma_{x_{it}}^2/\delta_{x_t}^2)$ random variable. The result will be that the i-th asset will not be represented by its proper variance, with potential adverse effects in the cross-section regression. The importance of this problem is illustrated by the following example.

Fig. 2 shows an example of the residual returns and raw earnings-to-price ratio (E/P) exposures of two companies, Tesoro Corporation (TSO) and Walgreen Co. (WAG). Tesoro is an oil company while Walgreen is a drug store. The same vertical axis ranges are used for the pair of residual returns plots and likewise for the E/P plots.

Now consider the same set of residual returns r_t and the cross-section standardized E/P exposures $\tilde{x}_t$ in Fig. 3. In computing the cross-section standardized exposures we used a common exposures data-cleaning practice to avoid undue influence of outliers. Specifically we first compute the cross-section mean and standard deviation of the 1% Winsorized raw exposures data. Then we standardize the exposures by subtracting the resulting mean and dividing by the resulting standard deviation. Finally any cross-section standardized exposures that are greater than 3 are set equal to 3, and any that are less than −3 are set equal to −3. This results for example in the constant values of −3 for the TSO cross-section standardized values between the end of 1985 and the middle of 1987 in Fig. 3.

Table 2 displays the sample standard deviations (volatilities) σ_r of the residual returns, the sample standard deviations (volatilities) $\sigma_{\tilde{x}}$ of the cross-section standardized E/P exposures and the ratios of the latter to the former, for both TSO and WAG.

Tesoro's residual return volatility is 1.94 times higher than that of Walgreen's, but Tesoro's E/P ratio volatility is 16.67 times higher than that of Walgreen's. Suppose, as is common practice, the factor returns f_t are computed using a Fama-MacBeth style cross-section regression of the raw residual return at a given time on the raw exposure at the previous time. Suppose further that predicted factor return $\hat{f}_t$ is based on the sample mean of the factor returns up to and including time t-1, which depend on exposures up to and including time t-2. Then the predicted residual return (alpha) for each asset is the product of $\hat{f}_t$ (which is the same for each stock) and the asset's cross-section standardized exposure $\tilde{x}_{t-1}$. Consequently the volatility of Tesoro's predicted residual return will be 16.67 times the volatility of Walgreen's predicted residual return. This is obviously a problematic prediction since the actual residual return volatility for Tesoro is only 1.94 times that of Walgreen.

To make matters worse, the expected value of x_{it} might be different for different assets, for example depending on the life cycle of a company, or depending on industry it belongs to. One might expect a lower earnings yield for a new growth company in high tech industry than for an established value company in consumer staple industry. Practitioners usually subtract different mean values for different industries to deal with this problem. However the details of this approach for different factors are beyond the scope of the current paper. In our empirical study we will for convenience simply assume that the x_{it} have a common expected value x_t for period t. In this regard see discussion point (D2) in Section 2.4.

[9] A GARCH(1,1) process for residual return r_{it} has conditional variance $\sigma_{it}^2 = \sigma_i^2(1 - \alpha - \beta) + \alpha r_{i,t-1}^2 + \beta\sigma_{i,t-1}^2$ and we used the fixed parameter values $\alpha = .1$, $\beta = .81$

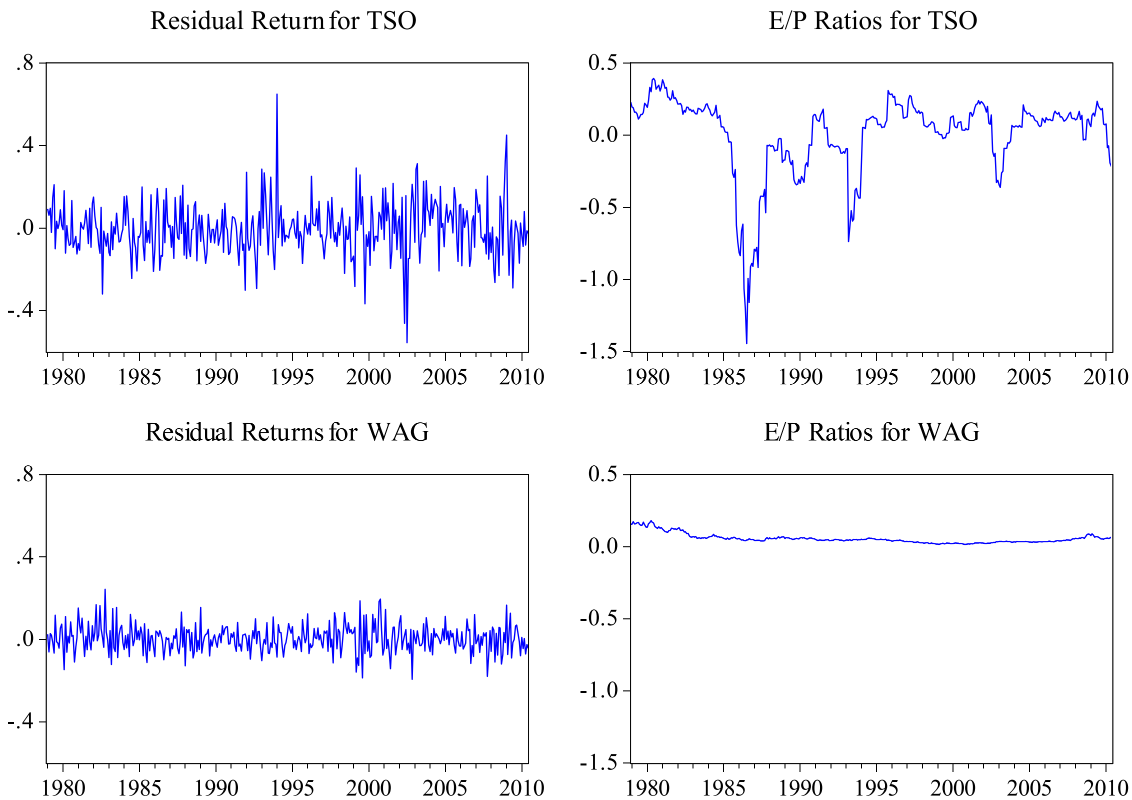

Fig. 2. Residual returns and E/P time series.

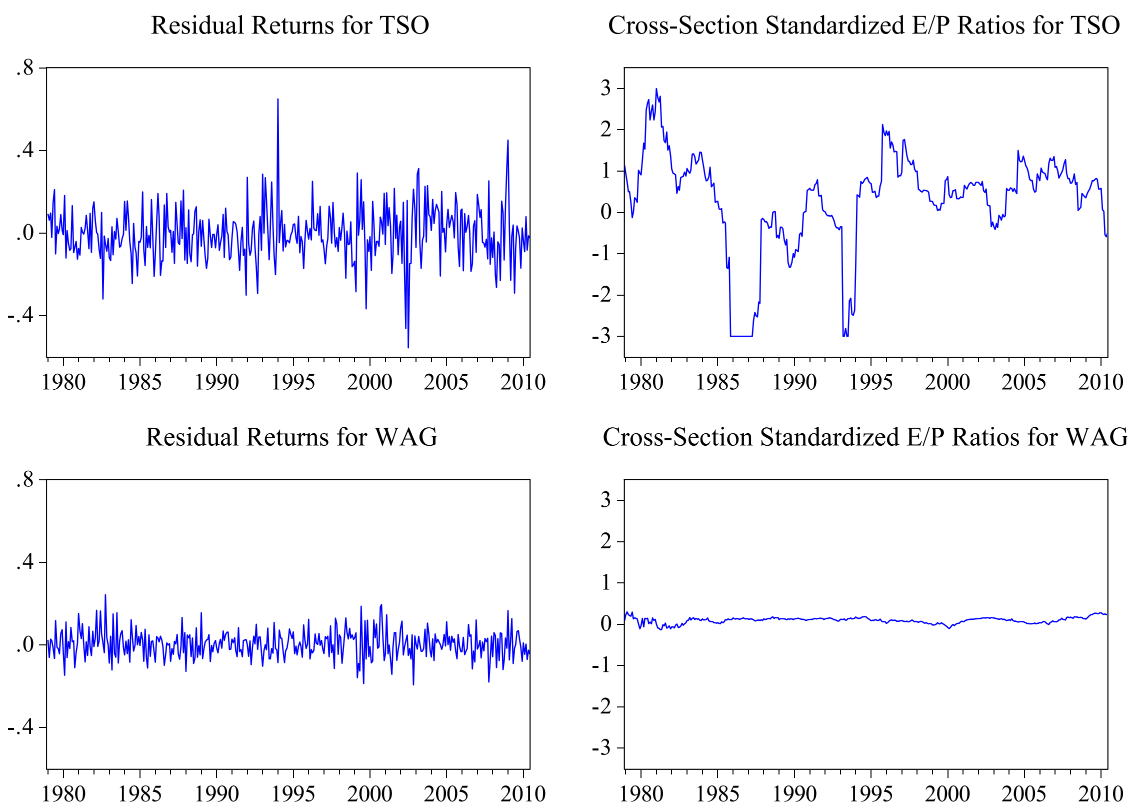

Fig. 3. Residual returns and cross-section standardized E/P ratios.

Table 2
TSO and WAG volatilities of residual returns and cross-section standardized E/P.

Ticker	σ_r	$\sigma_{\tilde{x}}$	$\sigma_{\tilde{x}}/\sigma_r$
TSO	0.128	1.217	9.51
WAG	0.066	0.073	1.11
TSO/WAG	1.94	16.67	

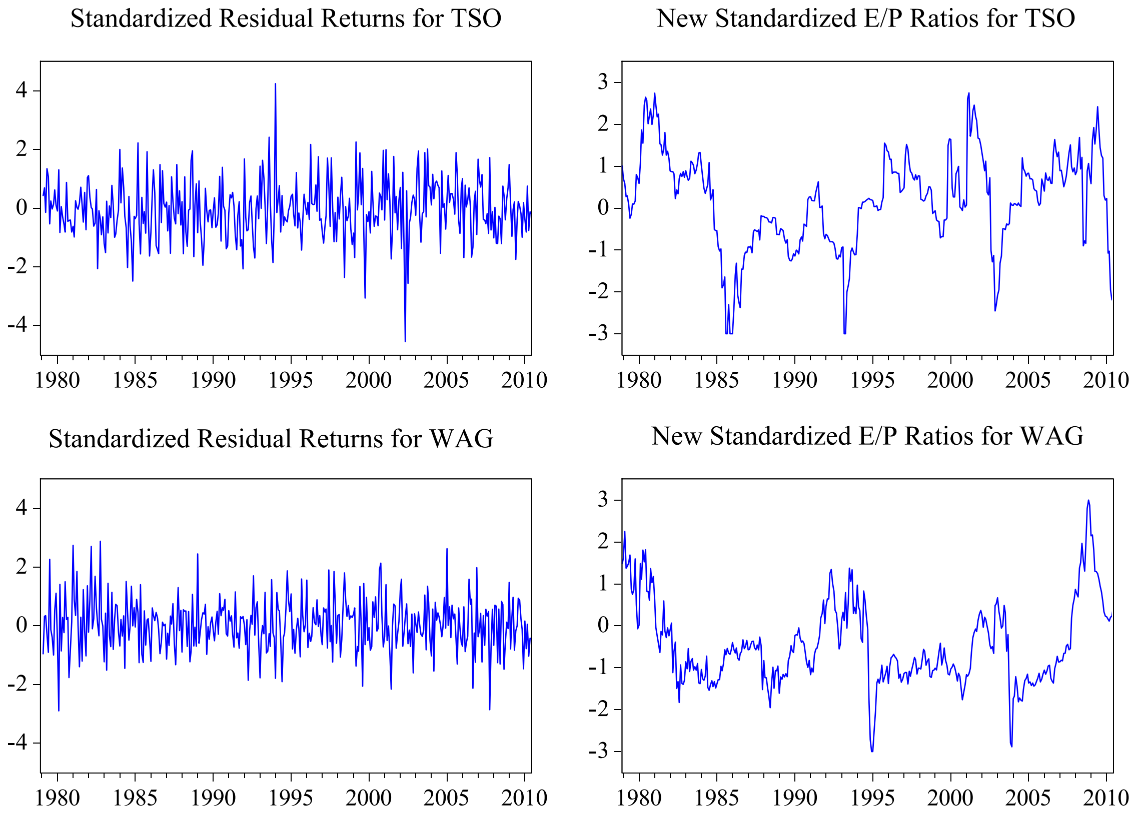

Fig. 4. Standardized residual returns and new standardized E/P ratios.

5.1.2. A better standardization of exposures

In order to avoid the problem caused by different assets having different exposure volatilities we standardize the exposures as follows. For each asset i we compute the time series $(x_{i,t-1} - \bar{x}_{t-1})^2$ where $\bar{x}_{t-1}$ is the cross-section mean of the exposures at time t-1. Then we use the time series $(x_{i,t-1} - \bar{x}_{t-1})^2$ to compute an EWMA estimate $s^2_{x_{i,t-1}}$ as follows.[10]

- For companies having at least five years of monthly history, use the full history to compute $s^2_{x_{i,t-1}}$.
- For companies that have less than five years of monthly data as of the date for which the EWMA estimate is needed we compute $s^2_{x_{i,t-1}}$ for simplicity as follows. If the company survived for more than five years we compute based on five years of the monthly $(x_{i,t-1} - \bar{x}_{t-1})^2$ values starting at the company's inception date. If the company survived for less than 5 years we compute $s^2_{x_{i,t-1}}$ as the sample mean of the monthly values of $(x_{i,t-1} - \bar{x}_{t-1})^2$ for the entire sample. We note that this method for short histories results in look-ahead bias and an alternative method without this feature should be used in practice.

Finally we obtained the standardized exposures $z_{i,t-1}$ by standardizing $x_{i,t-1}$ using $\bar{x}_{t-1}$ as mean and $s_{x_{i,t-1}}$ as exposure standard deviation

[10] The EWMA formula used is $s_{x_{i,t}} = \sqrt{0.1\,(x_{i,t} - \bar{x}_t)^2 + 0.9\, s^2_{x_{i,t-1}}}$.

Table 3
Volatilities of residual returns and new standardized E/P ratios.

Ticker	$\sigma_{\tilde{r}}$	σ_z	$\sigma_z/\sigma_{\tilde{r}}$
TSO	0.981	1.133	1.15
WAG	0.893	0.992	1.11
TSO/WAG	1.10	1.14	

$$z_{i,t-1} = \frac{x_{i,t-1} - \bar{x}_{t-1}}{s_{x_{i,t-1}}}. \tag{51}$$

With this standardization of the raw exposures, which we refer to as the "New" standardization, the $z_{i,t-1}$ should be close to a (0,1) random variable for each of the assets.

Now consider the Fig. 4 plots of standardized residual returns and standardized exposures using our new exposures standardization method. The standardized exposures for TSO and WAG now look reasonably like (0,1) random variables, as do the standardized returns. Table 3 displays the sample standard deviations (volatilities) $\sigma_{\tilde{r}}$ of the standardized residual returns $\tilde{r}_t$, the sample standard deviations (volatilities) σ_z of our new standardized E/P exposures z_t and the ratios of the latter to the former. Now Tesoro's standardized residual return volatility is only 1.10 times that of Walgreen's, and Tesoro's E/P ratio volatility is only 1.14 times that of Walgreen's. Now use of a common predictor value $\hat{f}_t$ to multiply the exposures $z_{TSO,t-1}$ and $z_{WAG,t-1}$ to compute predictors of $\tilde{r}_{TSO,t}$ and $\tilde{r}_{WAG,t}$ will result in predictor volatilities of similar size for TSO and WAG, thereby avoiding the problem revealed in Table 2 that occurred as a result of using cross-section standardized exposures.

5.2. The new model versus the industry standard model

In order to demonstrate the advantage of our model (24) based on standardization of residual returns and standardization of raw exposures on our investment universe from the Russell 3000, we study the behavior of IC, σ_{IC} and the various IR formulas for both our new model (NEW), and for the conventional model (ISM) that does not standardize returns and uses cross-section standardization of exposures. For both NEW and ISM models the residual returns are computed as described at the beginning of this section. For the NEW model the residual returns standard deviation estimates are computed as described in the first paragraph of Section 5.1.

Our new model (NEW) uses both the standardized residual returns $\tilde{r}_{it}$ and standardized exposures $z_{i,t-1}$ where the latter is based on cross-section centering and asset by asset volatility standardization given by (51). For our single factor model the factor returns are

$$f_t = \text{cov}(\tilde{r}_{it}, z_{i,t-1}) \tag{52}$$

and with (26) in mind we estimate the time series IC_t as the sample-based cross-section correlations

$$\widehat{\text{IC}}_t = \widehat{corr}(\tilde{r}_{it}, z_{i,t-1}). \tag{53}$$

Fig. 5 shows the time series plots of IC_t for our new model for the eight factors. The number of stocks used for these computations varies over time and the average number of stocks used for each factor is displayed in Table 4. The middle horizontal line is located at the sample mean of the IC_t time series, and the upper and lower horizontal lines represent the 95% confidence intervals at each time t. These confidence intervals indicate that the time variation of the information coefficients is so large that it cannot be simply attributed to estimation errors. Thus the factor returns are a random process as we assumed and a portfolio manager can incur a substantial amount of loss at any moment when she bets on any of these factors. However, as shown by the middle horizontal line in Fig. 5 and the corresponding values in Table 4, the average information coefficient for all these eight factors are all positive and one can make a positive average return by consistently betting on these factors over very long time periods.

In the industry standard model (ISM) the residual returns are not standardized and the exposures are standardized in a purely cross-section manner described in Section 5.1. So for the ISM model we estimate the time series IC_t as the sample-based cross-section correlations

$$\widehat{\text{IC}}_t = \widehat{corr}(r_{it}, \tilde{x}_{i,t-1}). \tag{54}$$

5.3. The empirical results

For each of the eight factors we did the following. We used the cross-section of residual returns to compute the IC_t time series values with (53) for the NEW model and with (54) for the ISM model for each of the eight factors using all the stocks in the corresponding Russell 3000 universe. Then we used those IC_t time series to compute the following:

The time series sample mean estimate $\widehat{\text{IC}}$ of the expected value of the IC = $E(\text{IC}_t)$ and the corresponding sample standard deviation estimate $\hat{\sigma}_{\text{IC}}$ of the volatility σ_{IC} of IC_t.

The Grinold and Kahn information ratio estimator with breadth BR taken to be number of assets N:

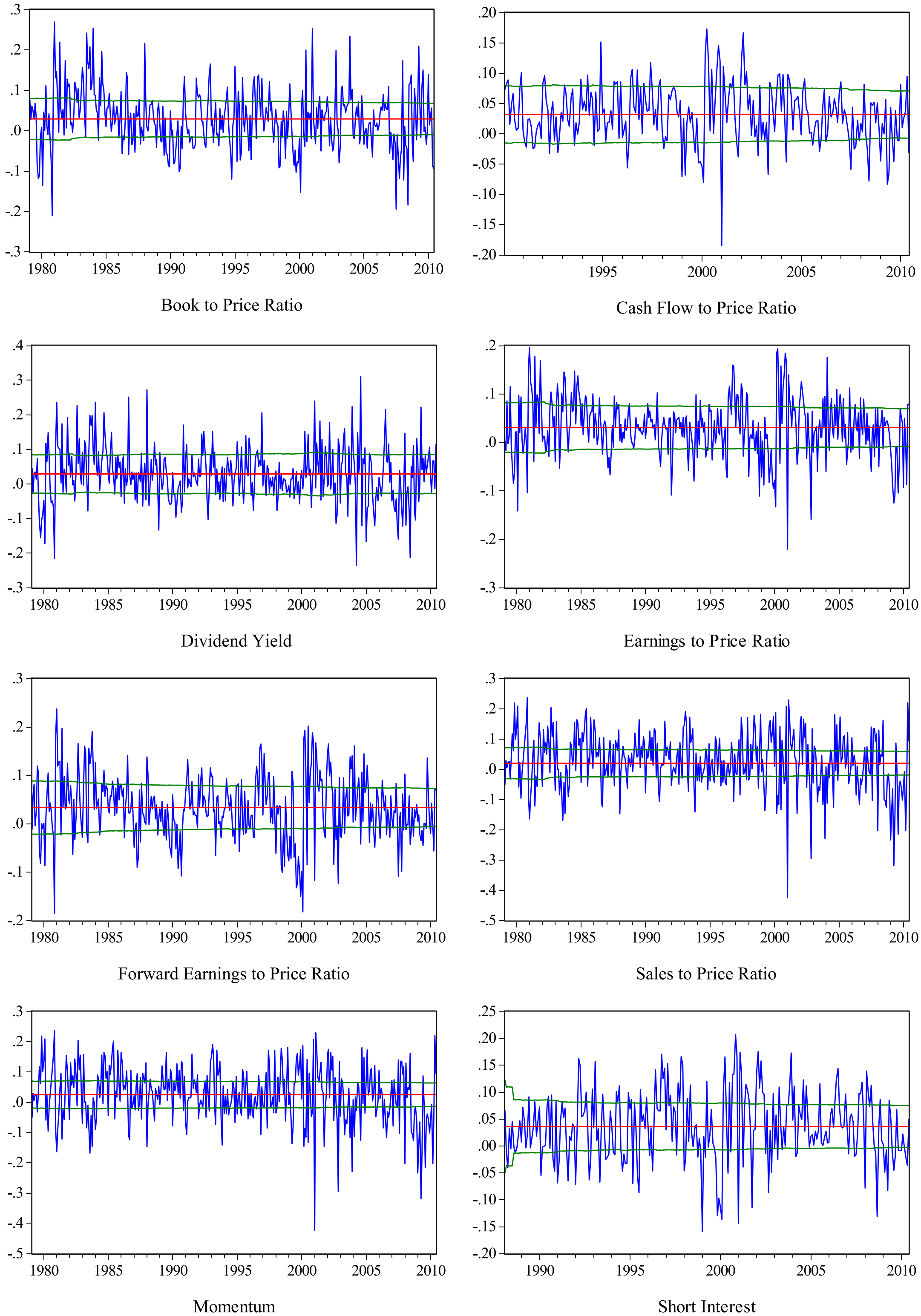

Fig. 5. Time series plot of new model IC_t estimates.

Table 4
IC and IR values for eight fundamental factors and two models.

Factor	Model	N	$\widehat{\mathrm{IC}}$	$\hat{\sigma}_{\mathrm{IC}}$	$\widehat{\mathrm{IR}}_{\mathrm{GK}}$	$\widehat{\mathrm{IR}}_{\infty}$	$\widehat{\mathrm{IR}}_{N}$	$\widehat{\mathrm{IR}}_{\mathrm{sim}}$
B/P	NEW	2088	0.029	0.074	1.337	0.394	0.378	0.395
B/P	ISM	2089	0.017	0.089	0.783	0.193	0.188	0.162
C/P	NEW	2006	0.032	0.049	1.432	0.650	0.592	0.523
C/P	ISM	2007	0.039	0.081	1.727	0.478	0.461	0.255
D/P	NEW	1244	0.029	0.080	1.014	0.359	0.338	0.318
D/P	ISM	1244	0.014	0.086	0.502	0.165	0.157	0.157
E/P	NEW	2095	0.031	0.061	1.411	0.504	0.475	0.421
E/P	ISM	2096	0.035	0.102	1.608	0.344	0.336	0.175
FE/P	NEW	1986	0.034	0.068	1.494	0.496	0.471	0.430
FE/P	ISM	1987	0.030	0.096	1.352	0.316	0.308	0.188
S/P	NEW	2078	0.020	0.068	0.927	0.301	0.286	0.287
S/P	ISM	2080	0.014	0.089	0.627	0.154	0.150	0.137
MOM	NEW	2151	0.025	0.095	1.162	0.263	0.257	0.217
MOM	ISM	2153	0.035	0.126	1.603	0.275	0.271	0.149
SHORT	NEW	2150	0.036	0.066	1.688	0.555	0.528	0.410
SHORT	ISM	2152	0.038	0.079	1.751	0.480	0.463	0.326

$$\widehat{\mathrm{IR}}_{\mathrm{GK}} = \widehat{\mathrm{IC}}\sqrt{N}.$$

The asymptotic information ratio (31) estimator for the limiting case for an infinite cross-section of assets:

$$\widehat{\mathrm{IR}}_{\infty} = \widehat{\mathrm{IC}}/\hat{\sigma}_{\mathrm{IC}}.$$

The finite cross-section of assets information ratio (30) estimator:

$$\widehat{\mathrm{IR}}_{N} = \frac{\widehat{\mathrm{IC}}}{\sqrt{(1-\widehat{\mathrm{IC}}^2-\hat{\sigma}_{\mathrm{IC}}^2)\big/N+\hat{\sigma}_{\mathrm{IC}}^2}}.$$

The information ratio $\mathrm{IR}_{\mathrm{sim}}$ for a simulated optimal portfolio for the NEW model and for the ISM model as follows.

For the NEW model we use the approximate active portolio weight formula (B-19):

$$\Delta w_{it} \approx \frac{\sigma_{A,t}}{\sigma_{\mathrm{IC}}}\frac{z_{i,t-1}}{N\sigma_{r_{it}}}$$

and corresponding portfolio active return $r_{A,t} = \sum_{i=1}^{N}\Delta w_{it} r_{it}$, where r_{it} is the residual return of security i at time t. The $z_{i,t-1}$ are standardized as in the last paragraph of Section 5.1. The unknown information coefficient volatility is replaced by an estimate $\hat{\sigma}_{\mathrm{IC}}$ as in the first paragraph above. Finally the portfolio tracking error is set at annualized value of $\sigma_{A,t} = 6\%$.

The procedure for the ISM model is similar to that of the NEW model except it is based on the different active weights formula

$$\Delta w_{it} \approx \frac{\sigma_{A,t}}{\sigma_{\mathrm{IC}}}\frac{\tilde{x}_{i,t-1}}{N}.$$

This formula does not have a factor $\sigma_{r_{it}}$ in the denominator because the ISM model does not use standardized residual returns, and the $\tilde{x}_{i,t-1}$ are conventional cross-section standardized exposures. The tracking error is again set at 6% on an annualized basis and $\hat{\sigma}_{\mathrm{IC}}$ is obtained just as for the NEW model.

The results of computing the above quantities are presented in Table 4 where some immediate conclusions are evident. First of all, the values of $\widehat{\mathrm{IR}}_{GK}$ range from a minimum of 2.2 times larger than $\widehat{\mathrm{IR}}_{\infty}$ to a maximum of 5.83 times larger than $\widehat{\mathrm{IR}}_{\infty}$, with a mean of 3.62. This confirms what we already know, namely that the assumption of a zero volatility IC_t is quite unreasonable and use of that assumption leads to grossly over-optimistic G & K information ratios.

We further note that in all cases $\widehat{\mathrm{IR}}_{\infty} > \widehat{\mathrm{IR}}_{N}$ as must be the case given their definitions. Furthermore for our NEW model we see that $\widehat{\mathrm{IR}}_{N} > \widehat{\mathrm{IR}}_{\mathrm{sim}}$ except for the S/P model where $\widehat{\mathrm{IR}}_{N}$ and $\widehat{\mathrm{IR}}_{\mathrm{sim}}$ are equal to two significant digits, and for B/P where $\widehat{\mathrm{IR}}_{N} < \widehat{\mathrm{IR}}_{\mathrm{sim}}$. This shows that while the finite N formula improves on the infinite N, the finite N formula appears to be typically an over-estimate. This over-estimation can be attributed to one or both of the following: (1) estimation error in $\widehat{\mathrm{IR}}_{N}$ and, (2) model misspecification in the sense that the data may require a multi-factor model. This issue is in need of further study.

Finally we note that for all of the factors except MOM, all of the values of $\widehat{\mathrm{IR}}_{\infty}$ and $\widehat{\mathrm{IR}}_{N}$ for our NEW model are substantially better than those for the conventional ISM model. And for all the factors our NEW model is better than the ISM model in terms of $\widehat{\mathrm{IR}}_{\mathrm{sim}}$. Across all the factors the ratios of $\widehat{\mathrm{IR}}_{N}$ for the NEW model to the $\widehat{\mathrm{IR}}_{N}$ for the ISM model range from 0.95 to 2.15 with a mean of 1.55, and for $\widehat{\mathrm{IR}}_{\mathrm{sim}}$ these ratios range from 1.26 to 2.44 with a mean of 2.00. These substantiate our choice of a factor model with time series standardization of both residual returns and proper standardization of exposures.

6. Concluding comments

We have derived a new version of the fundamental law of active management. The formula for the fundamental law is a function of the number of assets in the portfolio and the mean and variance of randomly time varying information coefficients. The only requirement on the latter is that they are realizations of a wide-sense stationary process.

As in a number of previous studies on the fundamental law, residual returns relative to a benchmark are used as inputs to portfolio optimization and the active management constraint is that the active portfolio weights sum to zero. However, our approach is based on a number of assumptions that are unique and important. First, we assume a random alpha that has zero expected value but generally non-zero conditional expected value, the latter of which may be exploited by superior forecasting. This assumption is very much in the spirit of Treynor and Black (1973) who in essence provided the initial impetus to the development of active portfolio management methods. Furthermore we make use of cross-section factor models with the following characteristics: (a) standardized residual returns in place of the usual non-standardized residual returns; (b) factor exposures are properly standardized based on a simple model for random exposures; (c) the time series of covariances between the standardized residuals and the factor model standardized exposures are assumed to be realizations of a wide-sense stationary stochastic process; (d) the cross-section factor model is used to compute conditional mean alpha forecasts and their related conditional covariance matrix.

The assumptions (a) and (b) for our factor model have the convenient feature that the error variances are constant in the cross-section. A consequence of (c) is that the factor returns and the resulting time series of information coefficients are stationary time series that have constant means and variances. The conditional forecast and covariance matrix in (d) are used as inputs to the calculation of mean-variance optimal active portfolio weights.

The fundamental law of active management is then derived using our framework and expressed in formula (30) for the case of a single factor model and a cross-section of assets of size N, with (31) as the corresponding asymptotic formula for an infinite cross-section size. The first of these two formulas show that the information ratio increases with increasing mean information coefficient IC, decreasing IC_t volatility σ_{IC} and increasing cross-section size N, all of which are intuitively appealing behaviors. It is to be noted that the denominator term in (30) that goes to zero like $1/N$ reflects model risk associated with the variability in fitting the factor model from observed data, which is also intuitively appealing. An important aspect of our results is that asymptotic IR formula (31) sets an absolute upper bound for the portfolio IR a portfolio manager can achieve with randomly time varying information coefficients.

It is important to keep in mind that making the unreasonable assumption of constant non-random information coefficients results in an information coefficient volatility $\sigma_{\mathrm{IC}} = 0$ for the single factor model. This results in the formula (32) which for typical IC values encountered in practice is well approximated by the G & K formula (33), both of which are proportional to $\sqrt{N}$ and hence unbounded as the number of assets tends to infinity. This is quite an unreasonable behavior, and it is reflected in the overly optimistic values of the G & K information ratio estimates values $\widehat{\mathrm{IR}}_{GK}$ reported in Table 4 of our Section 5 empirical study.

It should be noted that Qian and Hua (2004) derived a formula of the form (31) but without a clear econometric model basis for a randomly varying IC_t and without the use of a conditional covariance matrix for the conditional mean forecast, and with the overly restrictive and unrealistic assumption that the portfolio optimization uses constraints that neutralize all factor risks. On the other hand their work provided considerable stimulation for our derivation herein of the new fundamental law.

Our extension of the single-factor fundamental law to the multi-factor case resulted in information ratio formulas (39) and (41) as multivariate versions of (30) and (31). When the correlations between different groups of randomly time varying information coefficients is zero, Eq. (41) shows that the squared information ratio is the sum of the squared information ratios of the portfolios for each of the groups, a decomposition that was also established in G & K. More importantly our multi-factor results show that use of a multi-factor model that does not include all of the factors of a true model results in a lower information ratio than if all the factors are included in the model. Furthermore this is the case even when the mean values of the information coefficients associated with the missing factors are zero, but the correlation between the randomly time varying information coefficients of the missing factors and the included factors is not zero. This underscores the need for the portfolio manager to choose factor model exposures using statistical model selection methods. It remains as a topic of significant further research to study the empirical behavior of the multi-factor fundamental law formulas.

One insight from this paper is that the most important risk for a quant manger is the "strategy risk" of betting on certain alpha factors, i.e., including in a multi-factor model the exposures of selected factors whose information coefficients have relatively high volatility. Even though the “strategy risk” takes a small portion of the total risk for each individual asset, it is a common factor risk for the portfolio that cannot be diversified away. Portfolio managers should include the strategy risk in their risk model and try to play well (high mean information coefficients) and play precisely (low information coefficients volatilities).

Acknowledgements

We thank Roger Clarke, Harindra de Silva, Yuan Ding, Rob Engle, Russell Fuller, Christopher Green, John Kling, Wenling Lin, Jose Menchero, Edward Qian, Doug Stone, Wei Su, Yixiao Sun, Steven Thorley, Yining Tung, Jia Ye for helpful discussions and comments. We very much appreciate the contribution of an anonymous reviewer, whose very careful reviews and suggestions led to two substantial revisions of the paper to its present form. Last, but not least, we are very grateful to Richard Baillie for his thoughtful guidance leading to the publication of this paper.

Appendix A

For the single factor model the conditional mean forecast is

$$\boldsymbol{\alpha}_t = \boldsymbol{\Lambda}_t^{1/2} \cdot \mathrm{IC} \cdot \mathbf{z}_{t-1} \tag{A-1}$$

and conditional forecasting error covariance matrix is

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda}_t^{1/2}(\sigma_{\mathrm{IC}}^2 \mathbf{z}_{t-1}\mathbf{z}'_{t-1} + \sigma_\varepsilon^2\, \mathbf{I})\boldsymbol{\Lambda}_t^{1/2}. \tag{A-2}$$

The Sherman-Morrison matrix inversion formula gives

$$\boldsymbol{\Omega}_t^{-1} = \sigma_\varepsilon^{-2}\boldsymbol{\Lambda}_t^{-1/2}(\mathbf{I} - \phi\, \mathbf{z}_{t-1}\mathbf{z}'_{t-1})\boldsymbol{\Lambda}_t^{-1/2} \tag{A-3}$$

where

$$\phi = \frac{\sigma_{\mathrm{IC}}^2/\sigma_\varepsilon^2}{1 + \mathbf{z}'_{t-1}\mathbf{z}_{t-1}\sigma_{\mathrm{IC}}^2/\sigma_\varepsilon^2}. \tag{A-4}$$

Eq. (17) gives the general IR expression

$$\mathrm{IR} = E(\sqrt{\boldsymbol{\alpha}'_t\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\alpha}_t - \kappa\mathbf{1})}) = E(\sqrt{\boldsymbol{\alpha}'_t\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t - \boldsymbol{\kappa}\boldsymbol{\alpha}'_t\boldsymbol{\Omega}_t^{-1}\mathbf{1}}). \tag{A-5}$$

We obtain our fundamental law IR for the single factor model by analyzing the results of substituting the single factor model formulas (A-1) and (A-3) into (A-5). We proceed by evaluating the two terms under the radical sign separately, and in the process we will show that the second term is quite ignorable.

The first term is

$$\begin{aligned}
\boldsymbol{\alpha}'_t\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t &= \sigma_\varepsilon^{-2}\mathrm{IC}^2\ \mathbf{z}'_{t-1}\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\Lambda}_t^{-1/2}(\mathbf{I} - \phi\, \mathbf{z}_{t-1}\mathbf{z}'_{t-1})\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\Lambda}_t^{1/2}\, \mathbf{z}_{t-1} \\
&= \sigma_\varepsilon^{-2}\mathrm{IC}^2\ \mathbf{z}'_{t-1}(\mathbf{I} - \phi\, \mathbf{z}_{t-1}\mathbf{z}'_{t-1})\, \mathbf{z}_{t-1} \\
&= \sigma_\varepsilon^{-2}\mathrm{IC}^2\ \mathbf{z}'_{t-1}\mathbf{z}_{t-1}(1 - \phi\, \mathbf{z}'_{t-1}\mathbf{z}_{t-1}) \\
&= \frac{\mathrm{IC}^2}{\sigma_\varepsilon^2}\cdot\frac{\mathbf{z}'_{t-1}\mathbf{z}_{t-1}}{1 + \mathbf{z}'_{t-1}\mathbf{z}_{t-1}\sigma_{\mathrm{IC}}^2/\sigma_\varepsilon^2} \\
&= \frac{\mathrm{IC}^2}{\sigma_\varepsilon^2/\mathbf{z}'_{t-1}\mathbf{z}_{t-1} + \sigma_{\mathrm{IC}}^2} \\
&= \frac{\mathrm{IC}^2}{\sigma_\varepsilon^2/(N \cdot A_N) + \sigma_{\mathrm{IC}}^2}
\end{aligned} \tag{A-6}$$

where

$$A_N \triangleq \frac{1}{N}\mathbf{z}'_{t-1}\mathbf{z}_{t-1} = \frac{1}{N}\sum\nolimits_{i=1}^N z_{i,t-1}^2.$$

By the law of large numbers for identically distributed uncorrelated random variables we have

$$A_N \overset{P}{\to} E(z_{i,t-1}^2) = 1.$$

As for the second term under the radical sign we have:

$$\begin{aligned}
\kappa\boldsymbol{\alpha}'_t\boldsymbol{\Omega}_t^{-1}\mathbf{1} &= (\boldsymbol{\alpha}'_t\boldsymbol{\Omega}_t^{-1}\mathbf{1})^2/(\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\mathbf{1}) \\
&= \left(\frac{\mathrm{IC}}{\sigma_\varepsilon^2/\mathbf{z}'_{t-1}\mathbf{z}_{t-1} + \sigma_{\mathrm{IC}}^2}\cdot\frac{\mathbf{z}'_{t-1}\boldsymbol{\Lambda}_t^{-1/2}\mathbf{1}}{\mathbf{z}'_{t-1}\mathbf{z}_{t-1}}\right)^2\cdot\frac{1}{\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\mathbf{1}} \\
&= \left(\frac{\mathrm{IC}}{\sigma_\varepsilon^2/(N \cdot A_N) + \sigma_{\mathrm{IC}}^2}\cdot\frac{B_N}{A_N}\right)^2\cdot\frac{1}{\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\mathbf{1}}
\end{aligned} \tag{A-7}$$

where

$$B_N \triangleq \frac{1}{N}\mathbf{z}'_{t-1}\boldsymbol{\Lambda}_t^{-1/2}\mathbf{1} = \frac{1}{N}\sum\nolimits_{i=1}^N (z_{i,t-1}/\sigma_{r_{it}}).$$

By assumption (A1) of Section 2 we have $\sigma_{r_{it}} \geq b_{lo} > 0$ and by (A2) we have $E(z_{i,t-1}) = 0$, so by the law of large numbers for uncorrelated random variables with bounded variances that have a uniform upper bound we have $B_N \overset{P}{\to} 0$.The denominator of the second factor in (A-7) is

$$\begin{aligned}
\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\mathbf{1} &= \sigma_\varepsilon^{-2}\mathbf{1}'\boldsymbol{\Lambda}_t^{-1/2}(\mathbf{I} - \phi\ \mathbf{z}_{t-1}\mathbf{z}_{t-1}')\boldsymbol{\Lambda}_t^{-1/2}\mathbf{1} \\
&= \sigma_\varepsilon^{-2}\left(\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1} - \frac{\sigma_{\mathrm{IC}}^2}{\sigma_\varepsilon^2 + \mathbf{z}_{t-1}'\mathbf{z}_{t-1}\cdot\sigma_{\mathrm{IC}}^2}\cdot(\mathbf{z}_{t-1}'\boldsymbol{\Lambda}_t^{-1/2}\mathbf{1})^2\right) \\
&= \sigma_\varepsilon^{-2}\left(\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1} - \frac{\sigma_{\mathrm{IC}}^2}{\sigma_\varepsilon^2/(N\cdot A_N) + \sigma_{\mathrm{IC}}^2}\cdot\frac{N^2\cdot B_N^2}{N\cdot A_N}\right) \\
&= N\sigma_\varepsilon^{-2}\cdot\left(C_N - \frac{\sigma_{\mathrm{IC}}^2}{\sigma_\varepsilon^2/(N\cdot A_N) + \sigma_{\mathrm{IC}}^2}\cdot\frac{B_N^2}{A_N}\right)
\end{aligned} \tag{A-8}$$

where

$$C_N \triangleq \frac{1}{N}\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1} = \frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{1}{\sigma_{r_{it}}^2}.$$

Note that $\sigma_{r_{it}} \le b_{up} < \infty$ implies that $C_N \ge b_{up}^{-2}$. Since we have convergence in probability of B_N to zero and A_N to one, for large N we have

$$\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\mathbf{1} \sim N\sigma_\varepsilon^{-2}\cdot C_N \ge N\sigma_\varepsilon^{-2}\cdot b_{up}^{-2}.$$

Thus for large N the second term $\kappa\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1}$ under the radical sign in (A-5) behaves like

$$\kappa\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1} \sim \frac{1}{N}\cdot\left(\frac{\mathrm{IC}}{\sigma_\varepsilon^2/(N\cdot A_N) + \sigma_{\mathrm{IC}}^2}\cdot\frac{B_N}{A_N}\right)^2$$

which converges to zero considerably faster than B_N, and as such is ignorable compared with the first term $\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t$.

In order to confirm the above claim we carried out extensive simulations to show that the term $\kappa\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1} = (\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1})^2/(\mathbf{1}'\boldsymbol{\Omega}_t^{-1}\mathbf{1})$ is completely negligible compared with $\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t$. With specified values for IC, σ_{IC}, and $\boldsymbol{\Lambda}_t = diag(\sigma_{r_{it}}^2)$ we generate 1000 replicates of vector $\mathbf{z}_{t-1}$ consisting of N independent standard normal random variables. For each replicate we compute $\boldsymbol{\alpha}_t$ and $\boldsymbol{\Omega}_t^{-1}$ as given by (A-1) and (A-3) and thereby compute 1000 values of $\kappa\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1}$. For example, with IC = .03, $\sigma_{\mathrm{IC}} = .10$, and $\boldsymbol{\Lambda}_t = diag(\sigma_{r_{it}}^2)$ with the $\sigma_{r_{it}}$ drawn from a uniform distribution on [0.15, 0.35] we get the statistics for the 1000 replicates shown in the table below. Typical values for the first term $\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t$ for the values of N in the table are around 0.05 to 0.15. So by comparison the second term is completely negligible.

N	50	100	500	1000
Min	3.4E-09	5.5E-11	2.0E-11	2.5E-13
Median	0.00016	0.00011	1.0E-05	3.1E-06
Mean	0.00038	0.00022	2.3E-05	7.1E-06
Stdev	0.00055	0.00031	3.5E-05	1.0E-05
Max	0.00540	0.00278	0.00031	7.7E-05

Thus we can re-write (A-5) as

$$\mathrm{IR} = E\left(\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t}\right) \tag{A-9}$$

and using (A-6) and Lemma 2 gives

$$\mathrm{IR} = \mathrm{IC}\cdot E\left(\frac{1}{\sqrt{\sigma_\varepsilon^2/\mathbf{z}_{t-1}'\mathbf{z}_{t-1} + \sigma_{\mathrm{IC}}^2}}\right) = \mathrm{IC}\cdot E\left(\frac{1}{\sqrt{(1 - \mathrm{IC}^2 - \sigma_{\mathrm{IC}}^2)/\mathbf{z}_{t-1}'\mathbf{z}_{t-1} + \sigma_{\mathrm{IC}}^2}}\right). \tag{A-10}$$

Letting $\gamma = \sigma_{\mathrm{IC}}^2/\sigma_\varepsilon^2$ the above may be written in the form

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sigma_\varepsilon}\cdot E\left(\frac{1}{\sqrt{1/\mathbf{z}_{t-1}'\mathbf{z}_{t-1} + \gamma}}\right). \tag{A-11}$$

Under our assumption (A2) the random variable $\mathbf{z}_{t-1}'\mathbf{z}_{t-1} = \sum_{i=1}^{N} z_{i,t-1}^2$ has expected value N and $\mathrm{var}(\mathbf{z}_{t-1}'\mathbf{z}_{t-1}) = a\cdot N$, where $a \triangleq m_4 - 1$ with m_4 the fourth moment of $z_{i,t-1}$. For normally distributed exposures $a = 2$. Thus we can express $\mathbf{z}_{t-1}'\mathbf{z}_{t-1}$ as

$$\mathbf{z}_{t-1}'\mathbf{z}_{t-1} = N + \xi_N \tag{A-12}$$

where $E(\xi_N) = 0$ and $\mathrm{var}(\xi_N) = E(\xi_N^2) = a\cdot N$. The expression inside the expectation operator in (A-11) may be written as

$$g(\xi_N) = ((N + \xi_N)^{-1} + \gamma)^{-1/2} \tag{A-13}$$

and a three-term Taylor expansion at $E(\xi_N) = 0$ ignoring the remainder term gives

$$g(\xi_N) = (1/N + \gamma)^{-1/2} + \frac{1}{2}(1/N + \gamma)^{-3/2}\frac{\xi_N}{N^2} + (1/N + \gamma)^{-3/2}\left(\frac{3}{4N}(1/N + \gamma)^{-1} - 1\right)\frac{\xi_N^2}{2N^3}.$$

Thus

$$E(g(\xi_N)) = (1/N + \gamma)^{-1/2} + (1/N + \gamma)^{-3/2}\left(\frac{3}{4N}(1/N + \gamma)^{-1} - 1\right)\frac{a}{2N^2}. \tag{A-14}$$

The second term is of order $1/N^2$ and even for a very small actively managed portfolio with $N = 50$ assets we have $1/N^2 = 0.0004$, and for portfolios with more assets the second term will contribute even less. Noting that the missing remainder term will be of smaller order than $1/N^2$ indicates that it is quite reasonable to drop the second term in (A-14) and use the following in place of (A-10) and hence as the expression (30) in Proposition 1:

$$\mathrm{IR} = \frac{\mathrm{IC}}{\sqrt{\sigma_\epsilon^2/N + \sigma_{\mathrm{IC}}^2}} = \frac{\mathrm{IC}}{\sqrt{(1 - \mathrm{IC}^2 - \sigma_{\mathrm{IC}}^2)/N + \sigma_{\mathrm{IC}}^2}}. \tag{A-15}$$

The relative error in our approximation is

$$RE = \frac{1}{(1/N + \gamma)}\cdot\left(\frac{3}{4N}\cdot\frac{1}{(1/N + \gamma)} - 1\right)\cdot\frac{a}{2N^2}.$$

Since $\frac{3}{4N}(1/N + \gamma)^{-1} - 1 < 0$ and a is positive, the relative error is negative. Thus the Proposition 1 result (30) always over estimates the IR.

Evaluation of the above relative error expression for typical values of IC and σ_{IC} such as those occurring in Table 5 of Section 5 shows that the relative errors are sufficiently small to justify our result (30). For example with IC = .03, $\sigma_{\mathrm{IC}} = .1$ and $a = 2$ one obtains the following values of RE for $N = 50,\ 100,\ 500,\ 1000$:

N	50	100	500	1000
Relative error	−0.00667	−0.00312	−0.00029	−0.00008

So for the case of 50 assets, which is a small portfolio, our fundamental law formula (30) gives a relative error of less than 0.67% and for portfolios with at least 100 assets the relative error is less than 0.32%.

To check the above error sizes against reality we used 1000 replicates of the N-dimensional random vector $\mathbf{z}_{t-1}$ with standard normal components to simulate the expected value of

$$E\left(\frac{\mathrm{IC}}{\sqrt{(1 - \mathrm{IC}^2 - \sigma_{\mathrm{IC}}^2)/\mathbf{z}'_{t-1}\mathbf{z}_{t-1} + \sigma_{\mathrm{IC}}^2}}\right).$$

The results are compared with our formula (30) for $N = 50$, 100, 500, 1000 in the table below.

N	50	100	500	1000
Simulated Exact IR Formula	0.172306	0.212172	0.274024	0.286144
Our Fundamental Law IR	0.173838	0.212712	0.274110	0.286181
Relative Error	−0.00881	−0.00254	−0.00031	−0.00012

The values in the above table show that our Fundamental Law IR formula works quite well down to sample size 50 where the formula values differ from the simulation values only in the third digit, with differences at most in the fourth digit for sample sizes 100 and larger.

Appendix B

For the multi-factor model we proceed similarly as in Appendix A, noting that now the conditional mean predictor is

$$\boldsymbol{\alpha}_t = \boldsymbol{\Lambda}_t^{1/2}\mathbf{Z}_{t-1}\mathbf{IC} \tag{B-1}$$

and the conditional forecasting error covariance matrix is the positive definite matrix

$$\mathbf{\Omega}_t = \mathbf{\Lambda}_t^{1/2}(\mathbf{Z}_{t-1}\mathbf{\Sigma}_{\mathbf{IC}}\mathbf{Z}'_{t-1} + \mathbf{\Sigma}_\varepsilon)\mathbf{\Lambda}_t^{1/2}. \tag{B-2}$$

In this case the more general Woodbury formula[11] may be used to obtain

$$\mathbf{\Omega}_t^{-1} = \mathbf{\Lambda}_t^{-1/2}(\mathbf{\Sigma}_\varepsilon^{-1} - \mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1})\mathbf{\Lambda}_t^{-1/2}. \tag{B-3}$$

Using the conditional mean forecast formula we have

$$\begin{aligned}
\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\alpha}_t &= \mathbf{IC}'\mathbf{Z}'_{t-1}(\mathbf{\Sigma}_\varepsilon^{-1} - \mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1})\mathbf{Z}_{t-1}\mathbf{IC}\\
&= \mathbf{IC}'\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{I} - (\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})\mathbf{IC}\\
&= \mathbf{IC}'\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{\Sigma}_{\mathbf{IC}}^{-1}\mathbf{IC}\\
&= \mathbf{IC}'(\mathbf{\Sigma}_{\mathbf{IC}}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})(\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1})^{-1}\mathbf{IC}\\
&= \mathbf{IC}'((\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1} + \mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}\\
&= \mathbf{IC}'((\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}/N)^{-1}/N + \mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}\\
&= \mathbf{IC}'(\sigma_\varepsilon^2\cdot\mathbf{A}_N^{-1}/N + \mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}
\end{aligned} \tag{B-4}$$

where $\mathbf{A}_N \triangleq \frac{1}{N}\mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}$ with elements $\mathbf{A}_{N,kl} = \frac{1}{N}\sum_{i=1}^N z_{ik,t-1}z_{il,t-1}$, $k, l = 1, \cdots, K$. Since the exposures $z_{ik,t-1}$, $z_{il,t-1}$ have mean zero and variance one and are uncorrelated for $k \neq l$ it follows from the standard law of large numbers that $\mathbf{A}_N \xrightarrow{P} \mathbf{I}$ and correspondingly $\mathbf{A}_N^{-1} \xrightarrow{P} \mathbf{I}$. Based on the single factor model analysis in Appendix A it is reasonable to replace $\mathbf{A}_N^{-1}$ with $\mathbf{I}$ in the last line of (B-4). This results in

$$\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\alpha}_t = \mathbf{IC}'(\ \mathbf{I}\sigma_\varepsilon^2/\ N\ + \mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC} \tag{B-5}$$

as a finite N approximation for $\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\alpha}_t$. For an infinite number of assets N we have

$$\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\alpha}_t = \mathbf{IC}'\mathbf{\Sigma}_{\mathbf{IC}}^{-1}\mathbf{IC}. \tag{B-6}$$

Now for the term $\boldsymbol{\alpha}'_t\mathbf{\Omega}_t^{-1}\mathbf{1}$ which may be written

$$\begin{aligned}
\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\mathbf{1} &= \mathbf{IC}'\mathbf{Z}'_{t-1}(\mathbf{\Sigma}_\varepsilon^{-1} - \mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1})\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \mathbf{IC}'(\mathbf{I} - \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1})\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \mathbf{IC}'\mathbf{\Sigma}_{\mathbf{IC}}^{-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \mathbf{IC}'((\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})\mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \mathbf{IC}'(\mathbf{I}\sigma_\varepsilon^2 + \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}\mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \mathbf{IC}'\ (\mathbf{I}\sigma_\varepsilon^2 + N\cdot\mathbf{A}_N\mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{b}_N\cdot N\\
&= \mathbf{IC}'(\mathbf{I}\sigma_\varepsilon^2/N + \mathbf{A}_N\mathbf{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{b}_N
\end{aligned} \tag{B-7}$$

where $\mathbf{A}_N$ is as before and the vector $\mathbf{b}_N$ has elements

$$b_{N,k} = \frac{1}{N}\sum_{i=1}^N (z_{ik,t-1}/\sigma_{r_{it}})\ ,\ k = 1, \cdots, K. \tag{B-8}$$

By a law of large numbers for uncorrelated random variables with differing variances, each $b_{N,k}$ converges to zero in probability, and we already know that $\mathbf{A}_N$ converges to the identity matrix in probability. Thus $\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\mathbf{1}$ converges to zero in probability and $(\boldsymbol{\alpha}_t'\mathbf{\Omega}_t^{-1}\mathbf{1})^2$ converges to zero in probability even more rapidly. Since $\mathbf{\Omega}_t$ is positive definite it follows that $\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1}$ is positive for every finite N. So we just have to show that as $N \to \infty$, $\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1}$ is bounded away from zero. We have

$$\begin{aligned}
\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1} &= \mathbf{1}'\mathbf{\Lambda}_t^{-1/2}(\mathbf{\Sigma}_\varepsilon^{-1} - \mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\mathbf{\Sigma}_\varepsilon^{-1})\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \sigma_\varepsilon^{-2}\mathbf{1}'\mathbf{\Lambda}_t^{-1/2}(\mathbf{I} - \mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \sigma_\varepsilon^{-2}\mathbf{Z}'_{t-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}'_{t-1}\sigma_\varepsilon^{-2})\mathbf{\Lambda}_t^{-1/2}\mathbf{1}\\
&= \sigma_\varepsilon^{-2}\mathbf{1}'\mathbf{\Lambda}_t^{-1/2}(\mathbf{I} - \mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \sigma_\varepsilon^{-2}\cdot N\cdot\mathbf{A}_N)^{-1}\mathbf{Z}'_{t-1}\sigma_\varepsilon^{-2})\mathbf{\Lambda}_t^{-1/2}\mathbf{1}
\end{aligned} \tag{B-9}$$

which is positive for every finite N. Since $\mathbf{A}_N$ converges to the identity matrix in probability it follows that

$$\mathbf{Z}_{t-1}(\mathbf{\Sigma}_{\mathbf{IC}}^{-1} + \sigma_\varepsilon^{-2}\cdot N\cdot\mathbf{A}_N)^{-1}\mathbf{Z}'_{t-1} \xrightarrow{P} \mathbf{0} \tag{B-10}$$

and so

$$\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1} \xrightarrow{P} \sigma_\varepsilon^{-2}\mathbf{1}'\mathbf{\Lambda}_t^{-1}\mathbf{1}. \tag{B-11}$$

[11] The formula is $(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})\mathbf{VA}^{-1}$ where $\mathbf{A}$, $\mathbf{U}$, $\mathbf{C}$, $\mathbf{V}$ are of dimensions $n \times n$, $n \times k$, $k \times k$ and $k \times n$ respectively. This result is due to Woodbury (1950), “Inverting modified matrices”, Memorandum Rept. 42, Statistical Research Group, Princeton University, Princeton, NJ. The formula reduces to the Sherman-Morrison formula when $\mathbf{C}$ is a scalar and $\mathbf{U}$ and $\mathbf{V}$ have dimensions $n \times 1$, $1 \times n$ respectively.

But

$$\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1} = \sum_{i=1}^{N} \sigma_{r_{it}}^{-2} \geq N/b_{up}^2 > 0 \tag{B-12}$$

and so $\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1}$ is positive and grows linearly with N. Since $(\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1})^2$ converges to zero in probability more rapidly than $\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1}$ and $\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1}$ has a lower bound that grows linearly with N, the term $(\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1})^2/(\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1})$ in (17) goes to zero so rapidly as to be very negligible relative to $\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t$ (see empirical results for the single factor case in Appendix A). Thus we can drop $(\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{1})^2/(\mathbf{1}'\boldsymbol{\Lambda}_t^{-1}\mathbf{1})$ in (17) and using (B-5) gives the finite number of assets N multi-factor result

$$\mathrm{IR} = \sqrt{\mathbf{IC}'(\,\mathbf{I}\sigma_\varepsilon^2/N \;+ \boldsymbol{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}} \tag{B-13}$$

where

$$\sigma_\varepsilon^2 = 1 - \sum_{k=1}^{K} (\mathrm{IC}_k^2 + \sigma_{\mathrm{IC},k}^2). \tag{B-14}$$

The asymptotic information ratio is

$$\mathrm{IR} = \sqrt{\mathbf{IC}'\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1}\mathbf{IC}}\,. \tag{B-15}$$

For a one factor model, the above result simplifies to the one factor fundamental law in (30).

Finally, for the multi-factor case we derive the explicit form of the optimal active portfolio weights whose general formula is Eq. (14). By the same argument as above we can drop the term involving κ in (14) and obtain the following

$$\Delta\mathbf{w}_t = \sigma_{A,t}\frac{\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t}{\sqrt{\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t}}. \tag{B-16}$$

The numerator of the above expression is

$$\begin{aligned}
\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t &= \boldsymbol{\Lambda}_t^{-1/2}(\boldsymbol{\Sigma}_\varepsilon^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1} + \mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1})\mathbf{Z}_{t-1}\mathbf{IC}\\
&= \boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{I} - (\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1} \;+ \mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})\mathbf{IC}\\
&= \boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1} \;+ \mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1}\mathbf{IC}\\
&= \boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1}(\mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1}((\mathbf{Z}_{t-1}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Z}_{t-1})^{-1} + \boldsymbol{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}\\
&= \boldsymbol{\Lambda}_t^{-1/2}\mathbf{Z}_{t-1}/N(\sigma_\varepsilon^2/\,N\;\mathbf{I} + \boldsymbol{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}.
\end{aligned} \tag{B-17}$$

Recalling the expression (B-5) for $\boldsymbol{\alpha}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\alpha}_t$ we see that

$$\Delta\mathbf{w}_t = \frac{\sigma_{A,t}}{N}\frac{\boldsymbol{\Lambda}_t^{-1/2}\mathbf{Z}_{t-1}(\sigma_\varepsilon^2/\,N\;\mathbf{I} + \boldsymbol{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}}{\sqrt{\mathbf{IC}'(\sigma_\varepsilon^2/\,N\;\mathbf{I} + \boldsymbol{\Sigma}_{\mathbf{IC}})^{-1}\mathbf{IC}}} \approx \frac{\sigma_{A,t}}{N}\frac{\boldsymbol{\Lambda}_t^{-1/2}\mathbf{Z}_{t-1}\,\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1}\mathbf{IC}}{\sqrt{\mathbf{IC}'\boldsymbol{\Sigma}_{\mathbf{IC}}^{-1}\mathbf{IC}}}. \tag{B-18}$$

For a single factor model the optimal weight in (B-18) reduces to:

$$\Delta w_{it} = \frac{\sigma_{A,t}}{\sqrt{\sigma_\varepsilon^2/N + \sigma_{\mathrm{IC}}^2}}\frac{z_{i,t-1}}{N\sigma_{r_{it}}} \approx \frac{\sigma_{A,t}}{\sigma_{\mathrm{IC}}}\frac{z_{i,t-1}}{N\sigma_{r_{it}}}\;. \tag{B-19}$$

The optimal weight is positively related to the target portfolio tracking error $\sigma_{A,t}$, and inversely related to: (a) the size of the selection universe N, (b) the residual return volatility $\sigma_{r_{it}}$, and (c) the "strategy risk"σ_{IC}. When the correlation between the residual returns and the signal is noisier the "strategy risk" is higher (σ_{IC} large) and the optimal active weight will be smaller and vice versa. The signal IC volatility acts as a "valve" to control the optimal leverage of the portfolio. In reality, most portfolios built using a risk model that does not incorporate "strategy risk" as we do in this paper will have the problem of over-leveraging, and if no other constraints are imposed the portfolio will be wiped out sooner or later.

References

Bloch, M., Guerard, J., Markowitz, H., Todd, P., Xu, G., 1993. A comparison of some aspects of the U.S. and Japanese equity markets. Jpn. World Econ. 5, 3–26.
Buckle, David, 2004. How to calculate breadth: an evolution of the fundamental law of active portfolio management. J. Asset Manag. 4, 393–405.
Clarke, Roger, Silva, Harindra de, Thorley, Steven, 2002. Portfolio constraints and the fundamental law of active management. Financ. Anal. J. 58 (5), 48–66.
Clarke, Roger, Silva, Harindra de, Thorley, Steven, 2006. The fundamental law of active management. J. Invest. Manag. 4 (3), 54–72.
Connor, Gregory, Korajczyk, Robert A., 2010. Factor models of asset returns. In: Cont, Rama (Ed.), Encyclopedia of Quantitative Finance. Wiley, Chicester.
Fama, Eugene F., MacBeth, James D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81 (3), 607–636.
Fama, Eugene F., French, R. Kenneth, 1992. The cross-section of expected stock returns. J. Financ. 47 (2), 427–465.
Guerard, John B., Markowitz, Harry, Xu, Ganlin, 2013. Global stock selection modeling and efficient portfolio construction and management. J. Invest., 1–8.
Grinold, Richard C., 1989. The fundamental law of active management. J. Portf. Manag. 15 (3), 30–38.
Grinold, Richard C., 1994. Alpha is volatility times IC times score. J. Portf. Manag. 20 (4), 9–16.
Grinold, Richard C., 2007. Dynamic portfolio analysis. J. Portf. Manag. 34 (1), 12–26.
Grinold, Richard C., Kahn, Ronald N., 2000. Active Portfolio Management 2nd ed.. McGraw-Hill, New York.

Goodwin, Thomas H., 1998. The information ratio. Financ. Anal. J. 54 (4), 34–43.
Huberman, Guy, Wang, Zhenyu, 2008. Arbitrage pricing theory. In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics2nd ed.. Palgrave Macmillan, Basingstoke.
Lee, Jyh-Huei, Stefek, Dan, 2008. Do risk factors eat alphas. J. Portf. Manag. 34 (4), 12–25.
Levy, Haim, Markowitz, Harry M., 1979. Approximating expected utility by a function of mean and variance. Am. Econ. Rev. 69, 308–317.
Magnus, Jan R., Neudecker, Heinz, 2002. Matrix Differential Calculus with Applications in Statistics and Econometrics Revised Edition. John Wiley & Sons, New York.
Markowitz, Harry M., 1952. Portfolio selection. J. Financ. 7 (1), 77–91.
Menchero, Jose, Mitra, Indrajit, 2008. The structure of hybrid factor models. J. Invest. Manag. 6 (3), 35–47.
Qian, Edward, Hua, Ronald, 2004. Active risk and information ratio. J. Invest. Manag. 2 (3), 20–34.
Qian, E., Hua, R., Sorensen, E.H., 2007. Quantitative Equity Portfolio Management: modern Techniques and Applications. CRC Press, London.
Rosenberg, Barr, 1974. Extra-market components of covariance in security returns. J. Financ. Quant. Anal. 9, 263–274.
Ross, S., 1976. The arbitrage theory of capital asset pricing. J. Econ. Theory 13, 341–360.
Treynor, Jack L., Black, Fischer, 1973. How to use security analysis to improve portfolio selection. J. Bus. 46 (1), 66–86.
Ye, Jia, 2008. How variation in signal quality affects performance. Financ. Anal. J. 64 (4), 48–61.
Zangari, Peter, 2003. Equity risk factor models. In: Letterman, Robert (Ed.), Modern Investment Management: An Equilibrium Approach. Wiley.