CHAPTER 7

Multifactor Alpha Models

In Chapter 4 (see also Qian & Hua 2004), we presented an analytic framework to evaluate individual alpha factors based on the risk-adjusted information coefficient (IC). The ratio of average IC to the standard deviation of IC serves as a proxy for the information ratio (IR) of active strategies that employ the alpha factors. We then devoted the next two chapters to the examination of several alpha factors on an individual basis. In practice, alpha models almost always employ multiple factors instead of a single one. So then, the question naturally arises: how to blend these factors optimally into a composite alpha model? The combination of these factors is not restricted to quantitative factors. For instance, some investment firms conduct both fundamental and quantitative researches. How to combine them into a single forecasting process, in terms of ranking or scores, presents a similar challenge.

In this chapter, we extend the analytic framework to derive factor weights in a multifactor alpha model. Our objective is to maximize the IR of the multifactor model. The approach is similar to a mean–variance optimization. The difference is that we now replace a portfolio of stocks with a portfolio of factors. Thus, average IC and standard deviation of IC resemble the expected return and risk of dollar neutral, risk-neutral factor portfolios. In addition, correlations between ICs of different factor portfolios also play an essential role in delivering the diversification benefits. It is important to note that the correlation between ICs is not the same as the correlation between factor scores. The former is the correlation of returns to factor portfolios across time, whereas the latter is the cross-sectional correlation of factor scores at a given time. We will show that the correlations among ICs play a crucial role in determining the optimal alpha model weights, whereas correlations among factor scores play a secondary role. Theoretically, it is tempting to assume that the two are identical, but empirical evidence seems to prove the contrary.

This chapter consists of four sections. In the first section, we derive the analytical expression of the composite IC of a multifactor alpha model for a single period. We define a *multifactor model* as one that linearly combines scores of individual alpha factors to create a composite forecast (i.e., a composite score), and a composite IC is the IC of the composite score. The efficacy (or the expected performance) of a multifactor alpha model becomes the IR of its single-period ICs through time. A similar approach is illustrated in Chapter 4. In the second section, the analytical expression of a composite IR is derived with the assumption that cross-sectional factor-score correlations do not change over time. This time invariant assumption makes analytical derivations tractable, so we can solve for the optimal model weighting that achieves the highest IR of the composite forecast. In the third section, we discuss the important difference between cross-sectional factor score correlation and time-series IC correlation in the context of multifactor model building. We also suggest a practical procedure to deal with the time variability of factor-score correlations. In the last section, we examine the statistical linkage between our model optimization framework and the Fama–MacBeth regression procedure. Specifically, we provide cautionary notes to practitioners who would like to apply a Fama–MacBeth-like regression framework to derive optimal model weights.

7.1 SINGLE-PERIOD COMPOSITE IC OF A MULTIFACTOR MODEL

As in Chapter 4, we will first consider a single-period excess return of a multifactor model, which is a linear combination of M factors $\left(\mathbf{F}_1,\mathbf{F}_2,\ldots,\mathbf{F}_M\right)$ with the weight vector $\mathbf{v}=\left(v_1,v_2,\cdots,v_M\right)'$. The weight vector, once selected, shall remain constant over time. To put it differently, we are solving for the optimal weighting of a constant linear multifactor model. There are more complex alpha models that could be nonlinear and/or dynamic. We shall cover them in later chapters.

To link model performance to realistic portfolio implementation, we assume all factors are risk-adjusted according to the analytical framework illustrated in Chapter 4. Therefore, the composite risk-adjusted factor is a linear combination:

$$\mathbf{F}_c = \sum_{i=1}^{M} v_i \mathbf{F}_i \,. \tag{7.1}$$

The composite will also be risk-adjusted in the sense that the associated active portfolio will be neutral to all risk factors and is mean–variance optimal. Now we treat the composite factor $\mathbf{F}_c$ as a single factor and use the analytic framework presented in Chapter 4.

Recall from Chapter 4 that the single-period excess return of an alpha factor is expressed as a function of the covariance between the factor and the risk-adjusted return. To clarify the notation, $\mathbf{F}_{c,t}$ represents the risk-adjusted composite factor available at the beginning of period t, whereas $\mathbf{R}_t$ is the risk-adjusted return during period t.

$$\begin{aligned}\alpha_t &= \frac{(N-1)}{\lambda_t}\operatorname{cov}\left(\mathbf{F}_{c,t},\mathbf{R}_t\right)\\ &= \frac{(N-1)}{\lambda_t}\operatorname{corr}\left(\mathbf{F}_{c,t},\mathbf{R}_t\right)\operatorname{dis}\left(\mathbf{F}_{c,t}\right)\operatorname{dis}(\mathbf{R}_t)\end{aligned} \tag{7.2}$$

The covariance between the composite factor and the risk-adjusted return is a linear combination of covariances between individual factors and the risk-adjusted return:

$$\begin{aligned}\operatorname{cov}\left(\mathbf{F}_{c,t},\mathbf{R}_t\right) &= \operatorname{cov}\left(\sum_{i=1}^{M} v_i \mathbf{F}_{i,t},\mathbf{R}_t\right) = \sum_{i=1}^{M} v_i \operatorname{cov}\left(\mathbf{F}_{i,t},\mathbf{R}_t\right)\\ &= \left[\sum_{i=1}^{M} v_i IC_{i,t}\operatorname{dis}\left(\mathbf{F}_{i,t}\right)\right]\operatorname{dis}(\mathbf{R}_t)\end{aligned} \tag{7.3}$$

In the second line of the preceding equation, we have expressed the covariances in terms of ICs and dispersions. Also recall from Chapter 4 that the risk-aversion parameter is calibrated such that the active portfolio would have a targeted tracking error. The relationship in the case of a composite alpha factor is

$$\lambda_t = \frac{\sqrt{N-1}\,\mathrm{dis}\left(\mathbf{F}_{c,t}\right)}{\sigma_{\mathrm{model}}}. \tag{7.4}$$

The dispersion of the composite factor depends on the model weights and cross-sectional covariances among different factor scores. Denoting the cross-sectional covariance between two factors by $\phi_{ij,t} = \mathrm{cov}\left(\mathbf{F}_{i,t}, \mathbf{F}_{j,t}\right)$ and the factor covariance matrix by $\boldsymbol{\Phi}_t = \left(\phi_{ij,t}\right)_{i,j=1}^{M}$, the dispersion of the composite is given by

$$\mathrm{dis}\left(\mathbf{F}_{c,t}\right) = \sqrt{\mathbf{v}'\boldsymbol{\Phi}_t\mathbf{v}}\,. \tag{7.5}$$

Substituting Equation 7.5, Equation 7.4, and Equation 7.3 into Equation 7.2 yields

$$\alpha_t = IC_{c,t}\sqrt{N-1}\,\sigma_{\mathrm{model}}\,\mathrm{dis}(\mathbf{R}_t)\,. \tag{7.6}$$

Further,

$$IC_{c,t} = \mathrm{corr}\left(\mathbf{F}_{c,t}, \mathbf{R}_t\right) = \frac{\sum_{i=1}^{M} v_i IC_{i,t}\,\mathrm{dis}\left(\mathbf{F}_{i,t}\right)}{\sqrt{\mathbf{v}'\boldsymbol{\Phi}_t\mathbf{v}}}. \tag{7.7}$$

Equation 7.6 provides the excess return of a multifactor alpha model. It is essentially of the same form as in the single-factor case, except that the IC is that of a composite factor given in (7.7) instead of a single one. The composite IC is a linear combination of individual factor ICs, and the weights are factor weight v_i times the ratio of individual factor dispersion to composite factor dispersion. Among the four terms in (7.6), the number of stocks, the target tracking error, and the dispersion of risk-adjusted returns have either little or no time-series variation, so we shall assume that they are constant throughout the remainder of the chapter. The composite IC, on the other hand, has many time-varying components, including the ICs of the underlying alpha factors $IC_{i,t}$, their cross-sectional dispersions $\mathrm{dis}\left(\mathbf{F}_{i,t}\right)$, and their covariance matrix $\boldsymbol{\Phi}_t$.

Example 7.1

Suppose we have two factors $\mathbf{F}_1$ and $\mathbf{F}_2$. In a given period, we have $\text{dis}(\mathbf{F}_1)=1$ and $\text{dis}(\mathbf{F}_2)=0.5$, and the factor correlation is 0.5. Then the factor covariance matrix is

$$\mathbf{\Phi}=\begin{pmatrix} 1 & 0.5\cdot 1\cdot 0.5 \\ 0.5\cdot 1\cdot 0.5 & 0.5^2 \end{pmatrix}=\begin{pmatrix} 1 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}.$$

Suppose we equally weight these two factors; the dispersion of the composite factor is

$$\text{dis}(\mathbf{F}_c)=\sqrt{\mathbf{v}'\mathbf{\Phi}\mathbf{v}}=\left[\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}\begin{pmatrix} 1 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\right]^{1/2}$$

$$=\sqrt{0.5^2+0.25\cdot 0.5^2+2\cdot 0.25\cdot 0.5^2}=0.66$$

Example 7.2

Suppose that, in the given period, the ICs of factor 1 and factor 2 are 0.15 and 0.20, respectively. Then the IC of the composite factor is

$$IC_c=\frac{\sum_{i=1}^{M} v_i IC_{i,t}\,\text{dis}(\mathbf{F}_{i,t})}{\sqrt{\mathbf{v}'\mathbf{\Phi}_t\mathbf{v}}}=\frac{0.5\cdot 0.15\cdot 1+0.5\cdot 0.20\cdot 0.5}{0.66}=0.11+0.08=0.19\,.$$

In this case, the composite IC is greater than the IC of factor 1 but less than that of factor 2.

The previous examples illustrate the relationship between the composite IC and individual ICs for a single period. The major purpose of optimal alpha modeling is to maximize the IR over multiple periods, which depends not only on the average IC but also on the standard deviation of IC. It seems highly unlikely that there exists a full analytic solution for the weight vector $\mathbf{v}$ that maximizes the IR based on (7.6) because $\mathbf{v}$ appears in a quadratic form in the denominator. There are several possible approaches to solving this problem. One involves analytical approximation, and another involves transformation of alpha factors into orthogonal factors. We shall start with analytical approximation by assuming the factor correlation to be constant through time. Factor orthogonalization and factor-score correlations are also discussed in the second half of this chapter.

7.2 OPTIMAL ALPHA MODEL: AN ANALYTICAL DERIVATION

In this section, we derive an analytical expression of the optimal model weighting that achieves the highest information ratio, under the assumption that the factor covariance matrix stays unchanged over time. We first explore how factor standardization affects the IC of a composite factor. Then, the analytical expression of IR is derived for a composite multifactor alpha model, linking the composite IR to the time-series of ICs of each individual alpha factor. Based on this expression of composite IR, we solve analytically for the optimal model weighting that achieves the highest composite IR. In this derivation, we assume that model weighting is also time invariant. Lastly, we provide a brief discussion of why maximizing the single-period IC of a composite model does not achieve optimality.

7.2.1 Factor Standardization

If we assume that the factor covariance matrix is time invariant, the composite IC becomes a constant linear combination of model weights and individual ICs. To simplify things further, we standardize all individual factors such that their dispersion is always unity over time, i.e., $\text{dis}\left(\mathbf{F}_{i,t}\right)=1$, for all i,t. It is common to standardize all factors in practice, and there are several potential benefits for doing so. First, it "equalizes" the contribution of individual factors to the overall model for a given set of model weights. Second, it immunizes the composite model from changes in the dispersions of the factors, thus reducing portfolio turnovers associated with such changes. More importantly, there is little direct empirical evidence indicating that such turnover adds value. Note the following:

- Standardizing individual factors before combining them into an alpha model amounts to rescaling the model weights putting factors in the same units for comparison. Moreover, as the dispersions of factors change over time, the rescaling weights are also time varying. In other words, standardizing factors actually leads to implicit time-varying alpha models.

Example 7.3

We will standardize factor 2 in Example 7.1, whose original dispersion for the given period is 0.5, by multiplying it by 2. The first factor is already standardized. Suppose we still equally weight the two standardized factors; the effective weights on the original factors are 1/3 and 2/3. Suppose also that during the next period, the dispersion of factor 1 changes to 0.5, whereas the dispersion of factor 2 changes to 1. We would standardize the factor 1 by doubling it while leaving factor 2 untouched. In this period, an equally weighted model of the standardized factor would imply an effective weight of 2/3 and 1/3 on the original factors.

With factor standardization, the composite IC for time t is

$$IC_{c,t} = \frac{1}{\sqrt{\mathbf{v}'\mathbf{\Phi}\mathbf{v}}} \sum_{i=1}^{M} v_i IC_{i,t} = \frac{1}{\tau} \sum_{i=1}^{M} v_i IC_{i,t} \,. \tag{7.8}$$

The covariance matrix $\mathbf{\Phi}$ reduces to the correlation matrix of factors because all factors are standardized. The composite the IC can be seen as a linear combination of the ICs of the underlying factors scaled by a constant τ, which is the dispersion of the composite factor (7.5). Another important feature of Equation 7.8 is that the composite IC remains unchanged if the factor weights are all scaled by the same constant.

7.2.2 IR of the Composite IC

We now calculate the expected IC and the standard deviation of IC to obtain the IR. We start with a two-factor example.

Example 7.4

If there are two factors, then we have

$$IC_{c,t} = \frac{1}{\sqrt{v_1^2 + v_2^2 + 2v_1v_2\rho_{12}}} \left(v_1 IC_{1,t} + v_2 IC_{2,t} \right) = \frac{1}{\tau} \left(v_1 IC_{1,t} + v_2 IC_{2,t} \right). \tag{7.9}$$

The correlation between the two factors is ρ_{12}, which, for the moment, is assumed to be constant over time. The expected composite IC is a linear combination of individual ICs is

$$\overline{IC}_c = \frac{1}{\tau} \left(v_1 \overline{IC}_1 + v_2 \overline{IC}_2 \right), \tag{7.10}$$

and the standard deviation of the IC is

$$\begin{aligned} \mathrm{std}\left(IC_c\right) &= \frac{1}{\tau} \mathrm{std}\left(v_1 IC_{1,t} + v_2 IC_{2,t} \right) \\ &= \frac{1}{\tau} \sqrt{v_1^2 \sigma_{IC_1}^2 + v_2^2 \sigma_{IC_2}^2 + 2 v_1 v_2 \rho_{12,IC} \sigma_{IC_1} \sigma_{IC_2}} \end{aligned} \tag{7.11}$$

The IC correlation between the two factors is denoted by $\rho_{12,IC}$, and the standard deviations of ICs are σ_{IC_1} and σ_{IC_2}. The IR, in this case the ratio of average IC to the standard deviation of IC, is

$$IR_c = \frac{\left(v_1\overline{IC}_1 + v_2\overline{IC}_2\right)}{\sqrt{v_1^2\sigma_{IC_1}^2 + v_2^2\sigma_{IC_2}^2 + 2v_1v_2\rho_{12,IC}\sigma_{IC_1}\sigma_{IC_2}}}. \tag{7.12}$$

For a general model with M factors, we can denote the average IC by a vector $\overline{\mathbf{IC}} = \left(\overline{IC}_1, \overline{IC}_2, \cdots, \overline{IC}_M\right)'$, and the IC covariances by matrix $\mathbf{\Sigma}_{IC} = \left(\rho_{ij,IC}\right)_{i,j=1}^{M}$. Then the average and standard deviation of a composite IC are

$$\begin{aligned} \overline{IC}_c &= \frac{1}{\tau}\sum_{i=1}^{M} v_i\overline{IC}_i = \frac{1}{\tau}\mathbf{v}'\cdot\overline{\mathbf{IC}} \\ \operatorname{std}\left(IC_c\right) &= \frac{1}{\tau}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} v_iv_j\rho_{ij,IC}\sigma_{IC_i}\sigma_{IC_i}} = \frac{1}{\tau}\sqrt{\mathbf{v}'\cdot\mathbf{\Sigma}_{IC}\cdot\mathbf{v}} \end{aligned} \tag{7.13}$$

and the IR is

$$IR_c = \frac{\sum_{i=1}^{M} v_i\overline{IC}_i}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} v_iv_j\rho_{ij,IC}\sigma_{IC_i}\sigma_{IC_i}}} = \frac{\mathbf{v}'\cdot\overline{\mathbf{IC}}}{\sqrt{\mathbf{v}'\cdot\mathbf{\Sigma}_{IC}\cdot\mathbf{v}}}. \tag{7.14}$$

- The scale constant τ — the dispersion of the composite factor, which depends on cross-sectional factor-score correlations — has completely dropped out of the IR equation. However, the time-series IC correlations remain, and the IC correlation matrix determines the standard deviation of composite IC over time, and thus its active risk.

7.2.3 Optimal Model Weights

We can now find the optimal model weights that maximize the IR (7.14) of the composite alpha factor. We note that IR in (7.14) assumes that the cross-sectional factor-score correlation matrix is a constant through time. As we can see, although the IR optimization problem is similar to mean–variance optimization, there are important differences. The objective function is the mean/standard deviation ratio, and there is no risk-aversion parameter. As a result, any constant multiple of optimal weights will also be optimal because they give rise to the same IR. In theory, there is no need for the weight to sum up to 100%. However, in practice, we often do so customarily.

This is an unconstrained optimization. Taking the partial derivative of (7.14) with respect to the weights yields

$$\frac{\partial\left(IR_c\right)}{\partial\mathbf{v}}=\frac{\overline{\mathbf{IC}}}{\sqrt{\mathbf{v}'\cdot\mathbf{\Sigma}_{IC}\cdot\mathbf{v}}}-\frac{\left(\mathbf{v}'\cdot\overline{\mathbf{IC}}\right)\mathbf{\Sigma}_{IC}\cdot\mathbf{v}}{\left(\mathbf{v}'\cdot\mathbf{\Sigma}_{IC}\cdot\mathbf{v}\right)^{3/2}}. \tag{7.15}$$

Equating the partial derivatives to zero, we have

$$\left(\mathbf{v}'\cdot\mathbf{\Sigma}_{IC}\cdot\mathbf{v}\right)\overline{\mathbf{IC}}=\left(\mathbf{v}'\cdot\overline{\mathbf{IC}}\right)\mathbf{\Sigma}_{IC}\cdot\mathbf{v}. \tag{7.16}$$

The solution for the optimal weights is

$$\mathbf{v}^*=s\mathbf{\Sigma}_{IC}^{-1}\overline{\mathbf{IC}}, \tag{7.17}$$

where s is an arbitrary, generally positive constant. We can select s such that the sum of its optimal weights is 1. Substituting the optimal weights into (7.14) gives the optimal IR:

$$IR^*=\sqrt{\overline{\mathbf{IC}}'\cdot\mathbf{\Sigma}_{IC}^{-1}\cdot\overline{\mathbf{IC}}}. \tag{7.18}$$

- The optimal weight (7.17) is akin to the mean–variance solution for the optimal portfolio of securities including cash. It is identical to the solution of optimal manager selections for investment consultants, where the "managers" in this case are alpha factors. This indicates that the weight of an alpha factor in the composite depends not only on its own risk/return trade-off but also on its IC correlation with other factors' ICs.
- The optimal weight $\mathbf{v}^*$ can also be derived from an OLS regression without an intercept term. Britten-Jones (1998) shows that mean–variance (MV) optimal weights in general can be obtained this way.

One of the benefits of this alternative approach is that we can obtain standard errors for the optimal weights. We leave the proof as an exercise (see Problem 7.4).

Example 7.5

We illustrate the optimal model weights in a two-factor case in which

$$\begin{aligned} v_1 &= \frac{s}{1-\rho_{12,IC}^2}\left(\frac{\overline{IC}_1}{\sigma_{IC_1}^2} - \frac{\rho_{12,IC}\overline{IC}_2}{\sigma_{IC_1}\sigma_{IC_2}}\right) \\ v_2 &= \frac{s}{1-\rho_{12,IC}^2}\left(\frac{\overline{IC}_2}{\sigma_{IC_2}^2} - \frac{\rho_{12,IC}\overline{IC}_1}{\sigma_{IC_1}\sigma_{IC_2}}\right) \end{aligned}. \tag{7.19}$$

Equation 7.19 states that the optimal weight of a factor is determined by two terms. The first term is the ratio of the average IC to the variance of IC. The second term, carrying a negative sign, is proportional to the IC correlation and the average IC of the other factor. Therefore, if a factor has high IC correlations with other factors, then its model weight will be negatively affected. On the other hand, if a factor has low and/or negative IC correlations with other factors, its model weight will be positively affected.

For a model with two factors, the optimal IR can also be explicitly written as

$$IR^* = \frac{\sqrt{IR_1^2 + IR_2^2 - 2\rho_{12,IC}IR_1IR_2}}{\sqrt{1-\rho_{12,IC}^2}}. \tag{7.20}$$

For two factors with given IRs, the optimal IR will be higher if their IC correlation is lower. Figure 7.1 plots the optimal IR as a function of IC correlation for given values of two individual IRs. The two IRs are 1.0 and 0.5, respectively. As the IC correlation changes from −0.5 to 0.5, the optimal IR declines from 1.5 to 1.0. When the IC correlation is at −0.5, there are strong diversification benefits between the two factors, and the combined optimal IR is much higher than both individual IRs. However, as the IC correlation increases, the diversification benefit shrinks. When it reaches 0.5 and above, the benefit disappears entirely unless one is willing to bet against one of the factors (see Problem 7.6), i.e., when the optimal weight becomes negative.

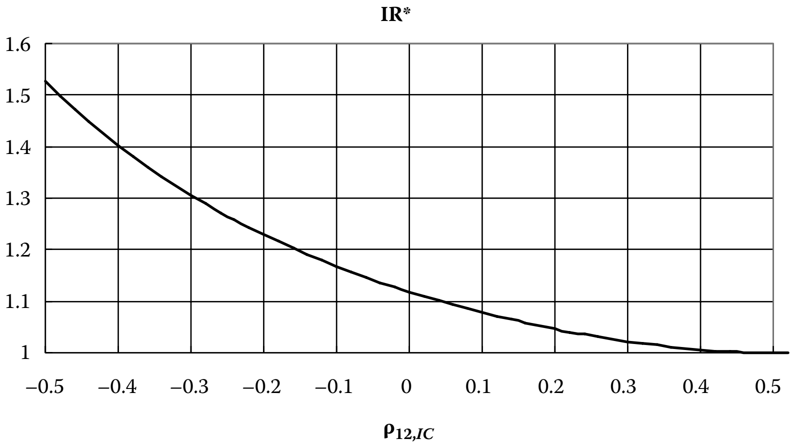

FIGURE 7.1. The optimal IR as a function of IC correlation between the two factors whose IRs are 1.0 and 0.5, respectively.

Although such a factor model is theoretically correct, in practice it is highly improbable to implement such a solution. This is so because, when the IC correlation is high and positive, the optimal model will try to arbitrage one factor against another, i.e., place positive weight on the factor with higher IR, and negative weight on the factor with lower IR. Thus, the outcome of such a model is extremely sensitive to the estimation accuracy of the IR difference. If the model happens to be wrong in this regard, it would put the wrong weights on the wrong factors.

7.2.4 An Empirical Example

To illustrate an empirical application of Equation 7.17, we select one factor from each factor category discussed in Chapter 5: cash flow from operation to enterprise value (CFO2EV) from the value category, external financing (XF) from the quality category, and the 9-month price momentum (Ret9) from the momentum category. For each factor, we calculated the risk-adjusted IC on a quarterly basis using the Russell 3000 as the stock universe. The time span of our data is from 1987 to 2004 — 72 quarters in total. We also compute the average IC and the standard deviation of IC for the three factors so that we can derive the optimal alpha model weights based on the three factors.

The average ICs and the standard deviation of ICs are listed in Table 7.1 together with the annualized IR. Because we use quarterly data, the annualized IR is simply twice the ratio of average IC to the standard deviation

TABLE 7.1 Average IC and Standard Deviation of IC for the Three Factors

	CFO2EV	XF	Ret9
Average IC	0.06	0.04	0.05
Standard deviation	0.05	0.04	0.09
Annualized IR	2.09	1.91	1.10

TABLE 7.2 Weights of Alpha Models and Corresponding IR

	IR	CFO2EV	XF	Ret9
$\mathbf{w}_1$	2.68	38%	50%	12%
$\mathbf{w}^*$	3.23	69%	–1%	32%

of IC. As we can see from this table, both the value factor CFO2EV and the quality factor XF have high IR mainly due to a low standard deviation of IC, i.e., the excess returns associated with these two factors tend to exhibit low volatility. On the other hand, the momentum factor has the same level of average IC as the other two, but its standard deviation is almost twice as high, resulting in lower IR for the factor.

With standard deviations of IC and the IC correlation matrix (in Table 7.4), we construct the IC covariance matrix and then derive the optimal alpha model that maximizes IR, using (7.17). The weights of the optimal model are shown as $\mathbf{w}^*$ in Table 7.2. In this case, we have 69% in CFO2EV and 32% in Ret7, but –1% in XF. The XF factor itself has an IR of 1.91, but because it is highly correlated with the factor CFO2EV, which has a higher IR and lower correlation with Ret9, the XF factor gets no weight in the optimal alpha model. To see the importance of IC correlation more directly, we also derive another set of weights with a diagonal IC covariance matrix by letting IC correlations be zero. This is shown as $\mathbf{w}_1$ in Table 7.2 and has 50, 38, and 12% in XF, CFO2EV, and Ret9, respectively. However, the IR of this model is only 2.68, whereas the maximum IR with $\mathbf{w}^*$ is 3.23.

7.2.5 Maximum Single-Period IC

We have found the optimal model weights $\mathbf{v}$ that maximize the multiperiod IR. One could also focus on model weights that maximize the single-period IC. The optimal weights for a single-period IC depend on the average ICs and the factor correlation matrix $\mathbf{\Phi}$.

From (7.8), we take the partial derivative with respect to $\mathbf{v}$ to obtain the optimality condition. Following steps similar to (7.16) and (7.17), we obtain

$$\mathbf{v} = s\mathbf{\Phi}_t^{-1}\overline{\mathbf{IC}}. \tag{7.21}$$

The solution is proportional to the inverse of the factor covariance (or correlation) matrix times the IC.

If the factor correlation matrix remains constant over time, (7.21) is also the solution that achieves *the maximum average IC* over multiple periods. However, the efficacy of an alpha model is not in the average IC but in the ratio of the average IC to the standard deviation of IC. The weights in (7.21) totally ignore the standard deviation of IC. Therefore, there is no guarantee that its IR would be high. A prime example of factors with high average IC but high standard deviation of IC is the 1-month price reversal factor. In addition, the 1-month reversal factor tends to have low factor correlation with other low-frequency factors. Hence, a model that maximizes the average IC would have significant weight in the 1-month price reversal factor. However, such a model is likely to have a low IR and, to make matters worse, extremely high turnover. We shall discuss the subject of portfolio turnover in detail in later chapters.

7.3 FACTOR CORRELATION VS. IC CORRELATION

The optimal model weights depend strongly on IC correlations but not on factor correlations. We have shown that, when we assume that the factor correlations stay constant over time, it completely drops out of the analysis as far as IR is concerned. Although it is important to distinguish between them, the two are in fact interrelated. In this section we analyze their relationship.

7.3.1 Relationship in a Single Period

We continue to use the two-factor case as an example. Suppose that, for a single period, the two standardized factors have a factor correlation $\phi_{12,t} = \mathrm{corr}\left(\mathbf{F}_{1,t}, \mathbf{F}_{2,t}\right)$. The ICs of the two factors for the period will be constrained by the factor correlation. Imagine the case where the factor correlation is unity; then we know that the two factors are essentially identical and the two ICs must be the same. On the other hand, if the factor correlation is –1, then the two ICs must be the opposite of each other. However, when the factor correlation falls somewhere between these two extreme cases, it leads to a much looser constraint on the two ICs.

For general cases, the two ICs — $IC_{1,t}$ and $IC_{2,t}$ — together with $\phi_{12,t}$ forms a 3×3 correlation matrix:

$$C=\begin{pmatrix} 1 & IC_{1,t} & IC_{2,t} \\ IC_{1,t} & 1 & \phi_{12,t} \\ IC_{2,t} & \phi_{12,t} & 1 \end{pmatrix}. \tag{7.22}$$

Because C has to be positive definite, its determinant must be nonnegative. We have

$$\begin{aligned}\det C &= \begin{vmatrix} 1 & \phi_{12,t} \\ \phi_{12,t} & 1 \end{vmatrix} - IC_{1,t}\begin{vmatrix} IC_{1,t} & \phi_{12,t} \\ IC_{2,t} & 1 \end{vmatrix} + IC_{2,t}\begin{vmatrix} IC_{1,t} & 1 \\ IC_{2,t} & \phi_{12,t} \end{vmatrix} \\ &= 1-\phi_{12,t}^2 - IC_{1,t}^2 - IC_{2,t}^2 + 2\phi_{12,t} IC_{1,t} IC_{2,t} \ge 0 \end{aligned} \tag{7.23}$$

or

$$IC_{1,t}^2 + IC_{2,t}^2 - 2\phi_{12,t} IC_{1,t} IC_{2,t} + \phi_{12,t}^2 - 1 \le 0 . \tag{7.24}$$

For a given factor correlation, the expression on the left side describes an ellipse on the $\left(IC_{1,t}, IC_{2,t}\right)$-plane, and the two ICs must lie inside the ellipse. Figure 7.2 plots the ellipse and the region within for a factor correlation of 0.5. The major axis of the ellipse lies on the line $IC_{1,t} = IC_{2,t}$, and the minor axis on the line $IC_{1,t} = -IC_{2,t}$. This is true as long as $\phi_{12,t} \ge 0$. When the factor correlation is negative, the two axes switch places. Statistically, the two ICs can be anywhere inside the ellipse. As seen from the graph, the possibilities are numerous: they can be both positive, both negative, or have opposite signs.

Another way to look at the influence of the factor correlation on the two ICs is to express IC_2 in terms of IC_1, ϕ_{12}, and a residual IC, $IC_{\varepsilon 2,1}$, as

$$IC_2 = \phi_{12} \cdot IC_1 + \sqrt{1-\phi_{12}^2} \cdot IC_{\varepsilon 2,1} . \tag{7.25}$$

Here, we suppress the subscript t for clarity. The residual IC, $IC_{\varepsilon 2,1}$, is the correlation between security returns and the residual factor score of $\mathbf{F}_2$ after netting out $\mathbf{F}_1$. Because the correlation between the two factors is ϕ_{12} and the two factors are standardized, the residual factor, $\boldsymbol{\varepsilon}_{2,1}$, is simply $\boldsymbol{\varepsilon}_{2,1} = \mathbf{F}_2 - \phi_{12}\mathbf{F}_1$ and it is orthogonal to $\mathbf{F}_1$. It is easy to prove that the correlation $IC_{\varepsilon 2,1}$ between the residual factor $\boldsymbol{\varepsilon}_{2,1}$, and the return is related to other terms by (7.25). Furthermore, as $\boldsymbol{\varepsilon}_{2,1}$ is orthogonal to $\mathbf{F}_1$, the residual

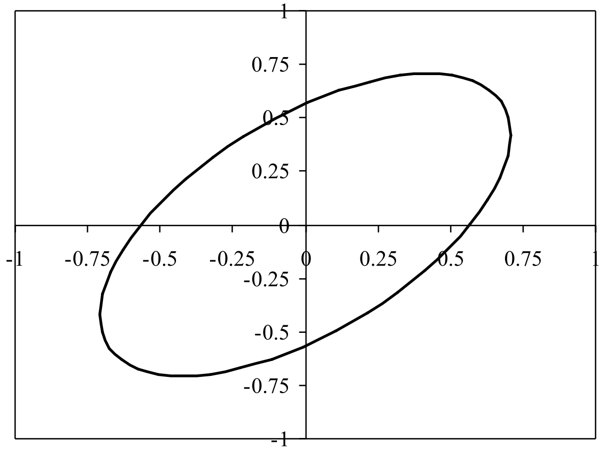

FIGURE 7.2. Feasible region of IC for two factors with correlation of 0.5.

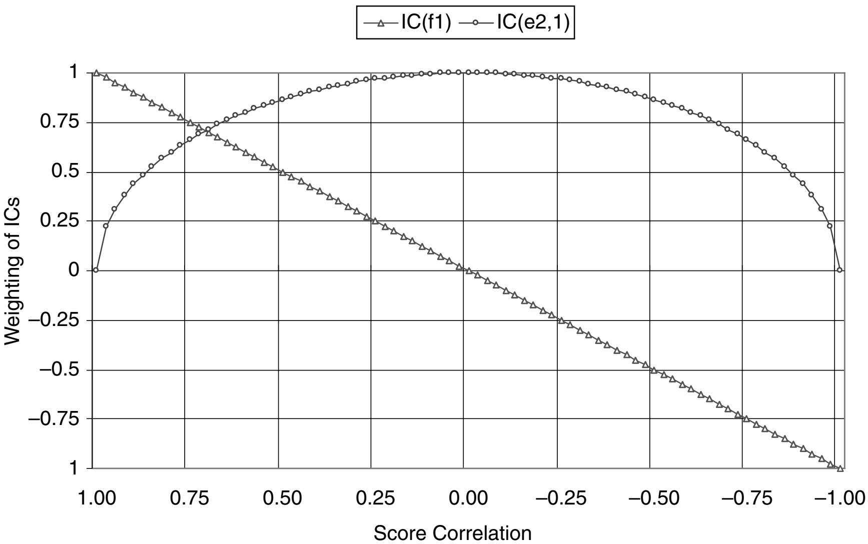

FIGURE 7.3. Weighting of ICs with score correlations.

correlation $IC_{\varepsilon 2,1}$ is completely free, i.e., it can be any number between –1 and 1. Based on (7.25), IC_2 can be as high as $IC_2 = \phi_{12} \cdot IC_1 + \sqrt{1-\phi_{12}^2}$ and as low as $IC_2 = \phi_{12} \cdot IC_1 - \sqrt{1-\phi_{12}^2}$.

We can also interpret IC_2 as a weighted, linear combination of IC_1 and $IC_{\varepsilon 2,1}$ whose weighting is a function of the score correlation, ϕ_{12} . Figure 7.3 shows how the weighting of IC_1 and $IC_{\varepsilon 2,1}$ varies with ϕ_{12} . The influence of IC_1 is linearly proportional to ϕ_{12}, ranging from 1 to −1, whereas the influence of $IC_{\varepsilon 2,1}$ is not only always positive but also a concave function. As such, $IC_{\varepsilon 2,1}$ generally exhibits more influence in determining IC_2 than IC_1. For example, when ϕ_{12} is equal to 0.975 — extremely close to a perfect-score correlation — the weights for IC_1 and $IC_{\varepsilon 2,1}$ are 0.975 and 0.222, respectively, implying that $IC_{\varepsilon 2,1}$ still commands a material influence. In contrast, when factor scores are close to being uncorrelated, such as ϕ_{12} being equal to 0.025, the weights for IC_1 and $IC_{\varepsilon 2,1}$ are 0.025 and 0.9997, respectively. In this instance, the influence of IC_1 is no longer material.

7.3.2 Multiperiod IC Correlations

The discussion so far has focused on the ICs and factor correlation of a single period, and they are calculated based on a cross section of two risk-adjusted forecast vectors and risk-adjusted returns of N stocks. As we extend from a single period to multiple periods, all three correlation coefficients in matrix (7.22) fluctuate, forming time-series or distributions. For instance, $IC_{1,t}$ and $IC_{2,t}$ each has sample (theoretical) and empirical distributions. Our interest is on the statistical properties of their distribution.

One of the major findings from Chapter 4 is that, even though the naive estimation for the standard deviation of IC is $1/\sqrt{N}$ or the sampling error, with N being the number of stocks, empirically the IC standard deviation for the majority of alpha factors we considered, is much higher than the naive estimation. With two or more factors, we are interested in the correlation between their ICs over time because they play a crucial role in determining the IR of multifactor alpha models. In this section, we first present a naive estimation of the IC correlation and then examine IC correlations empirically.

One naive estimate of IC correlation follows the general theory of sample covariance matrix based on a multivariate normal distribution. Under certain assumptions, the sample covariance matrix follows a Wishart distribution (see Muirhead 1982), and the covariance between the ICs is given by the following equation:

$$\operatorname{cov}\left(IC_{1,t}, IC_{2,t}\right) = \frac{1}{N}\left(\bar{\phi}_{12} + \overline{IC}_1 \cdot \overline{IC}_2\right). \tag{7.26}$$

The left-hand side is the covariance between the two ICs. On the right-hand side, N is the number of stocks; the barred variables are the averages of factor correlations and the averages of ICs. In practice, the average IC of the alpha factors is usually small. We approximate Equation 7.26 by

$$\operatorname{cov}\left(IC_{1,t}, IC_{2,t}\right) = \operatorname{std}\left(IC_1\right)\operatorname{std}\left(IC_2\right)\operatorname{corr}\left(IC_1, IC_2\right) \approx \frac{1}{N}\bar{\phi}_{12}. \quad (7.27)$$

Therefore, we have

$$\operatorname{corr}\left(IC_1, IC_2\right) \approx \frac{\bar{\phi}_{12}}{N\operatorname{std}\left(IC_1\right)\operatorname{std}\left(IC_2\right)}. \quad (7.28)$$

Equation 7.28 is the naïve estimation of the IC correlation. Furthermore, when the standard deviations of ICs are solely due to sampling error, they are equal to $1/\sqrt{N}$, i.e., $\operatorname{std}\left(IC_1\right) = \operatorname{std}\left(IC_2\right) = 1/\sqrt{N}$. If that were the case, then the IC correlation would be approximately the same as the average factor correlation, i.e., $\operatorname{corr}\left(IC_1, IC_2\right) \approx \bar{\phi}_{12}$.

When the standard deviations of ICs are greater than the sampling error, the IC correlation, as demonstrated in Chapter 4 and according to (7.28), should be *in theory* of the same sign as the factor correlation but less than the factor correlation. For models with more than two factors, Equation 7.28 applies to every pairwise IC correlation.

- Previous researchers seem to have focused solely on factor correlation, ignoring IC correlation. For analysis of multiperiod IR, we have established a theoretical link between the IC correlation and the factor correlation, which is only valid under the most ideal assumptions. Although the link provides some theoretical justification for previous research using factor correlation, it also highlights their limitation.

Example 7.6

If the average factor correlation is 0.5, $N = 1000$, and if the standard deviations of both ICs are $1/\sqrt{N}$, i.e., 0.032, then the IC correlation should also be 0.5. However, if the standard deviations of IC are 0.04 and 0.05, respectively, the IC correlation should be $0.5/(1000 \times 0.04 \times 0.05) = 0.25$, half the factor correlation.

TABLE 7.3 Average and Standard Deviation of Factor Correlations

Average (Stdev)	CFO2EV	XF	Ret9
CFO2EV	1.00 (0.00)	0.31 (0.09)	–0.04 (0.10)
XF		1.00 (0.00)	0.06 (0.06)
Ret9			1.00 (0.00)

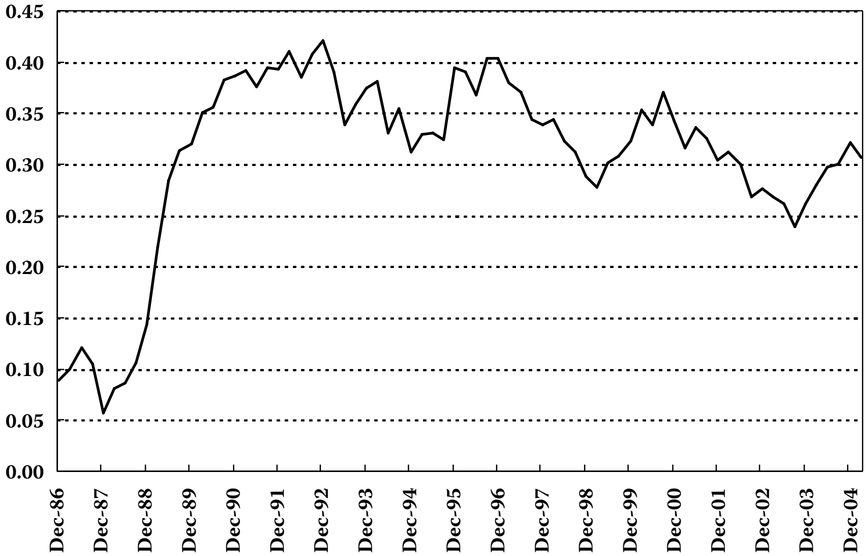

FIGURE 7.4. Quarterly factor correlations between CFO2EV and XF.

7.3.3 Empirical Examination of Factor Correlation and IC Correlation

It is probably safe to say that, in reality, many simplifying assumptions underlying theoretical models of the stock market break down. For instance, stock returns are generally not normally distributed. We also saw another example in Chapter 4 in the standard deviation of IC. We will now examine another case concerning the IC correlation.

Continuing the empirical example in the last section, Table 7.3 shows the average and standard deviation of factor correlations over the entire period. It is interesting to note that the correlation between CFO2EV and XF has an average of 0.31 and a standard deviation of 0.09, so it is significantly positive. The correlation between CFO2EV and Ret9 is slightly negative, whereas the correlation between XF and Ret9 is slightly positive. Figure 7.4 plots the time series of the factor correlations between CFO2EV and XF. It is initially low in 1987 and then increases to around 0.4 in 1990. Since then it has been fluctuating between 0.3 and 0.4.

TABLE 7.4 The IC Correlations of Three Factors

	CFO2EV	XF	Ret9
CFO2EV	1.00	0.73	−0.50
XF		1.00	−0.22
Ret9			1.00

TABLE 7.5 Sampling Errors of Time-Series IC Correlations

	ρ	std(ρ)	2-std Interval
ρ(IC_XF, IC_CFO2EV)	0.73	0.08	(0.56, 0.89)
ρ(IC_RET9, IC_CFO2EV)	−0.50	0.10	(−0.71, −0.29)
ρ(IC_RET9, IC_XF)	−0.22	0.12	(−0.45, 0.02)

The correlations of risk-adjusted ICs for the three factors are presented in Table 7.4. We note that they are significantly different from the factor correlations seen in Table 7.3. For example, the IC correlation between CFO2EV and XF is 0.73, which is significantly higher than the average factor correlation of 0.31, indicating that the diversification benefit between these two factors is not as strong as it would seem. On the other hand, the IC correlation between CFO2EV and Ret9 is −0.5, which is significantly lower than the factor correlation between the two. This seems to be a general phenomenon for value factors and price momentum factors as the IC diversification between them is significantly better than what the factor correlation would otherwise indicate. Lastly, the IC correlation between the quality factor XF and the price momentum factor Ret9 is slightly negative.

In our example, two out of the three IC correlations are significantly different from the factor correlations even if we take into account the variability of factor correlations over the entire period. We can calculate the confidence interval of IC correlations to provide another perspective. The standard deviation of IC correlation is approximately given by in the sample IC and the number of quarters Q (Keeping, 1995)

$$\text{std}(\rho_{IC}) = \frac{(1-\rho_{IC}^2)}{\sqrt{Q-1}}\sqrt{1+\frac{11\rho_{IC}^2}{2(Q-1)}}\,. \tag{7.29}$$

Table 7.5 shows the sampling error of the time-series IC correlations as well as their two standard deviation confidence intervals. All three cross-sectional score correlations fall out of their corresponding confidence interval. In fact, for the first two pairs, their average factor correlations lie outside the three standard deviations confidence interval.

7.4 COMPOSITE ALPHA MODEL WITH ORTHOGONALIZED FACTORS

Our analysis so far has focused on building composite models with the risk-adjusted factors. We have shown that the optimal weights of factors depend on average ICs and the covariance matrix of ICs. This provides important insights into factor diversification: factors with low IC correlations are more desirable than factors with high IC correlation, as the previous example illustrates.

We have made several simplifying assumptions, though. First, we standardized all risk-adjusted factors so that their cross-sectional dispersions remain unity. Second, we assumed that correlations among factors are constant over time. These assumptions made the problem of optimizing IR analytically tractable and led to our solution for the optimal weights and insight about factor diversification.

However, factor correlations are time varying, as we have shown in the last section in Figure 7.4. The fact that the variation in factor correlations is relatively small compared to the IC volatility justifies our approximation approach. Nevertheless, it would be desirable to derive a solution without this simplification. We can do so with orthogonalized factors. Factor orthogonalization can be viewed as another step in preprocessing factors along with factor standardization. When the procedure is carried out in every time period, the factor correlations will always be zero and thus constant.

When the factors are both orthogonal and standardized, the single-period IC of a composite (7.8) reduces to

$$IC_{c,t} = \frac{1}{\sqrt{\mathbf{v}' \cdot \mathbf{v}}} \sum_{i=1}^{M} v_i IC_{i,t} \ . \tag{7.30}$$

Because the ICs are now the only terms that vary in time, the IR of the model will be exactly that of (7.14), and the previous solution of optimal weights applies without any approximation.

7.4.1 Gram–Schmidt Procedure

A common mathematical technique, the Gram–Schmidt procedure sequentially makes each factor orthogonal to previously orthogonalized factors. Suppose we have M factors $\left(\mathbf{F}_1,\mathbf{F}_2,\cdots,\mathbf{F}_M\right)$ that have been standardized. With no particular order, the first factor $\mathbf{F}_1$ will be the first orthogonal factor, i.e., $\mathbf{F}_1^o=\mathbf{F}_1$, with the superscript denoting orthogonalized factors. Then the second orthogonal factor is defined as

$$\mathbf{F}_2^o=\frac{1}{\sqrt{1-\hat{\rho}_{21}^2}}\left(\mathbf{F}_2-\hat{\rho}_{21}^2\mathbf{F}_1^o\right), \tag{7.31}$$

where $\hat{\rho}_{21}=\rho_{21}$ is the cross-sectional correlation between $\mathbf{F}_2$ and $\mathbf{F}_1^o$, which is the same as the correlation between $\mathbf{F}_2$ and $\mathbf{F}_1$. The orthogonlized factor $\mathbf{F}_2^o$ is the factor $\mathbf{F}_2$ with the effect of $\mathbf{F}_1^o$ taken out. The ratio $1\Big/\sqrt{1-\hat{\rho}_{21}^2}$ makes $\mathbf{F}_2^o$ standardized. Moving on to the third factor, let $\hat{\rho}_{31}$ and $\hat{\rho}_{32}$ be the correlation between $\mathbf{F}_3$ and $\mathbf{F}_1^o$ and $\mathbf{F}_3$ and $\mathbf{F}_2^o$, respectively, which are calculated after we have derived the orthogonalized factor. Then,

$$\mathbf{F}_3^o=\frac{1}{\sqrt{1-\hat{\rho}_{32}^2-\hat{\rho}_{31}^2}}\left(\mathbf{F}_3-\hat{\rho}_{32}\mathbf{F}_2^o-\hat{\rho}_{31}\mathbf{F}_1^o\right) \tag{7.32}$$

is a standardized factor orthogonal to both $\mathbf{F}_2^o$ and $\mathbf{F}_1^o$. In general, suppose $\left(\mathbf{F}_1^o,\cdots,\mathbf{F}_{p-1}^o\right)$ are orthogonalized factors; then, for the factor $\mathbf{F}_p$, we first calculate its correlations with $\left(\mathbf{F}_1^o,\cdots,\mathbf{F}_{p-1}^o\right)$ and denote them by $\left(\hat{\rho}_{p1},\cdots,\hat{\rho}_{p,p-1}\right)$. The orthogonalized factor is given by

$$\mathbf{F}_p^o=\frac{1}{\sqrt{1-\hat{\rho}_{p1}^2-\hat{\rho}_{p2}^2-\cdots-\hat{\rho}_{p,p-1}^2}}\left(\mathbf{F}_p-\hat{\rho}_{p1}\mathbf{F}_1^o-\hat{\rho}_{p2}\mathbf{F}_2^o-\cdots-\hat{\rho}_{p,p-1}\mathbf{F}_{p-1}^o\right). \tag{7.33}$$

The factor $\mathbf{F}_p^o$ is proportional to the component of $\mathbf{F}_p$, which is uncorrelated with the previous orthogonlized factors.

Orthogonal factors produced by the Gram–Schmidt procedure can attest whether or not the original factors have independent information about forward returns. This is true if the IC of an orthogonalized factor is still positive and significant. However, if the IC of an orthogonalized factor becomes insignificant or even changes sign, its weight in the optimal model will likely change dramatically.

7.4.2 Optimal Model with the Gram–Schmidt Procedure

How do we combine the orthogonalized factors into an optimal alpha model? Recall the solution for weights of the optimal alpha model that is

TABLE 7.6 Average IC and Standard Deviation of IC for the Three Orthogonalized Factors

	CFO2EV.o	XF.o	Ret9.o
Average IC	0.06	0.02	0.05
Standard deviation	0.05	0.03	0.09
Annualized IR	2.09	1.36	1.15

given by $\mathbf{v}^* = s\boldsymbol{\Sigma}_{IC}^{-1}\overline{\mathbf{IC}}$ in (7.17), where $\boldsymbol{\Sigma}_{IC}^{-1}$ is the inverse of the IC covariance matrix, $\overline{\mathbf{IC}}$ is the average IC of the factors, and s is a scalar. The optimal model of orthogonalized factors follows the same form. We illustrate it with the three factors used in the previous example: cash flow from operating to enterprise value (CFO2EV), external financing (XF), and 9-month return (Ret9). In the Gram–Schmidt procedure, we have picked CFO2EV as the first factor, XF as the second, and Ret9 as the third.

Table 7.6 lists the average IC, the standard deviation of the orthogonalized factors, and the IR. As CFO2EV is the first factor, the orthogonalized version CFO2EV.o is the same as the original factor. The second factor XF.o differs significantly from the original factor. Compared to Table 7.3, both the average IC and the standard deviation of IC decrease, and the IR is less than that of the original factor. The reason is that the factor correlation between XF and CFO2EV is reasonably high, and hence the orthogonalization procedure greatly affects XF. On the other hand, the last factor Ret9 has little correlation with the other two factors, so Ret.o is almost the same as Ret9.

- As the example shows, the Gram–Schmidt procedure affects factors that have high correlations with other factors. This is especially true for factors in the same factor category: for example, earning yield and dividend yield in the value category.

Table 7.7 shows the IC correlations of the orthogonalized factors. In general, we should expect ICs of the orthogonalized factors to be less correlated than the original factors because their factor correlations are constructed to be zero. This seems to be true for two pairs of factors. Factors CFO2EV.o and XF.o have IC correlation of 0.34 compared to the IC correlation of 0.73 for CFO2EV and XF. Factors XF.o and Ret9.o have IC correlation of −0.03 compared to the IC correlation of −0.22 for the original factors (Table 7.2). However, the other IC correlation between CFO2EV.o and Ret9.o shows no change.

TABLE 7.7 The IC Correlations of Three Orthogonalized Factors

	CFO2EV.o	XF.o	Ret9.o
CFO2EV.o	1.00	0.34	−0.50
XF.o		1.00	−0.03
Ret9.o			1.00

TABLE 7.8 Weights of Alpha Models and Corresponding IR Based on the Three Orthogonalized Factors

	IR	CFO2EV.o	XF.o	Ret9.o
$\mathbf{w}_1$	2.85	40%	47%	13%
$\mathbf{w}^*$	3.30	61%	9%	30%

Table 7.8 shows the sets of weights of optimal alpha models based on the orthogonalized factors — one with the full IC covariance matrix and the other with diagonal IC covariance matrix. Compared to Table 7.4, the optimal weight $\mathbf{w}^*$ has a positive 9% in XF.o, and the IR increases slightly. The IR of $\mathbf{w}_1$ shows greater improvement from that of Table 7.4 because the IC correlations of the orthogonalized factors play a lesser role in determining the optimal IR. Note the following:

- Another method of factor orthogonalization is principal component analysis, or PCA. The principal components (PC) of $\left(\mathbf{F}_1,\mathbf{F}_2,\cdots,\mathbf{F}_M\right)$ are their linear combinations. The first PC is the linear combination of $\left(\mathbf{F}_1,\mathbf{F}_2,\cdots,\mathbf{F}_M\right)$ that has the largest cross-sectional dispersion, and the second PC is the combination of $\left(\mathbf{F}_1,\mathbf{F}_2,\cdots,\mathbf{F}_M\right)$ uncorrelated to the first PC that has the largest cross-sectional dispersion, and so on. The PCA technique is theoretically appealing, but it has one practical difficulty. Because principal components are unique up to a change in signs, one has to ensure that "same" PCs are selected over time. This could be a challenge if the correlation structure of factors changes drastically over time.

7.5 FAMA–MACBETH REGRESSION AND OPTIMAL ALPHA MODEL

Although most practitioners recognize the benefit of combining multiple alpha sources in terms of IR improvement, their approaches to construct a multifactor alpha model vary widely. The analytical framework developed so far in this book relies on the risk-adjusted ICs of individual factors and their correlations. One of the key facts for a multifactor alpha model is that the excess returns from individual factors are essentially additive; the overall excess return is a linear combination of individual excess returns, whereas the factor correlations enter the linear combination through a scaling factor.

There are practitioners who employ other statistical framework and derive forecasts based on empirical asset pricing back-test procedure, such as the Fama–MacBeth (1973) regression, which consists of a series of cross-sectional OLS regressions. Even though the Fama–MacBeth regression is simple to implement and intuitively appealing, it is used in most asset pricing studies to ascertain whether a factor is priced. The question is whether it provides an analytical foundation for combining multiple alpha sources.

To answer this question, we should first give an economic interpretation of the regression coefficients in a cross-sectional OLS regression. The key question is whether the regression coefficients represent the excess returns of certain active portfolios, and, if they do, what are the alpha factors behind these active portfolios?

7.5.1 Univariate OLS Regression

When there is just one independent factor in the cross-sectional regression, the interpretation is straightforward. Suppose the regression takes the form

$$\mathbf{r}_t = \alpha_t + \beta_t \mathbf{f}_t . \tag{7.34}$$

Then the coefficient is

$$\beta_t = \frac{\operatorname{cov}\left(\mathbf{r}_t, \mathbf{f}_t\right)}{\operatorname{var}\left(\mathbf{f}_t\right)} = \frac{\operatorname{corr}\left(\mathbf{r}_t, \mathbf{f}_t\right)\operatorname{dis}\left(\mathbf{r}_t\right)}{\operatorname{dis}\left(\mathbf{f}_t\right)} . \tag{7.35}$$

When the factor is standardized, the regression coefficient is IC times the dispersion of realized returns, i.e.,

$$\beta_t = \operatorname{corr}\left(\mathbf{r}_t, \mathbf{f}_t\right)\operatorname{dis}\left(\mathbf{r}_t\right) . \tag{7.36}$$

Comparing Equation 7.36 with Equation 7.6, we see that, in this case, the regression coefficient is proportional to the excess return of an active portfolio based on the factor.

7.5.2 OLS Regression with Multiple Factors

When there are multiple factors, the OLS regression coefficients are no longer the ICs of individual factors, unless the factors are uncorrelated. However, what are their economic interpretations in the context of excess returns? To develop insight into this question, we consider the case with two factors and derive the coefficients explicitly. The regression equation is

$$\mathbf{r}_t = \alpha_t + \beta_{1,t}\mathbf{f}_{1,t} + \beta_{2,t}\mathbf{f}_{2,t} . \tag{7.37}$$

The coefficients in terms of variances and covariances are given by

$$\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}^{-1} \begin{pmatrix} IC_1 \\ IC_2 \end{pmatrix} \mathrm{dis}\left(\mathbf{r}_t\right). \tag{7.38}$$

Again, we have assumed that the factors are standardized, with variance being 1, and ρ_t denotes the factor or score correlation. Inverting the matrix and multiplying the ICs gives

$$\begin{aligned} \beta_1 &= \frac{1}{1-\rho^2}\left(IC_1 - \rho IC_2\right)\mathrm{dis}\left(\mathbf{r}\right) \\ \beta_2 &= \frac{1}{1-\rho^2}\left(IC_2 - \rho IC_1\right)\mathrm{dis}\left(\mathbf{r}\right) \end{aligned} \tag{7.39}$$

We have suppressed subscript t for clarity. The coefficients are combinations of ICs, with the factor correlation entering as one of the weights. When the two factors are uncorrelated, the coefficients are identical to the univariate regression coefficients.

The economic interpretation of β_1 is the *marginal* return contribution of $\mathbf{f}_1$ after netting out the influence of $\mathbf{f}_2$. Similarly, β_2 represents the *marginal* return contribution of $\mathbf{f}_2$ after controlling the influence of $\mathbf{f}_1$. To see this, we note that both β_1 and β_2 can be derived from two separate univariate OLS regressions with cross-sectional return as the dependent variable. For instance, to derive β_1, we first regress $\mathbf{f}_1$ against $\mathbf{f}_2$:

$$\mathbf{f}_1 = \rho\mathbf{f}_2 + \boldsymbol{\varepsilon}_{1,2} . \tag{7.40}$$

The residual is then $\boldsymbol{\varepsilon}_{1,2} = \mathbf{f}_1 - \rho\mathbf{f}_2$. To be consistent with factor standardization, we standardize the residual so that its cross-sectional dispersion is unity:

$$\tilde{\boldsymbol{\varepsilon}}_{1,2} = \frac{\mathbf{f}_1 - \rho\mathbf{f}_2}{\sqrt{1-\rho^2}}. \tag{7.41}$$

ICs of both $\tilde{\boldsymbol{\varepsilon}}_{1,2}$ (standardized residual) and $\boldsymbol{\varepsilon}_{1,2}$ (raw residual) are the same:

$$\widetilde{IC_1} = \frac{\operatorname{cov}\left(\tilde{\boldsymbol{\varepsilon}}_{1,2}, \mathbf{r}\right)}{\operatorname{dis}\left(\mathbf{r}\right)} = \frac{IC_1 - \rho IC_2}{\sqrt{1-\rho^2}}. \tag{7.42}$$

In the second univariate regression, let $\beta_{\mathbf{r},\boldsymbol{\varepsilon}_{1,2}}$ be the coefficient estimate of a cross-sectional regression, wherein the cross-sectional return, $\mathbf{r}_t$, is the dependent variable, and raw residual of $\boldsymbol{\varepsilon}_{1,2}$ is the independent variable. As the following equation shows, $\beta_{\mathbf{r},\boldsymbol{\varepsilon}_{1,2}}$ is exactly the same as β_1

$$\beta_{\mathbf{r},\boldsymbol{\varepsilon}_{1,2}} = \frac{\operatorname{cov}(\boldsymbol{\varepsilon}_{1,2}, \mathbf{r})}{\operatorname{var}(\boldsymbol{\varepsilon}_{1,2})} = \frac{\operatorname{cov}(\mathbf{f}_1 - \rho\mathbf{f}_2, \mathbf{r})}{1-\rho^2} = \frac{IC_1 - \rho \cdot IC_2}{1-\rho^2} \cdot \operatorname{dis}(\mathbf{r}) = \beta_1. \tag{7.43}$$

Similarly, the IC of factor 2 with factor 1 regressed out is

$$\widetilde{IC_2} = \frac{IC_2 - \rho IC_1}{\sqrt{1-\rho^2}}. \tag{7.44}$$

Comparing Equation 7.39, Equation 7.42, and Equation 7.44 shows that multivariate regression coefficients are related to *residual ICs* as

$$\begin{aligned} \beta_1 &= \frac{1}{\sqrt{1-\rho^2}} \widetilde{IC_1} \operatorname{dis}\left(\mathbf{r}\right) = \beta_{\mathbf{r},\boldsymbol{\varepsilon}_{1,2}} \\ \beta_2 &= \frac{1}{\sqrt{1-\rho^2}} \widetilde{IC_2} \operatorname{dis}\left(\mathbf{r}\right) = \beta_{\mathbf{r},\boldsymbol{\varepsilon}_{2,1}} \end{aligned} \tag{7.45}$$

- The residual IC is, in essence, the information coefficient of a composite factor whose weights are related to the factor correlation. For example, $\widetilde{IC_1}$ is the IC of factor $\tilde{\boldsymbol{\varepsilon}}_{1,2} = \left(\mathbf{f}_1 - \rho\mathbf{f}_2\right) \big/ \sqrt{1-\rho^2}$. Depending on the factor correlation, the residual IC could be very different from the IC of the individual factor.

Example 7.7

Suppose $IC_1 = 0.2$, $IC_2 = 0.1$, and $\rho = 0.8$. Then the residual ICs are $\widetilde{IC_1} = \left(0.2 - 0.8 \cdot 0.1\right) \big/ \sqrt{1-0.8^2} = 0.2$ and $\widetilde{IC_2} = \left(0.1 - 0.8 \cdot 0.2\right) \big/ \sqrt{1-0.8^2} =$ -0.1. Even though both factors have positive ICs, one residual IC is positive and the other is negative! This is due to the high correlation between the two factors. If the correlation is reduced to 0.5 from 0.8, the residual ICs are $\widetilde{IC_1} = 0.17$ and $\widetilde{IC_2} = 0.0$, respectively. The second factor is rendered as having no information.

When the factor correlation is negative, the residual ICs are going to be higher than the original ICs. The lesson is that one should not interpret multivariate regression coefficients as returns to alpha factors; instead, they are *marginal* returns to alpha factors after netting out influences from other factors. Especially, they should not be used in performance attribution of alpha factors. This is particularly problematic or simply wrong when the factors from the same category have high correlations, as we have seen in Chapter 5. For instance, earnings yield and cash flow yield tend to have high factor-score correlation, as both are constructed with the price as the denominator. Just because one worked better than the other in terms of higher IC, we cannot conclude that the lesser one had a negative contribution to the portfolio return.

7.5.3 Fama–MacBeth Regression and Asset Pricing Tests

Fama–Macbeth regression is commonly used by academic researchers to ascertain whether a factor is priced by the market through time after controlling for other known, priced factors such as beta, book-to-price, size, or price momentum. The procedure consists of a series of multiple OLS regressions for each cross section of securities. In each regression, cross-sectional returns form the dependent variable; and independent variables consist of two parts: control variables and a set of tested factors. Control variables are deployed to ensure that the tested pricing phenomenon was not subsumed by other known pricing phenomena. In other words, it is a test of whether the factor in question provides incremental pricing information. For illustrative purpose, let us assume that $\mathbf{f}_1$ is a control variable and $\mathbf{f}_2$ is the factor in question. Each cross-sectional regression at time t is formulated as $\mathbf{r}_t = \alpha_t + \beta_{1,t}\mathbf{f}_{1,t} + \beta_{2,t}\mathbf{f}_{2,t}$. Factor $\mathbf{f}_2$ is considered as a priced factor if its time series t-stat $t = \overline{\beta}_2 / \text{std}(\beta_{2,t})$ is significantly different from zero. In other words, should $t(\beta_{2,t})$ be significantly different from zero, then $\mathbf{f}_2$ is said to be priced by the market after controlling for the known asset pricing phenomenon of $\mathbf{f}_1$.

Equation 7.45 shows this residual effect directly because it connects the OLS regression coefficients to the ICs of residual factors. When factor correlation ρ is stable and the return dispersion is constant, it is easily seen that the Fama–MacBeth t-stat is proportional to the IR of residual factors.

The interpretation of multivariate regression coefficients as coefficients of univariate regressions of return vs. residual factors provides critical insight into the results of the Fama–MacBeth regression. It turns out that this interpretation remains true as we add control variables (or risk factors) and more alpha factors into the OLS regression. Suppose we have

$$\mathbf{r} = \alpha + b_1\mathbf{I}_1 + \cdots + b_K\mathbf{I}_K + \beta_1\mathbf{f}_1 + \cdots + \beta_L\mathbf{f}_L, \tag{7.46}$$

where $\left(\mathbf{I}_1, \cdots, \mathbf{I}_K\right)$ are control variables and $\left(\mathbf{f}_1, \cdots, \mathbf{f}_L\right)$ are alpha factors, then the coefficient β_j can be obtained in the following steps for each cross section at a given time t, and these steps are repeated through time to derive a time series of estimates of β_j (see appendix for proof).

- Step 1: We regress factor $\mathbf{f}_j$ against all control variables and remaining alpha factors simultaneously.
- Step 2: We take the residual of the regression in Step 1 and run a univariate regression of returns against the residual to obtain β_j.

Similar to Equation 7.45, the coefficient β_j is related to the IC of the residual, the dispersion of the actual return, and the dispersion of the residual.

- There is a connection between the residual IC and the IC of the purified alpha in Chapter 4. The purified alpha is an alpha signal with the risk factors regressed out. The residual IC that is contained in the multivariate regression (7.46) is the IC of an alpha signal with not only the risk factors but also all other alpha factors regressed out. It is an alpha signal so "pure" that it is orthogonal to both risk factors and other alpha factors.

7.5.4 Multifactor Model through Fama–MacBeth Regression

Although multivariate regression coefficients should be interpreted as return sensitivities to residual factor scores, a naive application of the Fama–MacBeth regression in deriving factor returns and optimal model weighting would result in erroneous model estimation due to factor-score correlations. There are two methods to alleviate the problem. First, recall if the factors are uncorrelated, and then the coefficients become sensitive to the factors and proportional to the factors' ICs. Thus, one simple way to avoid the collinear problem is to sequentially orthogonalize factor scores through the Gram–Schmidt procedure before each cross-sectional OLS regression. Then, using the coefficients, we can estimate the average ICs and covariances of IC to derive the optimal alpha model. This is the same model derived under the Gram–Schmidt procedure.

In the second method, one may choose not to orthogonalize the factors. Given the interpretation of regression coefficients in the Fama–MacBeth regression, one can still construct a multifactor model using the regression coefficients based on residual ICs. As we have shown, the residual IC can be easily derived from the Fama–MacBeth regression coefficients. We can find optimal weights that maximize the IR of the residual ICs, i.e., the average of residual IC to its standard deviation. This is similar to our approach of finding optimal weights based on the ICs of individual factors. However, there is one crucial difference. Models constructed through the Fama–MacBeth regression coefficients are no longer models for the original factors. Rather, they should be used as models of the residual factors. To apply the weights of the model, one must first find the residual factors by performing multivariate regression on each factor against all other factors and compute a weighted sum of the residual factors as the composite model.

The procedure to find the optimal weights of residual factors is analogous to the previous procedure for the original factors. We shall not repeat it here. We focus instead on the connection between the two sets of models: the model that maximizes the IR of the original factors and the model that maximizes the IR of the residual factors. First, it should be noted that the optimal model of the residual factors could be transformed into a model of the original factors because the residual factors themselves are linear combination of the original factors. For instance, for two-factor cases, the residual factors are $\tilde{\boldsymbol{\varepsilon}}_{1,2} = \left(\mathbf{f}_1 - \rho \mathbf{f}_2\right) / \sqrt{1-\rho^2}$ and $\tilde{\boldsymbol{\varepsilon}}_{2,1} = \left(\mathbf{f}_2 - \rho \mathbf{f}_1\right) / \sqrt{1-\rho^2}$. If the model weights for the residual factors are $\tilde{v}_1$ and $\tilde{v}_2$, we have

$$\tilde{v}_1 \tilde{\boldsymbol{\varepsilon}}_{1,2} + \tilde{v}_2 \tilde{\boldsymbol{\varepsilon}}_{2,1} = \frac{\left(\tilde{v}_1 - \rho \tilde{v}_2\right)}{\sqrt{1-\rho^2}} \mathbf{f}_1 + \frac{\left(\tilde{v}_2 - \rho \tilde{v}_1\right)}{\sqrt{1-\rho^2}} \mathbf{f}_2 = v_1 \mathbf{f}_1 + v_2 \mathbf{f}_2 . \tag{7.47}$$

Conversely, a model of original factors can be transformed to a model of residual factors:

$$\tilde{v}_1 = \frac{v_1 + \rho v_2}{\sqrt{1-\rho^2}}, \tilde{v}_2 = \frac{v_2 + \rho v_1}{\sqrt{1-\rho^2}} . \tag{7.48}$$

Because of this linear transformation between the two sets of models, optimal models that maximize the information ratio utilizing either original factors or standardized residual factors are identical, provided that the factor correlations are constant over time. This is because the relationship between the residual IC and the original IC, and the relationship between the standardized residual factor and the original factors are identical (see, for example, Equations 7.41 and 7.42).

For the general case, denoting this constant linear relationship by matrix $\mathbf{P}$, we have

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{P} \cdot \mathbf{f} \text{ and } \mathbf{IC}_{\tilde{\boldsymbol{\varepsilon}}} = \mathbf{P} \cdot \mathbf{IC} . \tag{7.49}$$

The average residual IC and its covariance matrix are related to the average of the original IC and its covariance matrix by $\overline{\mathbf{IC}_{\tilde{\boldsymbol{\varepsilon}}}} = \mathbf{P} \cdot \overline{\mathbf{IC}}$ and $\boldsymbol{\Sigma}_{IC_\varepsilon} = \mathbf{P}' \boldsymbol{\Sigma}_{IC} \mathbf{P}$. The optimal weights (see Problem 7.9) for the residual factors are simply

$$\mathbf{v}_\varepsilon = \mathbf{P}^{-1} \boldsymbol{\Sigma}_{IC}^{-1} \overline{\mathbf{IC}} = \mathbf{P}^{-1} \mathbf{v} , \tag{7.50}$$

where $\mathbf{v} = \boldsymbol{\Sigma}_{IC}^{-1} \overline{\mathbf{IC}}$ is the optimal weights for the original factors. Therefore, the two composites with respective optimal weights are equal:

$$\mathbf{v}_{\boldsymbol{\varepsilon}}' \cdot \tilde{\boldsymbol{\varepsilon}} = \mathbf{v}' \mathbf{P}^{-1} \mathbf{P} \mathbf{f} = \mathbf{v}' \mathbf{f} . \tag{7.51}$$

- Another alternative for constructing a multifactor alpha model using Fama–MacBeth regression is to apply it directly to a predetermined combination of alpha factors plus risk factors from the outset (Yang, 2005). Unlike the multivariate setting, we now have just one composite alpha factor whose regression coefficient is directly linked to its IC after the effects of the risk factors are netted out. There is no residual effect involving other alpha factors. This is a version of purified alpha for a composite factor, and the regression coefficient is simply the multifactor IC times the dispersion of actual returns. When we carry out Fama–MacBeth regression over multiple time periods, the *t*-stat of the regression coefficient is a proxy of the IR for the predetermined combination of the alpha factors. This serves as a good indicator of portfolio performance for the given model. To find the optimal alpha model, however, we have to search for the optimal weights that maximize the *t*-stats of the regression coefficients by numerical means.

PROBLEMS

7.1 Calculate the dispersion and IC of the composite factor in Example 7.1 and 7.2 if the factor weights are 1/3 and 2/3, respectively.

7.2 Prove that the model weights that maximize single-period IC of (7.8) is (7.21).

7.3 Verify (7.17) to satisfy Equation 7.16. Find the value of s so that the sum of the model weights equals 1.

7.4 Assume that there are M alpha factors whose ICs are measured over T periods. We derive the optimal model weight $\mathbf{v}$ that maximizes IR by the following OLS regression:

$$\underset{(T\times 1)}{\mathbf{i}} = \underset{(T\times M)}{\mathbf{IC}} \times \underset{(M\times 1)}{\mathbf{v}} + \underset{(T\times 1)}{\mathbf{u}},$$

where $\mathbf{i}$ is a vector of ones — a constant dependent variable — $\mathbf{IC}$ is the observed IC matrix from the independent variables , $\mathbf{v}$ is the regression coefficients, and $\mathbf{u}$ is the error vector.

Prove that

(a) $\mathbf{v}=\left(\mathbf{IC}'\mathbf{IC}\right)^{-1}\left(\mathbf{IC}'\cdot\mathbf{i}\right)$;

(b) $\mathbf{IC}'\mathbf{IC}=\boldsymbol{\Sigma}_{\mathbf{IC}}+\overline{\mathbf{IC}}\cdot\overline{\mathbf{IC}}'$;

(c) $(\mathbf{IC}'\mathbf{IC})^{-1}=\boldsymbol{\Sigma}_{IC}^{-1}+\dfrac{\left(\boldsymbol{\Sigma}_{IC}^{-1}\overline{\mathbf{IC}}\cdot\overline{\mathbf{IC}}'\boldsymbol{\Sigma}_{IC}^{-1}\right)}{1+\overline{\mathbf{IC}}'\boldsymbol{\Sigma}_{IC}^{-1}\overline{\mathbf{IC}}}$;

(d) $\mathbf{v}=\dfrac{\boldsymbol{\Sigma}_{IC}^{-1}\overline{\mathbf{IC}}}{1+\overline{\mathbf{IC}}'\boldsymbol{\Sigma}_{IC}^{-1}\overline{\mathbf{IC}}}$.

7.5 Derive the optimal IR (7.20) for two-factor models.

7.6 Extend Figure 7.1 to the full range of IC correlation from –1 to 1. Show that, when the IC correlation is greater than 0.5, the optimal model weight of factor 2 is negative.

7.7 Prove that factor $\mathbf{F}_p^o$ in (7.33) is orthogonal to $\left(\mathbf{F}_1^o,\cdots,\mathbf{F}_{p-1}^o\right)$.

7.8 Given two residual terms $\boldsymbol{\varepsilon}_{1,2}=\mathbf{f}_1-\rho_t\mathbf{f}_2$ and $\boldsymbol{\varepsilon}_{2,1}=\mathbf{f}_2-\rho_t\mathbf{f}_1$, calculate their correlation coefficient.

7.9 Derive Equation 7.48.

7.10 (a) Suppose the standardized residual factors are related to the original factor through $\tilde{\boldsymbol{\varepsilon}}=\mathbf{P}\cdot\mathbf{f}$. Prove that $\mathbf{IC}_{\tilde{\boldsymbol{\varepsilon}}}=\mathbf{P}\cdot\mathbf{IC}$. (b) With averages and covariance matrix of residual ICs given by $\overline{\mathbf{IC}}_{\tilde{\boldsymbol{\varepsilon}}}=\mathbf{P}\cdot\overline{\mathbf{IC}}$ and $\boldsymbol{\Sigma}_{IC_\varepsilon}=\mathbf{P}'\boldsymbol{\Sigma}_{IC}\mathbf{P}$, show that the optimal weights for the standardized residual factors are related to the optimal weights for the original factors by $\mathbf{v}_\varepsilon=\mathbf{P}^{-1}\mathbf{v}$.

APPENDIX

In this appendix, we prove that a multivariate linear regression can be decomposed into two separate regressions: one between independent variables and the other between a dependent variable and the residual of the first regression. This property is inherent to the multivariate regression.

A7.1 INVERSE OF A PARTITIONED MATRIX

We first present the following result for the inverse of a nonsingular matrix. Given a square matrix $\boldsymbol{\Sigma}$, we partition it as block matrix in which the diagonal blocks $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are nonsingular square matrix:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}. \tag{7.52}$$

Define

$$\begin{aligned} \mathbf{\Sigma}_{11,2} &= \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \\ \mathbf{\Sigma}_{22,1} &= \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \end{aligned}. \tag{7.53}$$

Then the inverse is given by

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}_{11,2}^{-1} & -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22,1}^{-1} \\ -\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11,2}^{-1} & \mathbf{\Sigma}_{22,1}^{-1} \end{pmatrix}. \tag{7.54}$$

A7.2 DECOMPOSITION OF MULTIVARIATE REGRESSION

For a multivariate regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, the coefficient vector is given by $\boldsymbol{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$. Suppose all variables have zero mean. The covariance matrix of independent variables $\mathbf{x}$ is $\Sigma = \left(\sigma_{ij}\right)_{i,j=1}^{K}$, the standard deviation of the dependent variable y is σ_y, and the correlations between the independent variables and the dependent variable are $\left(s_1, \cdots, s_K\right)$. Then the regression coefficient can be written as

$$\boldsymbol{\beta} = \Sigma^{-1}\mathbf{s}. \tag{7.55}$$

The vector $\mathbf{s}$ consists of covariances between the independent variables and the dependent variable, i.e.,

$$\mathbf{s} = \left(s_1\sigma_1\sigma_y, \cdots, s_K\sigma_K\sigma_y\right)'. \tag{7.56}$$

We partition the independent variables into

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix},$$

where $\mathbf{x}_1$ consists of k_1 factors and $\mathbf{x}_2$ consists of k_2 factors, and $k_1 + k_2 = k$. The coefficient vector $\boldsymbol{\beta}$ and the vector $\mathbf{s}$ can also be partitioned into

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \ \mathbf{s} = \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix}.$$

The covariance matrix $\boldsymbol{\Sigma}$ can also be written as in (7.52), in which case $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are the covariance matrices for $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$ is the covariance matrix between $\mathbf{x}_1$ and $\mathbf{x}_2$. According to (7.55), we have

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{\Sigma}^{-1}\mathbf{s} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix}.$$

Using the inverse matrix (7.54) gives

$$\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11,2}^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22,1}^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11,2}^{-1} & \boldsymbol{\Sigma}_{22,1}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix}.$$

We now focus our attention on the coefficient $\boldsymbol{\beta}_1$ and obtain

$$\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}_{11,2}^{-1}\mathbf{s}_1 - \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22,1}^{-1}\mathbf{s}_2. \tag{7.57}$$

Next, we carry out the two-stage regression. First, we regress $\mathbf{x}_1$ against $\mathbf{x}_2$. As both dependent and independent variables are vectors in general, the regression coefficient is in fact a matrix in a form similar to (7.55) and it equals $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. Hence, the residual of this regression is

$$\boldsymbol{\varepsilon}_{1,2} = \mathbf{x}_1 - \mathbf{x}_2\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{7.58}$$

The second regression is to regress y vs. the residual $\boldsymbol{\varepsilon}_{1,2}$. Denoting the regression coefficient by $\tilde{\boldsymbol{\beta}}_1$, we can write its solution in the same form as (7.55), with the covariance matrix being that of the residuals and the vector $\mathbf{s}$ being the covariances between y and the residuals; i.e.,

$$\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_{1,2}}^{-1} \operatorname{cov}\left(y, \boldsymbol{\varepsilon}_{1,2}\right). \tag{7.59}$$

The covariance matrix of $\boldsymbol{\varepsilon}_{1,2}$ is

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_{1,2}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{11,2}. \qquad (7.60)$$

The covariances between y and the residuals are

$$\operatorname{cov}\left(y, \boldsymbol{\varepsilon}_{1,2}\right) = \mathbf{s}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{s}_2. \qquad (7.61)$$

Combining these, we have

$$\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{\Sigma}_{11,2}^{-1}\mathbf{s}_1 - \boldsymbol{\Sigma}_{11,2}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{s}_2. \qquad (7.62)$$

To prove $\boldsymbol{\beta}_1 = \tilde{\boldsymbol{\beta}}_1$ from Equation 7.57 and Equation 7.62, we need to prove that

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22,1}^{-1} = \boldsymbol{\Sigma}_{11,2}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1},$$

or

$$\boldsymbol{\Sigma}_{11,2}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{22,1}.$$

Substituting (7.53) into the preceding matrices gives

$$\left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right).$$

Multiplying the matrices leads to an identity

$$\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}. \qquad (7.63)$$

Equation 7.63 furnishes our proof for $\boldsymbol{\beta}_1 = \tilde{\boldsymbol{\beta}}_1$.

REFERENCES

Britten-Jones, M., Portfolio optimization and Bayesian regression, Q Group Conference, 1998.

Fama, E.F. and MacBeth, J.D., Risk, return and equilibrium: Empirical tests, *Journal of Political Economy*, Vol. 81, 607–636, 1973.