CHAPTER 9

Advanced Alpha Modeling Techniques

Quantitative equity portfolio management relies on both the alpha model and the risk model to construct a mean–variance efficient portfolio. The alpha model forecasts the excess return of each security by identifying pricing inefficiencies, whereas the risk model forecasts the covariance structure of the security return. The former delivers value added of active management in the form of portfolio returns in excess of its benchmarks; the latter provides portfolio risk control and diversification benefit. Although each plays a different role, both depend on the assumption of a return generating equation in constructing their forecasts.

In this chapter, we shall take a closer look at the return-generating equation behind most traditional quantitative models and present modeling techniques that provide a structured framework in relaxing many stringent assumptions behind the traditional approach. Specifically, we will first discuss three assumptions behind the commonly used return-generating equation: "one size fits all," "bigger is always better," and "time independence." We will then discuss various advanced modeling techniques that can achieve better alpha forecasts by relaxing the first two assumptions. Both assumptions are cross-sectional in nature. The techniques include contextual alpha modeling, sector modeling, and nonlinear effect modeling. We will address the third assumption in Chapter 10 by highlighting several time-varying modeling techniques.

9.1 THE RETURN-GENERATING EQUATION

Equation 9.1 postulates a generic return-generating equation, which expresses security returns in terms of exposures to factors. Security return is a linear combination of attributed returns to factors that possess cross-sectional explanatory power.

$$r_i = b_{i0} + b_{i1} I_1 + \cdots + b_{iK} I_K + \varepsilon_i . \tag{9.1}$$

In the equation, r_i is the return of stock i, $b_{i1},\cdots,b_{iK}$ are factor exposures of the stock, and $I_1, \cdots, I_K$ are factor returns. The residual portion of security return that is not attributed, is called *security specific return* and is expressed as ε_i. Note that in Equation 9.1 we dropped the subscript of time to simplify the notation. This equation serves as the core of risk models in Chapter 3. The covariance matrix of returns is given by

$$\mathbf{\Sigma} = \mathbf{B}\mathbf{\Sigma}_I\mathbf{B}' + \mathbf{S}, \tag{9.2}$$

where $\mathbf{\Sigma}_I$ is the factor return covariance matrix, **B** is the exposure matrix, and **S** is the diagonal specific variance matrix. Equation 9.2 forms the foundation of many commercially available risk models, such as BARRA, Northfield, or Citigroup GRAM. The only difference among them is the set of factors selected. For example, BARRA uses fundamental factors, whereas Northfield employs mostly macro economic factors.

Perhaps, due to its academic origin and popularity in commercial risk models, many active managers also adopt framework similar to (9.1) in constructing their proprietary alpha models. Specifically, they forecast expected return as

$$E(r_i) \propto f_{i1} v_1 + \cdots + f_{iM} v_M , \tag{9.3}$$

where $(f_{i1},\cdots,f_{iM})$ are cross-sectional alpha factors and $(v_1,\cdots,v_M)$ are the factor weights that are related to expected factor returns. Although methods of selecting the factor weights vary greatly among active managers (see Chapter 7 for the discussion), most methods conform to (9.3), which makes the following three unrealistic assumptions.

One size fits all: In Equation 9.3, the factor weights are the same for every security, thus making it a one-size-fits-all approach. However, most practitioners recognize the conditional nature of factor returns, and their intuitions find significant support from empirical research. For example, Daniel et al. (1999) find that momentum effects are stronger for growth stocks, and Asness (1997) finds that value strategies work, in general, but less so for stock with high momentum.

Bigger is always better: Because (9.3) is linear, it implies that the expected security return is linearly proportional to the factor exposure. For example, if buying cheap stocks is a good thing, then purchasing deep value securities must produce the best investment results. In reality, practitioners are often aware of the fact that deep value securities are often cheap for a reason. For example, Bruce and Morillo (2003) find that expected returns of securities with extreme factor values tend to break away from their linear expectations, sometimes in a fairly dramatic way.

Time independence: The last assumption deals with the constancy of factor weights over time, making it an unconditional model. In reality, factor returns change through time, depending on various macroeconomic regimes or even different calendar events. This time-varying behavior is ignored in (9.3).

In all, the linear one-size-fits-all return-generating equation provides a resilient foundation for risk models. However, the same equation is an inadequate foundation for forecasting the expected security return, mostly due to the linearity assumption. Such inadequacy is born out of the fact that security markets are quasi-efficient wherein many sophisticated managers try to arbitrage the same set of behavioral phenomenon. Simplistic alpha models such as (9.3) deliver inferior portfolio excess returns. In the rest of this chapter, we shall present several advanced modeling techniques.

9.2 CONTEXTUAL MODELING

In practice, linking a stock's ranking signal or factor to expected return and assigning it an appropriate weight is a matter of context. The application of a timely security selection criterion is conditional. Simply — it depends. For example, many researchers demonstrate that value, as a selection variable, is often conditional on the type of firm, other nonvalue factors, the investment horizon, or some other dimension. Sloan (2001) and Beneish et al. (2001) call this interdependency of security factors *contextual*.

Seasoned active managers know that value investing focuses on discovering cheap stocks with a balance of quality; at the same time, growth investing often seeks to balance positive momentum with quality and cheapness. This anecdotal assertion finds substantiation in prior academic studies. For example, Daniel and Titman (1999) find that momentum effects are stronger for growth stocks. Asness (1997) finds that value strategies work, in general, but less so for stocks with high momentum. In a particularly relevant study, Scott et al. (1999) focuses on prospect theory and investor overconfidence. They provide empirical evidence that rational value investors should emphasize cheapness (as in dogs), whereas growth investors should let winners run — with the prospect of future good news. Piotroski (2000) and Mohanram (2004) also demonstrate that one should focus on different sets of financial statement information when analyzing stocks with different book-to-price ratios. Taken together, these studies (and others) point to the importance of analyzing the efficacy of alpha factors within carefully selected security universes — the contextual analysis of active strategies.

9.2.1 Factor Categories

To illustrate contextual dynamics, we introduce five composite factors representing the set of investing philosophies discussed in Chapter 5. Table 9.1 describes the description of these composites. To capture the essence of the value investing that buys cheap stocks, we create the relative value (RV) factor, a composite encompassing two types of cheapness measures: the earnings yield and the asset value. We title this factor relative value because cheapness is gauged in the context of a peer group; and, in this study, we use sector as the peer group for comparison. Additionally, to represent the premise of the fundamental investing, we trace the analysis of the enterprise profitability, accrued to shareholders, into three composite factors: (1) the operating efficiency (OE) factor measuring management's ability to generate shareholder value, (2) the accounting accrual (AA) factor measuring the accuracy and the honesty of a company's financial reporting practice, and (3) the external financing (EF) factor measuring the hazard of self-serving management pursuing corporate expansions at the expense of shareholder wealth. Finally, the philosophy of riding market sentiment in momentum investing is captured in the momentum factor (MO), which consists of the measures of the intermediate-term price momentum, the earnings revision, and the earnings surprise.

TABLE 9.1 Definition of Factor Composites

Composite	Factors
Valuation (RV)	Book-to-price ratio
	Sales to enterprise value
	Earnings yield (historical)
	Earnings yield (IBES FY1)
	EBIT to enterprise value
Operating Efficiency (OE)	Increase in asset turnover ratio
	Level of operating leverage
	Cashflow-from-operation to sales
Accounting Accrual (AA)	Accounting accruals (balance sheet)
	Accounting accruals (cashflow statement)
External Financing (EF)	External financing to net operating assets
	Debt issuance to net operating assets
	Equity issuance to net operating assets
	Share count increase
Momentum (MO)	Six-month price momentum
	Nine-month earnings revision
	Earnings surprise score

Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

9.2.2 Security Contexts

We illustrate the interplay among factors along the dimensions of three risk characteristics: value, growth, and earning variability. Along each of these dimensions, we select two nonoverlapping security contexts with an equal number of stocks: one contains securities with high loadings of that risk characteristic, whereas the other includes securities with low loadings. Hence, six security contexts are defined, and they contain firms with high/low value measure, high/low growth rate, and high/low earnings variability.

We use the book-to-price ratio as our first risk dimension: value. The name value for the book-to-price ratio implies it associates with market inefficiency, but this is not relevant to the contextual analysis. What is relevant is the interpretation provided by Fama and French (1996), who associate the book-to-price ratio with the investment quality or financial condition of a company. Specifically, we can interpret a low book-to-price ratio as an indication of high quality and a high book-to-price ratio as low quality. Defined as such, high-quality companies are expected by investors to deliver superior returns on investment (ROI) and their *ex post* ROI typically exceeds the average ROI of a broad universe. In contrast, low-quality companies usually face a difficult operating environment and are expected to deliver inferior operating results. Different competitive standing, superior vs. inferior, often induces different challenges facing company management; one battles from a deteriorated competitive position to survive, whereas the other protects its competitive advantage by fending off competition. These intuitions are confirmed in the studies by Piotroski (2000) and Mohanram (2004). Therefore, we argue that investors should also focus their attentions on a different set of factors when evaluating the return appeal of companies with different book-to-price ratios.

Our second risk characteristic sorts companies based on their growth rate, creating contexts containing high-growth and low-growth companies. The rational behind this contextual dimension is well documented by Scott et al. (1999, 2003). Linking the behavioral science findings with the valuation theory, Scott et al. show that momentum investing (riding winners and looking for good news) is more important when selecting high-growth stocks, whereas selecting low-growth stocks should focus more on cheapness. The difference can be traced to how investors estimate the fair value of a business. The fair value estimate typically comprises two parts: the present value of existing business and the present value of future growth opportunities. For a low-growth company whose future growth prospect is limited, the value of its existing business dominates its fair value and, more importantly, valuation ratios (i.e., cash-flow yield or earnings yield) provide an accurate ranking of the relative cheapness of its existing business. In contrast, for high-growth companies, the majority of its fair value comes from the present value of future growth opportunities. As such, factors that are capable of predicting the quality of future growth play more prominent roles in determining the fair value. Combining this valuation reasoning with the observation that investors tend to under-react to news due to their overconfidence, Scott et al. (1999, 2003) show that earnings revision factor, a proxy of good news, is a consistent predictor of the excess returns of growth stocks.

Our last dimension differentiates companies along the earnings variability dimension. This contextual selection is inspired by the persistent predictability bias documented by Huberts and Fuller (1995). They show that sell-side analysts tend to provide overly optimistic forecasts for companies whose earnings are harder to predict, whereas their forecasts are more realistic, albeit still optimistic, for companies with stable earnings in the past. Das et al. (1998) provide a more rigorous examination of this phenomenon and derive the same conclusion. Lastly, Beckers et al. (2004) find the same bias in European analysts' forecasts. In all, if earnings forecasts are less trustworthy for companies whose earnings are more variable, it is our conjecture that investors should focus their attentions on the quality of earning and the competence of management to remedy the deficiency of earnings forecasts. Similarly, investors should rely more on analysts' forecasts when selecting stable-earning companies because these forecasts are more reliable.

9.3 MATHEMATICAL ANALYSIS OF CONTEXTUAL MODELING

The basic premise of contextual modeling is that the efficacies of alpha factors are different among stocks across the different contexts. By using different optimal weights across the contexts, we will achieve a higher overall information ratio.

9.3.1 A One-Factor Example

The following one-factor example provides some intuition to the approach. Suppose we have a single context that divides the stock universe into two halves: one high and one low. Let us also assume for the moment that we just have a single alpha factor. We are interested in how the factor performs overall if it performs differently in the two halves. According to Chapter 4, a single-period excess return is given by (Equation 4.19)

$$\alpha_t = \sum_{i=1}^{N} w_i r_i = \lambda^{-1} \sum_{i=1}^{N} F_i R_i \, , \tag{9.4}$$

where F_i is risk-adjusted forecast, R_i is the risk-adjusted return, N is the number of stocks, and λ is the risk-aversion parameter used to calibrate the portfolio to a targeted tracking error. Breaking the stock universe into two halves — high and low, according to the context — we rewrite (9.4) as

$$\alpha_t = \lambda^{-1} \sum_{i=1}^{N} F_i R_i = \lambda^{-1} \sum_{i \in H} F_i R + \lambda^{-1} \sum_{i \in L} F_i R \, . \tag{9.5}$$

Now, writing all three sums in terms of risk-adjusted ICs in the respective universe gives

$$N \cdot \mathrm{IC}\,\mathrm{dis}(\mathbf{F})\,\mathrm{dis}(\mathbf{R}) = \frac{N}{2} \times \mathrm{IC}_H \mathrm{dis}(\mathbf{F}_H)\,\mathrm{dis}(\mathbf{R}_H) + \frac{N}{2} \times \mathrm{IC}_L \mathrm{dis}(\mathbf{F}_L)\,\mathrm{dis}(\mathbf{R}_L). \tag{9.6}$$

For simplicity, we have omitted the subscript t. We shall assume all the dispersions of forecasts and return are the same, which leads to

$$\mathrm{IC} = \frac{1}{2} \cdot \mathrm{IC}_H + \frac{1}{2} \cdot \mathrm{IC}_L. \tag{9.7}$$

The overall IR is obtained by the ratio of average IC to the standard deviation of IC

$$\mathrm{IR} = \frac{\overline{\mathrm{IC}}_H + \overline{\mathrm{IC}}_L}{\sqrt{\sigma_H^2 + \sigma_L^2 + 2\rho_{H,L}\sigma_H\sigma_L}}. \tag{9.8}$$

Equation (9.8) gives the overall IR in terms of IC statistics in the high and low contexts.

Example 9.1

Suppose the factor only works in the high dimension, but not in the low dimension, i.e., $\overline{\mathrm{IC}}_L = 0$. Then

$$\mathrm{IR} = \frac{\overline{\mathrm{IC}}_H}{\sqrt{\sigma_H^2 + \sigma_L^2 + 2\rho_{H,L}\sigma_H\sigma_L}}. \tag{9.9}$$

If the correlation of ICs is not negative, this overall IR will be less than the IR of the factor in the high dimension alone, i.e.,

$$\mathrm{IR} < \mathrm{IR}_H = \frac{\overline{\mathrm{IC}}_H}{\sigma_H}. \tag{9.10}$$

For instance, if $\overline{\mathrm{IC}}_H = 0.1$, $\sigma_H = \sigma_L = 0.1$, $\rho_{H,L} = 0.2$, then the IR in the high dimension $\mathrm{IR}_H = 1$, but the overall IR is just 0.6.

This example illustrates the fact that when a factor does not add value in the low dimension, still using it would dilute the IR of the factor because it adds noise or risk without additional returns. The simple remedy for this problem is to not use the factor in the low dimension. In other words, we shall not take any exposure to the factor in the low dimension stock. In terms of factor weight, it is simply zero for low dimension stocks.

9.3.2 Optimal Factor Weights across the Context

Setting the factor to zero for the low dimension stocks in the previous example represents a simple solution, but it is not necessarily the optimal one. If we denote the factor weight by v_H and v_L in the high and low dimension, then the overall IR becomes

$$\mathrm{IR}=\frac{v_H\overline{\mathrm{IC}}_H+v_L\overline{\mathrm{IC}}_L}{\sqrt{v_H^2\sigma_H^2+v_L^2\sigma_L^2+2\rho_{H,L}\sigma_H\sigma_L v_H v_L}}. \tag{9.11}$$

The optimal weight can be found by the following

$$\begin{pmatrix} v_H^* \\ v_L^* \end{pmatrix} \propto \begin{pmatrix} \dfrac{\overline{\mathrm{IC}}_H}{\sigma_H^2}-\rho_{H,L}\dfrac{\overline{\mathrm{IC}}_L}{\sigma_H\sigma_L} \\ \dfrac{\overline{\mathrm{IC}}_L}{\sigma_L^2}-\rho_{H,L}\dfrac{\overline{\mathrm{IC}}_H}{\sigma_H\sigma_L} \end{pmatrix}. \tag{9.12}$$

With parameters in Example 9.1, the optimal weights are $v_H^*=125\%$ and $v_L^*=-25\%$. The optimal IR is at 1.02, slightly above the IR for the high dimension. Thus, the optimal weights would have us betting against the factor in the low dimension, not because of value-added (there is none since the average IC is zero), but because of reduced risk.

With multiple factors, the objective of contextual modeling is to maximize the overall IR with optimal weights of factors in high and low dimensions. There are M factors and the weights are $\mathbf{v}=\left(\mathbf{v}_H,\mathbf{v}_L\right)=\left(v_{1,H},v_{2,H},\cdots,v_{M,H},v_{1,L},v_{2,L},\cdots,v_{M,L}\right)'$. The vector of average IC is

$$\overline{\mathbf{IC}}=\left(\overline{\mathbf{IC}}_H,\overline{\mathbf{IC}}_L\right)=\left(\overline{IC}_{1,H},\overline{IC}_{2,H},\cdots,\overline{IC}_{M,H},\overline{IC}_{1,L},\overline{IC}_{2,L},\cdots,\overline{IC}_{M,L}\right)'$$

and the $2M\times 2M$ IC covariance matrix is $\mathbf{\Sigma}_{IC}$. The overall IR is given by

$$\mathrm{IR} = \frac{\mathbf{v}' \cdot \overline{\mathbf{IC}}}{\sqrt{\mathbf{v}' \cdot \mathbf{\Sigma}_{IC} \cdot \mathbf{v}}}. \tag{9.13}$$

The optimal weights are given by

$$\mathbf{v}^* \propto \mathbf{\Sigma}_{IC}^{-1} \cdot \overline{\mathbf{IC}}. \tag{9.14}$$

The proportional constant is determined by normalization of the weights.

9.4 EMPIRICAL EXAMINATION OF CONTEXTUAL APPROACH

In this section we present a series of empirical tests to illustrate the presence of contextual asset pricing. We use the Russell 1000 Index as the security universe, for the time period from December 1986 to September 2004. Data sources include (1) the Compustat quarterly database for financial characteristics; (2) the IBES US historical detail database for consensus earnings estimates; and (3) the BARRA US E3 database for price, return, and risk factor characteristics.

9.4.1 Risk-Adjusted ICs

We first compare the risk-adjusted ICs between sample partitions according to the BARRA definitions of value, growth, and earnings variability. Along these BARRA risk dimensions, we compare the average and the variance of IC, pertaining to the high and low security contexts, for each of the selected composite alpha factors.

Table 9.2 presents these comparisons (15 in all — 3 risk dimensions and 5 alpha measures). We calculate the two-sample *t*-test for the mean difference and the F-test for the variance difference. In Panel A, the return profile of the EF factor is significantly different between high- and low-value stocks. Both the two-sample *t*-test and the F-test are significant at 1% level. For low-value (low book-to-price ratio) stocks the IC is .015, as contrasted with an IC of .044 for high-value stocks. This demonstrates that the way the external financing factor is priced is indeed contextual dependent — more important for discounted firms than high-priced ones. (Note that discounted firm means high value, and high-priced firm refers to low value.) External financing costs and expected investment returns contribute to this contextual dependency. Dilution of shareholder wealth

TABLE 9.2 Comparison of Risk–Adjusted ICs in Different Risk Dimensions

Panel A Value Dimension

	Mean		**STD**		**Two-Sample t Test**		**F Test**			
	High	**Low**	**High**	**Low**	**t**	**p-Value**	**F**	**pval**	**df (num)**	**df (denom)**
RV	0.022	0.022	0.069	0.079	0.011	0.991	0.764	0.270	68	68
OE	0.032	0.040	0.047	0.037	−1.050	0.296	1.613	0.051	68	68
AA	0.027	0.042	0.043	0.050	−1.912	0.058	0.720	0.177	68	68
EF	0.044	0.015	0.041	0.057	3.460	0.001	0.504	0.005	68	68
MO	0.031	0.049	0.061	0.072	−1.577	0.117	0.711	0.163	68	68

Panel B Growth Dimension

	Mean		**STD**		**Two-Sample t Test**		**F Test**			
	High	**Low**	**High**	**Low**	**t**	**p-Value**	**F**	**pval**	**df (num)**	**df (denom)**
RV	0.003	0.034	0.113	0.062	−2.046	0.043	3.318	0.000	68	68
OE	0.061	0.019	0.043	0.042	5.702	0.000	1.037	0.883	68	68
AA	0.044	0.022	0.060	0.039	2.461	0.015	2.450	0.000	68	68
EF	0.028	0.017	0.054	0.043	1.274	0.205	1.567	0.066	68	68
MO	0.059	0.023	0.092	0.072	2.571	0.011	1.623	0.048	68	68

Panel C Variability Dimension

	Mean		**STD**		**Two-Sample t Test**		**F Test**			
	High	**Low**	**High**	**Low**	**t**	**p-Value**	**F**	**pval**	**df (num)**	**df (denom)**
RV	0.023	0.023	0.105	0.076	−0.025	0.980	1.911	0.008	68	68
OE	0.045	0.029	0.051	0.039	2.019	0.046	1.678	0.034	68	68
AA	0.033	0.032	0.049	0.036	0.151	0.880	1.848	0.012	68	68
EF	0.038	0.018	0.055	0.045	2.343	0.021	1.492	0.101	68	68
MO	0.034	0.038	0.094	0.074	−0.252	0.802	1.605	0.053	68	68

Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

is most likely to occur when the invested firm is traded at a discount and starts pursuing capital increases through external financing, because the proceed not only costs more to obtain but also generates lower returns to existing shareholders.

Panel B shows that investors reward high-growth companies for conservative accounting (AA), high OE, and better price and earnings performance (MO). In contrast, cheapness of share price (RV) is an important return driver for low-growth companies, with both the average and the standard deviation of ICs significantly different at 5% level when compared with high-growth companies. Our empirical results are consistent with the ones documented by Scott et al. (1999); and, in addition, we highlight the importance of conservative accounting and operating efficiency as important return drivers for high-growth companies. Consistent with Asness (1997), we find the average IC of momentum factor (MO) in the high-growth stocks is more than twice the size of the average in the low-growth stocks.

Panel C focuses on the earnings variability dimension. Operating efficiency (OE) and EF factors are more indicative of the future stock returns of companies with variable earnings, as shown in their two-sample *t*-tests, which are significant at a 5% level. On the other hand, RV and AA have almost identical average IC across the partitions. However, their standard deviations of ICs, the risk endogenous to the active strategies of applying RV and AA, are significantly different.

To summarize, Table 9.2 is generally consistent with the theory of rational pricing that is conditional. Using univariate average IC comparisons over the 1986–2003 period, we find that the market is more responsive to operating efficiency, conservative accounting, and positive earnings evidence when dealing with high-growth and/or high-priced firms than is the case with low growers. The market is much more focused on operating performance and shareholder-friendly managements when growth is at stake, and much less focused on cheapness of stock prices. Surveying the differences in IC averages and IC standard deviation across the three risk partitions, it appears that the growth dimension induces the most contextual difference, whereas the variability dimension induces the least.

9.4.2 IC Correlations

Table 9.3 reports the IC correlation matrices among the five composite factors in each of the six risk partitions. In each case, the numbers before and after the slash sign are correlations for higher (lower) partitions. Before we comment on the correlation difference across contexts, some

TABLE 9.3 Correlations of Risk-Adjusted ICs

Panel A Value Dimension

	OE	AA	EF	MO
RV	0.28/0.16	−0.22/0.21	−0.08/0.63	−0.11/−0.44
OE		0.42/0.50	0.16/0.24	0.24/0.19
AA			0.21/0.09	0.17/0.14
EF				0.18/−0.23

Panel B Growth Dimension

	OE	AA	EF	MO
RV	−0.22/0.19	0.14/−0.08	0.45/−0.08	−0.71/−0.25
OE		0.36/0.25	0.16/0.27	0.28/0.21
AA			0.23/0.21	−0.18/0.01
EF				−0.32/0.26

Panel C Variability Dimension

	OE	AA	EF	MO
RV	−0.16/0.12	−0.18/0.19	0.19/0.29	−0.60/−0.38
OE		0.30/0.37	0.26/0.38	0.48/0.10
AA			0.28/0.19	0.19/0.04
EF				0.05/−0.23

Note: In each cell, the number before the slash shows correlation of the high context and the number after the slash displays correlation for the low context.

Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

general patterns are worth noting. First, the IC correlation between RV and momentum (MO) is always negative, providing diversification benefit to an active strategy by including both factors. Second, the correlations among the three composite factors from the same quality category, i.e., OE, AA, and EF, are not only all positive in general, but they seem to be rather stable across the risk partitions. Third, the relative value factor tends to have small and often negative correlations with other factors. In all, the market generally prices quality and momentum concurrently, while rotating between cheapness and momentum, each at the expense of the other, due to perhaps changes in risk aversion.

Panel A compares the two correlation matrices derived from the high and low value contexts. The correlations between RV and AA and between RV and EF show the biggest differences. In high-value stocks, the two correlations are −0.22 and −0.08, respectively, whereas in low-value stocks the two correlations are considerably higher at 0.21 and 0.63, respectively. The other notable difference is the correlation between MO and EF. It is 0.18 in high-value stocks and −0.23 in low-value stocks. Along the growth dimension (Panel B), again the relative value causes most of the correlation differences. Its correlations with OE, AA, and EF all flip signs across the partition. The correlation between RV and MO is negative in both partitions, but it is remarkably low at −0.71 among high-growth stocks. Along the variability dimension (Panel C), the differences in correlation coefficients are smaller compared to those in Panel A and B. In aggregate, MO has lower correlation with other factors in low-variability stocks than in high-variability stocks.

9.4.3 Optimal Factor Weights and Their Differences

In this section, we solve for the optimal weights of the composite alpha factor using the IR maximization framework outlined in Chapter 7. We shall refer to a combination of alpha factors as an alpha model. In each of the six risk partitions, we find the optimal weights of the five composite factors using the IC averages and IC covariances over the whole sample period. Based on the differences of these inputs shown in Table 9.2 and Table 9.3, we naturally expect different alpha models in each high/low risk partition. However, are these weight differences statistically significant? We devise several ways to answer this question. In this section, we perform several direct tests on the optimal weights themselves. Later, we test the performance differences induced by weighting differences, focusing on their alpha-producing capabilities.

To test the statistical significance of the difference between the optimal weights, we adopt a bootstrapping procedure as follows, similar to the one introduced by Michaud (1998). We resample with replacement the historical ICs, jointly for all five composite alpha factors in each of the six security contexts. Similar to a bootstrapping procedure, we make the sample size the same as the number of time periods in the original sample. In each sample, we then calculate the average ICs and IC covariances of five factors along the different risk partitions, and derive IR-maximizing optimal weights. This is repeated one thousand times to obtain one thousand sets of optimal weight in each risk partition. By introducing sampling errors into the average ICs and the IC covariances, we translate the sampling errors of historical ICs into the sampling errors of model weighting. We deem a weight deviation significant if its magnitude is significantly larger than the sampling error.

TABLE 9.4 Resample Weights Comparison in Different Risk Dimensions

Panel A Value Dimension

	Mean		STD		Difference (High–Low)		
	High	Low	High	Low	Avg/Stdr	Avg	Stdr
RV	9.0	6.3	4.0	3.5	0.5	2.6	5.3
OE	16.7	46.4	6.0	8.9	–2.7	–29.7	10.8
AA	20.4	24.4	6.2	6.5	–0.4	–4.0	9.0
EF	43.0	5.1	7.9	4.8	4.1	37.9	9.3
MO	11.0	17.8	4.8	5.1	–1.0	–6.8	7.1

Panel B Growth Dimension

	Mean		STD		Difference (High–Low)		
	High	Low	High	Low	Avg/Stdr	Avg	Stdr
RV	3.7	22.8	2.4	7.3	–2.5	–19.1	7.6
OE	52.7	16.9	7.8	8.3	3.1	35.8	11.7
AA	16.7	33.3	5.0	8.8	–1.6	–16.6	10.1
EF	14.0	16.7	5.9	7.2	–0.3	–2.7	9.3
MO	12.9	10.3	4.0	5.0	0.4	2.6	6.3

Panel C Variability Dimension

	Mean		STD		Difference (High–Low)		
	High	Low	High	Low	Avg/Stdr	Avg	Stdr
RV	7.9	7.2	3.8	4.5	0.1	0.7	5.9
OE	36.1	27.0	7.4	6.5	0.9	9.1	10.0
AA	27.2	41.1	6.3	7.5	–1.4	–13.9	9.6
EF	22.5	10.5	6.6	5.1	1.4	12.0	8.4
MO	6.4	14.2	3.7	4.4	–1.4	–7.9	5.7

Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

The model weights can be compared individually for each of five factors or jointly for all five factors together. For individual comparison, Table 9.4 shows the average and the standard error of factor weights of 1000 bootstrapping samples, again across the 15 samples — 3 risk factor partitions and 5 alpha factors. We also show the difference in optimal weights across the three risk dimensions, in terms of average, standard error, and their ratio. This ratio can be similarly interpreted as a t-statistic, with a value of above 2 or below –2 indicating statistical significance in mean difference. The results in Table 9.4 are consistent with our interpretation of the univariate IC tests and correlation differences shown earlier. Note the following remarks:

- First, model weights of the high-growth context (Panel B) and the low-value context (Panel A) are remarkably similar. Perhaps, this points to a set of common challenges facing high-priced and high-growth firms, the most prominent of which is to maintain superior operating results captured by the OE factor. However, we note the reverse inference does not apply — model weights in the high-value and the low-growth contexts are quite different. In the high-value context, the most prominent weight (43%) is in EF factor, whereas in the low-growth context, the model weights are relatively equitable for all five factors. Note the relative value (RV) is weighted 23% here, whereas it never receives more than 10% elsewhere.

- Second, we notice that in the growth dimension (Panel B), whereas the RV factor's weight is substantially higher in the low-growth dimension than in the high-growth dimension, with a mean-standard error ratio of –2.5, consistent with the results by Scott et al. (1999); the MO factor's weight is only slightly higher in the high-growth half (12.9%) than in the lower half (10.3%). The reason for this is the higher strategy risk of the MO factor in the high-growth context (Table 9.2, Panel B) than in its counterpart in the low-growth context.

- Table 9.4 unveils primary return drivers for each security context, should they exist. To facilitate the discussion, let's delineate primary drivers as factors that are more than 40% of a model. Contextual partitioning plays a significant role in governing the primary return driver, as it shifts from OE for both high-priced and high-growth firms, to conservative EF for discounted firms and to honest management, gauged by conservative earnings reporting practice (AA), for firms with stable earning stream. These contextual dynamics further highlight the descriptive inadequacy of the one-size-fits-all assumption of traditional quantitative models.

- Across both the value and growth dimensions, there are two factors with significant weights, OE and EF in value and RV and OE in growth. However, across the variability dimension, none of the factors show significant weight difference.

Finally, we note the aggregated weight in the corporate quality category, i.e., the sum of weights in OE, AA, and EF accounts for over 70% of the model weight in almost all cases. This confirms the importance of financial statement analysis in active equity management.

9.4.4 Model Distance

Table 9.5 tests for significance in differences between the optimal weights jointly. For comparison, we first construct a static one-size-fits-all model without any contextual partitioning, using the same resampling procedure. The first row of Panel A shows the resampled efficient weights for this static model and the rest of Panel A show the weights from the previous section.

To compare the factor weights jointly, we employ two measures. The first measure is the distance between two models, defined as

$$d = \sqrt{\frac{\Delta \mathbf{w}' \cdot \Delta \mathbf{w}}{k}}, \tag{9.15}$$

where $\Delta \mathbf{w}$ is the difference in model weights, and k equals five, the number of factors in the model. It is the root mean square of the optimal weight differences. Panel B of Table 9.5 displays the distances between different pairs of models. Several interesting observations are worth noting. First, the static model is most similar to the high-variability contextual model and most dissimilar to the high-value contextual model. Second, when comparing the two contextual models pertaining to same risk dimension, the value dimension has the highest model distance followed by the growth dimension, whereas variability dimension has the smallest distance. Third, consistent with the observation above, the distance between the high-growth model and the low-value model is also very low.

Whereas the distance measure does not incorporate the sample error, our second measure does. Panel C and D of Table 9.5 provide the chi-square statistics between models and their p-value. Note the statistics are not symmetric, as we are testing whether the mean of the resampled weights of one model belongs to the ensemble of the resampled weights of another model. When the models are interchanged, the ensemble is also changed, resulting in a different chi-square statistic. (See Appendix A9.1 for a detailed technical note.) Panel D unveils three interesting findings. First, as shown on the first row (and the first column), the static model

TABLE 9.5 Pairwise Model Weight Comparison

Panel A: Model Weights of Resample Efficient Portfolios

		RV	OE	AA	EF	MO
One-size	R1000	2.5	41.6	36.3	13.0	6.5
Value	High	9.0	16.7	20.4	43.0	11.0
	Low	6.3	46.4	24.4	5.1	17.8
Growth	High	3.7	52.7	16.7	14.0	12.9
	Low	22.8	16.9	33.3	16.7	10.3
Variability	High	7.9	36.1	27.2	22.5	6.4
	Low	7.2	27.0	41.1	10.5	14.2

Panel B: Model Distance

		One-size	Value		Growth		Variable	
		R1000	High	Low	High	Low	High	Low
One-size	R1000	0.0	21.2	9.4	11.7	12.7	7.1	8.7
Value	High	21.2	0.0	24.4	23.2	14.6	14.7	20.0
	Low	9.4	24.4	0.0	7.1	16.9	11.7	13.2
Growth	High	11.7	23.2	7.1	0.0	19.8	11.2	17.8
	Low	12.7	14.6	16.9	19.8	0.0	10.7	7.4
Variability	High	7.1	14.7	11.7	11.2	10.7	0.0	11.0
	Low	8.7	20.0	13.2	17.8	7.4	11.0	0.0

Panel C: Chi-Squared Statistics

		One-size	Value		Growth		Variable	
		R1000	High	Low	High	Low	High	Low
One-size	R1000	0.0	31.8	13.2	19.8	13.5	5.6	7.7
Value	High	69.0	0.0	65.6	39.1	15.9	13.8	49.0
	Low	32.0	36.2	0.0	5.2	21.6	17.0	11.1
Growth	High	16.6	39.7	5.0	0.0	24.9	13.1	19.4
	Low	73.7	18.9	34.0	74.7	0.0	24.2	18.3
Variability	High	11.9	13.7	17.7	14.2	8.6	0.0	9.7
	Low	17.0	23.2	9.9	24.8	7.7	14.0	0.0

Panel D: p-Value of Chi-Squared Test

		One-size	Value		Growth		Variable	
		R1000	High	Low	High	Low	High	Low
One-size	R1000	1.000	0.000	0.010	0.001	0.009	0.235	0.103
Value	High	0.000	1.000	0.000	0.000	0.003	0.008	0.000
	Low	0.000	0.000	1.000	0.264	0.000	0.002	0.026
Growth	High	0.002	0.000	0.282	1.000	0.000	0.011	0.001
	Low	0.000	0.001	0.000	0.000	1.000	0.000	0.001
Variability	High	0.018	0.008	0.001	0.007	0.072	1.000	0.045
	Low	0.002	0.000	0.041	0.000	0.102	0.007	1.000

Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

is statistically different from the contextual models on the growth and the value dimensions at a 5% level. However, contextual models along the variability dimension are not statistically different from the static one. Second, when comparing model weights of the high and low contexts for each risk dimension, value and growth dimensions exhibit significant differences, whereas the variability dimension is questionable. Third, further substantiating the observation, shown in Table 9.3, that the high-growth model is similar to the low-value model, the p-value is either 0.28 when using the covariance from the low-value context or 0.26 when testing with the high-growth covariance; neither is significant.

9.4.5 Contextual Alpha Model

The results of the previous section confirm the benefits of the contextual approach in building quantitative alpha models, and part of the results concerning the value and growth dimensions should be applicable to portfolio mandates with styled benchmarks, as our partitions along these dimensions are partly consistent with how many styled benchmarks are defined. However, what about mandates with core benchmarks? In particular, can we build a contextual model based on our analysis that beats the one-size-fits-all model? In this section, we propose an approach in which factor weightings are dynamically selected and conditioned on the risk characteristics. Then, we compare the performance between contextual models constructed with this approach and the static model. As these models employ the same set of factors, this comparison provides some insight into added value of dynamic factor weightings.

To further illustrate the relevance of each risk dimension, we implement four variants of contextual model, named value, growth, variability, and comprehensive. The first three models are built with a single risk dimension (two security contexts) indicated by their names. For example, the growth contextual model derives its dynamic factor weightings from the high-growth and the low-growth contexts only. In a nutshell, the factor weighting for a particular stock is a linear combination of high-growth and low-growth model, and relative weights of the combination are determined by the stock's growth rate. The comprehensive contextual model takes into account all three contextual dimensions, thus generating return forecasts based on optimal weights from all six security contexts.

To provide a more efficient use of our limited data sample and to facilitate a fair performance comparison, we employ the cross-validation procedure. Specifically, we first divide our sample periods into ten subperiods chronologically with equal duration. We then elect one of the subperiods as the out-of-sample period, and the remaining nine subperiods become the in-sample period. Although efficient model weights (for both the static and contextual models) are estimated in the in-sample period through our IR optimization framework, the scores (forecasts) are computed based on the estimated factor weights for the out-of-sample periods wherein the model performance is also computed. This exercise is repeated ten times for each of the ten subperiods, whose out-of-sample results are then stringed together to calculate performance statistics. Although we realize this approach creates chronological inconsistency in terms of the sequencing of the in-sample, out-of-sample periods, it is free of potential bias caused by a particular choice of in-sample, out-of-sample periods.

9.5 PERFORMANCE OF CONTEXTUAL MODELS

9.5.1 Risk-Adjusted Portfolios

Table 9.6 compares model efficacy in terms of the excess returns generated by dollar-neutral portfolios, a comparison that incorporates realistic portfolio optimization constraints. Rebalanced on a quarterly basis, portfolios

TABLE 9.6 Performance Comparison of Optimal Dollar-Neutral Portfolios

Panel A: Model Performance

	Static	**Value**	**Growth**	**Variable**	**Comparison**
Alpha	7.41%	8.53%	8.54%	7.95%	8.57%
IR	1.56	1.63	1.66	1.54	1.72

Panel B: Pairwise Performance Comparison

	Static	**Value**	**Growth**	**Variable**	**Comparison**
Static		−1.13%	−1.13%	−0.54%	−1.16%
		(**−4.39)	(**−4.75)	(**−3.64)	(**−6.06)
Value	1.13%		0.00%	0.58%	−0.03%
	(**4.39)		(−0.02)	(*2.45)	(−0.23)
Growth	1.13%	0.00%		0.59%	−0.03%
	(**4.75)	(0.02)		(**3.34)	(−0.19)
Variability	0.54%	−0.58%	−0.59%		−0.62%
	(**3.64)	(*−2.45)	(**−3.34)		(**−4.46)
Comp.	1.16%	0.03%	0.03%	0.62%	
	(**6.06)	(0.23)	(0.19)	(**4.46)	

Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

are formed for each model aiming at the highest model score exposures, given that their annualized tracking error is targeted at 5% and they have no exposure to market beta and size. Panel A shows the excess return and IR of each model on an annual basis. Whereas the static model has the lowest excess return and the comprehensive model produces the highest excess return and IR, all models generate excellent performance.

We also compare model performance in a pairwise manner with the average and the t-statistic of performance differences through time. Specifically, each cell in Panel B represents the excess performance between the "active" model indicated by the row title and the "benchmark" model indicated by the column title. As shown on the first column of Panel B, contextual modeling enhances portfolio returns when compared to the static model. The enhancement of quarterly returns ranges from 1.16 to 0.54%. According to the t-statistic (number in parentheses), the comprehensive contextual model provides the most consistent out-performance with a t-statistic of 6.06, followed by the growth contextual model with a t-statistic of 4.75. Also worth noting is the observation that incorporating either the value or the growth dimension captures a significant portion of performance improvement, as the comprehensive implementation only outperforms both models by 3 bps annually, shown on the last row. Lastly, the superior *ex post* performance, delivered by the value and growth models, underscores the importance of the model distance test, which indicates a significant difference vs. the static model for models along the value and the growth dimensions, but not for the variability dimension. Perhaps, the model distance test provides a pathway of selecting contextual models that are likely to deliver better *ex post* returns.

9.5.2 Asset Pricing Tests (Fama–MacBeth Regression)

Table 9.7 documents the advantage of using contextual modeling from the asset pricing perspective. That is, incorporating contextual dependencies provides a better, more accurate description of how stocks are priced. Following the commonly accepted analytical framework employed by asset pricing studies, we apply the Fama–MacBeth regression to estimated returns to model scores through time on a quarterly basis.

Panel A answers the question as to whether contextual models contain relevant asset pricing information that is not captured by the static score. In this test, the dependent variable is a 3-month forward return, and the explanatory variables are beta, size, the static model score, and the residual contextual score (the contextual score netted out the static score). The

TABLE 9.7 Fama–MacBeth Regression Test

Panel A: Residual Contextual Scores vs. the Static Score

	Beta	Size	Static	Residual Comparison	Residual Value	Residual Growth	Residual Variability
Comprehensive	−0.262	−0.035	1.650	1.046			
	(−0.3)	(−0.1)	(12.9)	(6.8)			
Value	−0.288	−0.069	1.649		0.937		
	(−0.3)	(−0.2)	(13.0)		(6.6)		
Growth	−0.262	−0.018	1.653			0.970	
	(−0.3)	(−0.1)	(12.9)			(5.8)	
Variability	−0.223	−0.023	1.661				0.773
	(−0.2)	(−0.1)	(13.0)				(4.5)

Panel B: The Residual Static Score vs. Contextual Scores

	Beta	Size	Residual Static	Comparison	Value	Growth	Variability
Comprehensive	−0.263	−0.035	−0.559	1.915			
	(−0.3)	(−0.1)	(−3.8)	(14.2)			
Value	−0.287	−0.068	−0.274		1.913		
	(−0.3)	(−0.2)	(−2.1)		(13.1)		
Growth	−0.262	−0.018	−0.400			1.913	
	(−0.3)	(−0.1)	(−2.5)			(13.9)	
Variability	−0.224	−0.023	−0.445				1.797
	(−0.2)	(−0.1)	(−2.7)				(12.9)

Note: ( ) contains t-statistic.
Source: From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.

netting out allows for an orthogonal treatment, which distills the portion of asset pricing information exclusively contained in the contextual score, thus providing a measure that isolates the incremental value added by the contextual modeling . As shown in Panel A, the residual score of the comprehensive contextual model does indeed capture additional asset pricing information and its t-statistic is 6.8. Similar results are also found when the three risk-dimension specific models are tested and their t-statistics range from 6.6 to 4.5 — all significant at a 1% level.

Panel B shows the result of a complementary question to the one answered by Panel A. Is the static model statistically dominated by contextual models in the asset pricing test? In other words, does the static score add value when orthogonalized by contextual scores? To answer this question, we include the residual of static score and contextual scores in this set of Fama-MacBeth regressions. The residual score is computed by stripping the portion of variance of the static score that can be explained by the contextual score through OLS regression, the same procedure used in tests shown in Panel A. As shown in Panel B, the contextual score does provide return forecasts that dominate the forecasts of the static model statistically; and the return to the static score residual is not only negative but also statistically significant with a t-statistic of –3.8. Again, similar results are also found in tests of the three risk-dimension specific scores. The *t*-statistics in these three tests range from –2.1 to –2.7.

9.6 SECTOR VS. CONTEXTUAL MODELING

An alternative way to accommodate different sets of return drivers for each security is sector-based alpha modeling. This approach is fairly popular among quantitative practitioners, and it calls for a unique model for each sector, an approach that bears a strong resemblance to how fundamental research is typically organized in investment firms. A sector-oriented fundamental research makes intuitive sense. For fundamental research, it is more cost efficient to have fundamental analysts act as sector specialists who cover companies with similar business dynamics, as opposed to generalists who need to be experts in the full range of business models. Given that human mental capacity is limited, sector specialists should have a better chance of correctly processing categorically similar information. In comparison, when generalists face the challenge of reconciling a diverse spectrum of information, the ability to process it well is only reserved for the most experienced.

However, it is ambiguous why market inefficiencies should differ across sectors in general, simply because their business economics are different. In other words, it is hard to find a conjecture supporting the reason why investors' over- or underreaction to market information should differ for a car company when compared with a computer manufacturer.

On the other hand, some sectors are indeed different due to reasons related to regulation or significantly different business models. They confront company management with different challenges to add shareholder value, and perhaps warrant a separate model. In the U.S., for example, there are three broad sector categories: utilities, financials, and industrials. The industrial sector is a catch-all sector, which includes companies not belonging to either utility or financial sectors. Similar traits are shared among industrials companies.

Competitiveness: They belong to competitive industries wherein companies compete for business and to generate shareholder value.

Business economics: They share similar business economics. Goods are manufactured and services are rendered. A company's ability to create shareholder value depends on (1) its value add in *the value chain* and (2) the company's competitive standing to retain a portion of the added value.

Management challenges: To be successful, company management teams face similar challenges and engage in similar activities: working capital management, capital allocation decision, corporate financing activities, and business operation enhancement.

In contrast, the utility sector is primarily a regulated, cost-plus industry wherein company profits are both protected as well as capped by governmental regulations. As a result, operating efficiency loses its relevance in determining how competitive a company is. Capital allocation decisions are legislation driven rather than market driven.

The reason why the financial sector deserves a separate model is because of the significance of interest rates. As a result, many alpha factors that are relevant for industrial companies lose their meanings for the financial sector. For example, working capital is not relavent not only because financial companies do not produce inventories, but also because cash is part of the operating assets as cash is interest bearing. It is also an appealing proposition to model financial companies on the industry level — banks, life insurance, property and casualty, real estate investment trust (REIT), and diversified financials (such as brokers and investment managers). Many ratios are only meaningful for one particular financial industry, but not for others. For example, loan loss provision is a relevent matrix for banks, combined ratio is for insurance companies, and funds from operations (FFO) is for REITs.

Therefore, to isolate the appropriate return drivers and to achieve a more efficient forecast, quantitative alpha models should incorporate both

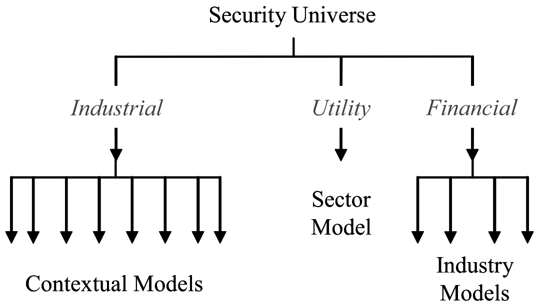

FIGURE 9.1. Modeling hierarchy.

contextual and sector modeling techniques. Figure 9.1 shows a modeling hierarchy that combines both sector modeling and contextual modeling techniques. There are two hierarchical levels: sector modeling being the first level and contextual modeling being the second. On the first level, a cross-section of securities is partitioned into three nonoverlapping sectors: industrial, financial, and utility. Within financial, securities are modeled on the industry level to reflect differences in business operations. Contextual modeling resides on the second level for industrial firms and forms overlapping contextual partitions to capture return idiosyncrasies rooted in behavioral differences. Note the following remarks:

- The combination of contextual and sector modeling enhances quantitative models with greater forecast accuracies (greater conviction in forecasts), a trait typically reserved for fundamental managers. Similar to fundamental research, these advanced forecasting techniques first categories companies based on their business environment and firm characteristics and then applies a set of relevant models to forecast their future returns individually. In doing so, a unique model is tailored for each security whose firm characteristics dictate each individual customization.

- Contextual modeling is a dynamic process over time and adapts to the progression of a company's life cycle. For example, many of today's successful firms (such as Microsoft) were very different a decade ago in terms of their firm characteristics, such as expected growth rate, value ratios, or earnings stability. As a firm evolves through time, its characteristics change and contextual approach adapts to this change by applying different models in forecasting the same security through time.

engage in these projects. On the other hand, companies without worthwhile projects should not spend at all, because spending CAPEX simply wastes shareholders' capitals. There are other links, such as future growth prospects or the cost of equity. For the interest of this section, we will use ROE as the link.

We now discuss each approach in detail.

Quadratic models: Here, we simply add a second-order term of the original factor to the linear model. In the case of a single factor, the model is

$$r = v_0 + v_1 F + v_2 F^2 + \varepsilon . \tag{9.16}$$

Combining a quadratic term with its linear counterpart can provide a better fit to a return response that exhibits nonlinear behavior. The shape of the function (9.16) depends on the signs of coefficients. Assume the coefficient of the linear term is positive. Then, the shape is concave if $v_2 < 0$ and convex if $v_2 > 0$. To model the CAPEX factor, we would have $v_2 < 0$. The expected return increases with the factor, reaches the maximum at $F = -v_1/2v_2$ and declines as the factor increases further. Companies with extremely high or low capital expenditures do not represent quality firms, whereas companies with reasonable, conservative capital expenditures do.

Conditional models: We can use another variable to partition the estimation universe into subgroups and construct linear models in each subgroup. In the case of CAPEX, we use ROE as the conditioning variable and create a dummy d_{high_roe}, which is binary –1 for companies with high historical ROE and 0 for companies with low historical ROE. Equation 9.17 isolates the dynamics of how CAPEX is priced for companies with high-ROE projects or those without.

$$r = v_0 + v_1 F_{capex} + v_2 d_{high\_roe} F_{capex} + \varepsilon . \tag{9.17}$$

For low-ROE companies, the model coefficient is v_1 and for high-ROE companies, the model coefficient is $v_1 + v_2$.

Interaction models: One can also use ROE together with CAPEX as an interaction term, i.e., the product of the two. Equation 9.18 shows a model of both ROE and CAPEX and their interaction. The interaction term captures the nonlinear effect. Assuming the coefficient v_3

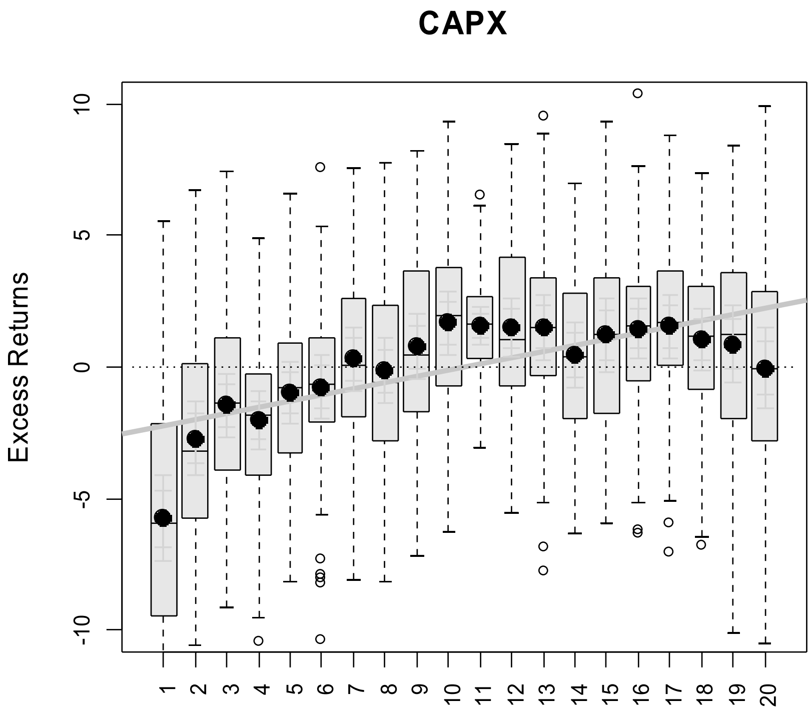

FIGURE 9.2. Fractile backtest of capital expenditure.

9.7.2 Nonlinear Effect Models

There are many ways to capture nonlinear effects. One simple way is to model the expected return using a polynomial by adding quadratic and even cubic terms of the factor values. The end result is still a linear model but with nonlinear factors. This approach is straightforward and flexible, but it often lacks economic intuition. With sufficient data mining, one runs the risk of finding a relationship that is statistically significant, but nonetheless spurious.

A better approach is to condition the factor value on other company attributes. In the case of CAPEX, we ask "What is the appropriate functional form that associates CAPEX with future security returns?" To answer this question, we go back to one of the primary philosophies outlined in Chapter 6. That is, we purchase quality companies that are expected to create shareholder value in the future. How does CAPEX relate to shareholder value generation? One of the important links between CAPEX and shareholder value is the expected ROE. Should a company have worthwhile projects (high-ROE projects), it is shareholder value enhancing to engage in these projects. On the other hand, companies without worthwhile projects should not spend at all, because spending CAPEX simply wastes shareholders' capitals. There are other links, such as future growth prospects or the cost of equity. For the interest of this section, we will use ROE as the link.

We now discuss each approach in detail.

Quadratic models: Here, we simply add a second-order term of the original factor to the linear model. In the case of a single factor, the model is

$$r = v_0 + v_1 F + v_2 F^2 + \varepsilon . \tag{9.16}$$

Combining a quadratic term with its linear counterpart can provide a better fit to a return response that exhibits nonlinear behavior. The shape of the function (9.16) depends on the signs of coefficients. Assume the coefficient of the linear term is positive. Then, the shape is concave if $v_2 < 0$ and convex if $v_2 > 0$. To model the CAPEX factor, we would have $v_2 < 0$. The expected return increases with the factor, reaches the maximum at $F = -v_1/2v_2$ and declines as the factor increases further. Companies with extremely high or low capital expenditures do not represent quality firms, whereas companies with reasonable, conservative capital expenditures do.

Conditional models: We can use another variable to partition the estimation universe into subgroups and construct linear models in each subgroup. In the case of CAPEX, we use ROE as the conditioning variable and create a dummy d_{high_roe}, which is binary –1 for companies with high historical ROE and 0 for companies with low historical ROE. Equation 9.17 isolates the dynamics of how CAPEX is priced for companies with high-ROE projects or those without.

$$r = v_0 + v_1 F_{capex} + v_2 d_{high\_roe} F_{capex} + \varepsilon . \tag{9.17}$$

For low-ROE companies, the model coefficient is v_1 and for high-ROE companies, the model coefficient is $v_1 + v_2$.

Interaction models: One can also use ROE together with CAPEX as an interaction term, i.e., the product of the two. Equation 9.18 shows a model of both ROE and CAPEX and their interaction. The interaction term captures the nonlinear effect. Assuming the coefficient v_3

is positive, the expected return is high for companies with high ROE and high CAPEX, and also for companies with low ROE and low CAPEX. However, the expected return is low for companies with high ROE and low CAPEX, and companies with low ROE and high CAPEX.

$$r = v_0 + v_1 F_{roe} + v_2 F_{capex} + v_3 F_{roe} F_{capex} + \varepsilon\,. \tag{9.18}$$

In general, it is common to see interaction variables in valuation-based factor return estimation, as valuation theory suggests that growth rate, return on invested capital, and cost of capital interact in product terms as well as their linear forms.

9.7.3 Linking CAPEX to Shareholder Value Creation

We combine quadratic and conditional models together to link capital expenditures and shareholder value creation. Specifically, Equation 9.8 shows a functional form that associates CAPEX and ROE with expected value creation and future return forecast.

$$r = v_0 + (v_1 F_{capex} + v_2 F_{capex}^2) + d_{high\_roe}(v_3 F_{capex} + v_4 F_{capex}^2) + \varepsilon\,. \tag{9.19}$$

Figure 9.3 shows the empirical estimation and compares the original CAPEX score (shown horizontally) with the transformed one (shown vertically). Because the universe is broken into high- and low-ROE companies,

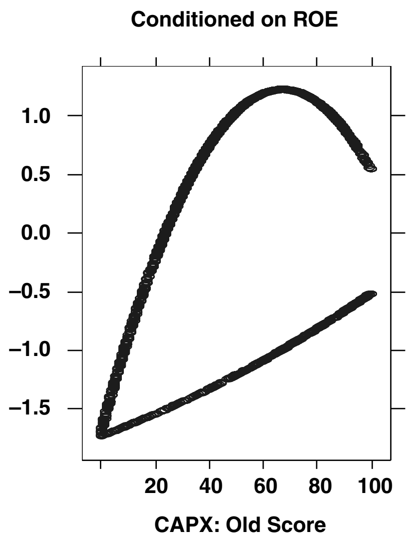

FIGURE 9.3. Transformation of the CAPEX factor.

two fitted lines are shown. The lower one represents low-ROE firms, whereas the upper one represents high-ROE firms. Obviously, high-ROE firms deliver higher returns than low-ROE firms. It is interesting to note that for firms without worthwhile projects, the return response is fairly linear. That is, lower (or even no) capital expenditures bode well, indeed, for low-ROE firms, as they will most likely waste shareholder capital. On the other hand, the return response for high-ROE firms is an upward-sloping, concave curve. The best firms are those who have high-ROE projects and spend conservatively on capital expenditures.

9.7.4 Related Practical Issues

When we introduce new variables to model nonlinear effects, it is important to consider their correlations with existing factors to avoid the multicolinearity problem. In practice, factors are either normalized z-scores or percentile. The former is approximately normally distributed with a restricted range from –3 to +3, and the latter is approximately uniformly distributed between 0 and 1.

Colinearity among factors: The correlation between the quadratic term and the linear term depends strongly on the distribution of the original factors. The correlation is minimal if the z-scores are used and the distribution is approximately normal (see Problem 9.5). On the other hand, the correlation is extremely high if the percentiles are used (see Problem 9.6). The high correlation subsequently results in an unstable estimation. Fortunately, we can use the Gram–Schmidt procedure to address this collinearity issue, as outlined in Chapter 7. The same is true for the correlation between the interaction term (product of two factors) and the original factors.

Conditional dummy: The aforementioned examples use a step function as the conditional dummy wherein there are only two possible values — 0 or 1. One issue with this approach is that the return forecast will change dramatically when a security is re-categorized from 0 to 1 or vice versa. To mitigate this problem, one can use a continuous step function as shown in Figure 9.4.

9.7.5 Nonlinear Effect vs. Contextual Model

Inquisitive reads may see that the conditional factor approach to nonlinear effect modeling is rather similar to the contextual modeling. They are

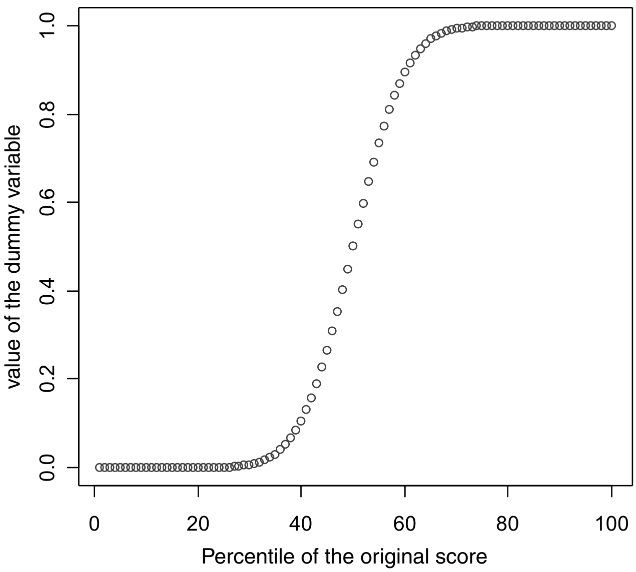

FIGURE 9.4. Continuous slope dummy.

both piecewise linear models. Specifically, both approaches first compartmentalize the cross-sectional security universe into homogeneous subgroups wherein securities tend to behave the same, and then form a set of piecewise linear models, one for each of the subgroups.

What makes them different and when should these approaches be applied? In general, the contextual modeling approach selects subgroups that are homogenous to many different alpha factors. For example, high-growth stocks' responses to cheapness, quality, and momentum are expected to differ from low-growth stocks. In this case, the contextual modeling approach is more appropriate. On the other hand, nonlinear effect modeling typically addresses one factor at a time, like the aforementioned CAPEX example. The security universe is partitioned into subgroups within each context that are expected to have different return responses to the original factor value.

The benefit of selecting the piecewise linear approach, instead of a full-bloom nonlinear modeling approach, is to maintain parsimonious parameterization. In addition, traditional linear statistics are more readily available, easier to understand, and more intuitive to interpret.

The benefit of a simultaneous estimation is the ability to capture different nonlinear effects across various contextual dimensions. In other words, nonlinear effects may also be contextually dependent. In addition, a simultaneous estimation will also deal with additional distributional issues, such as the correlation between a slope dummy and a contextual dimension. However, the argument against simultaneous estimation is overfitting, because the number of independent variables increases with the introduction of nonlinear terms, resulting in a dramatic decrease in the degrees of freedom.

9.7.6 Empirical Results

To compare the improvement in forecast efficacy, Figure 9.5 shows the decile returns of CAPEX factor for the Russell 2000 security universe. The panel on the left shows the decile performance of the original CAPEX factor and the panel on the right shows the transformed (new) CAPEX factor. Note that the factor return for the new CAPEX score is close to being linear, whereas the return for the original factor is clearly not. This supports our conjecture that a piecewise linear framework with parsimonious parameterization can provide enough flexibility to capture the nonlinear effects, without resorting to a full-bloom nonlinear model.

Modeling nonlinear effects has important implications for the performance of different portfolios. We note that most of the performance

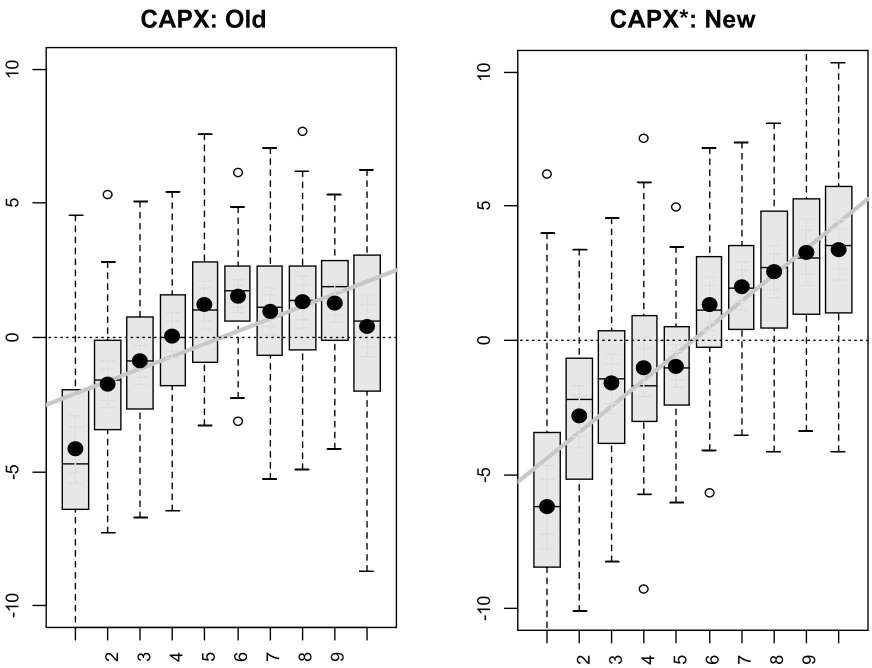

FIGURE 9.5. Performance comparison.

improvement for the new CAPEX factor comes from the long side or highly ranked stocks by CAPEX. As discussed before, CAPEX, in its original form, is effective in identifying losers due to the agency problem, but it does not add much value in picking winners. Therefore, the original factor is not very useful for long-only portfolios, as its benefits mostly come from "avoiding" losers for the long-only portfolios. The new CAPEX factor is now suited for long-only portfolios as well as long-short portfolios, because it symmetrically adds value both on the winner and the loser sides.

9.8 SUMMARY

In this chapter we highlighted two stringent assumptions behind a typical linear return forecasting model. These assumptions are not supported by empirical evidence and they impede the effectiveness of return forecasts. To improve return forecasting models, we introduced two advanced alpha modeling techniques: contextual alpha modeling and nonlinear effect modeling.

Both modeling approaches still utilize multifactor linear alpha models. However, a set of piecewise linear models are estimated and created simultaneously, one for each of the subuniverses that are carefully selected to ensure securities are homogenous within. When forecasting the future return of a security, different models are selected for each security dynamically, depending on the relevance between each model and the particular security. Relevance is governed by the security's attributes, such as growth rate, P/E ratio, or ROE. Nonlinear effects can be modeled in several different ways, including quadratic, conditional, or interaction models.

PROBLEMS

9.1 Find the condition under which the overall IR (9.9) is lower than the high dimension IR.

9.2 Derive the optimal weight (9.12) and calculate the optimal IR with parameters in Example 9.1.

9.3 Plot the function (9.16) for various values of coefficients. Prove that (a) the maximum return is at $F=-v_1/2v_2>0$ for $v_1<0$; (b) the minimum return is at $F=-v_1/2v_2<0$ for $v_1>0$. For the CAPEX factor, which case would apply?

9.4 Suppose factor mean and error mean are both zero in (9.16) and the factor is standardized. Then prove that $v_0+v_2=0$.

9.5 Suppose x is a normally distributed variable with zero mean. Prove that x and x^2 are uncorrelated.

9.6 Suppose x is uniformly distributed in the interval $\left[0,1\right]$. Prove that the correlation between x and x^2 is $\sqrt{15}/4 = 0.97$.

APPENDIX
A9.1 MODEL DISTANCE TEST

To gauge the significance of weighting difference — the likelihood of *not* attributing the cause solely to chance — we bootstrap the IC sample to simulate the inherent randomness of the weight estimation procedure by systematically introducing sampling errors into estimates. The bootstrapping procedure, similar to the one introduced by Michaud (1998), samples historical ICs, with replacement, one thousand times wherein one thousand sets of optimal weights are derived, one for each sample. This exercise is repeated for each security context to generate the set of resample weightings and the average of these weightings. We coin this average, $\mathbf{v}$, as the efficient factor weights — a convention dubbed by Michaud (1998). To illustrate how model distance is determined and tested, let us assume that $\mathbf{v}_1$ and $\mathbf{V}_1$ are the vector of efficient factor weights and the ensemble of resampled model weightings for the first security context, respectively, and that $\mathbf{v}_2$ and $\mathbf{V}_2$ are those for the second context. The vector of weighting difference is simply the difference between $\mathbf{v}_1$ and $\mathbf{v}_2$, $\Delta\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2$.

The equation below shows the chi-squared statistic when the weighting difference is tested against the sampling error generated from the second security context. The degree of freedom for this chi-squared test is the number of factors minus one, because factor weights sum up to 100%.

$$\chi^2 = \Delta\mathbf{v}' \cdot \mathbf{\Lambda}^{-1} \cdot \Delta\mathbf{v}\,, \tag{9.20}$$

where $\mathbf{\Lambda}^{-1}$ is the inverse of the covariance matrix for either $\mathbf{V}_1$ or $\mathbf{V}_2$.

As different covariance matrix, estimated from either $\mathbf{V}_1$ or $\mathbf{V}_2$, can be selected to compute the chi-squared statistic, significance test results may vary depending on the relative "tightness" of these covariances, albeit the same weighting difference is in question. Figure 9.6 shows a two-dimensional schematic plot of factor weights for a visual demonstration. The weighting difference is significance when using the covariance of $\mathbf{V}_2$ whose distribution on the right is tighter while the result is not significant with $\mathbf{V}_1$'s more diffused distribution. The dashed circles are the loci of significant distances for the two distributions, respectively.

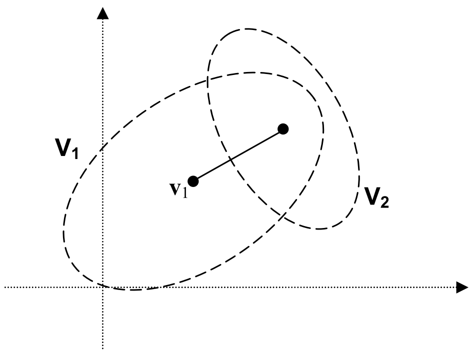

FIGURE 9.6. A two-dimensional projection of ensembles of optimal model weights. (From Sorensen, E.H., Hua, R., and Qian, E., *Journal of Portfolio Management*, Vol. 32, No. 1, 23–36, Fall 2005. With permission.)

REFERENCES

Asness, C.S., The interaction of value and momentum strategies, *Financial Analysts Journal*, 29–36, March–April 1997.

Beneish, M.D., Lee, C.M.C., and Tarpley, R.L., Contextual fundamental analysis through the prediction of extreme returns, *Review of Accounting Studies*, Vol. 6, 65–189, 2001.

Beckers, S., Steliaros, M., and Thomson, A., Bias in European analysts' earnings forecast, *Financial Analysts Journal*, 74–85, March–April 2004.

Black, F. and Litterman, R., Asset allocation: Combining investor views with market equilibrium, Goldman, Sachs & Co., Fixed Income Research, September 1990.

Black, F. and Litterman, R., Global portfolio optimization, *Financial Analyst Journal*, 28–43, September–October 1992.

Bruce, B. and Morillo, D., Extreme investing: Using extreme data to find value in small cap stocks, PanAgora working paper, May 2003.

Daniel, K. and Titman, S., Market efficiency in an irrational world, *Financial Analysts Journal*, 28–40, November–December 1999.

Das, S., Levine, C.B., and Sivaramakrishnan, K., Earnings predictability and bias in analysts' earnings forecasts, *The Accounting Review*, 277–294, April 1998.

Fama, E.F. and French, K.R., Multifactor explanations of asset pricing anomalies, *Journal of Finance*, Vol. 51, 55–84, 1996.

Heaton, J.B., Managerial optimism and corporate finance, *Financial Management*, 33–45, Summer 2002.

Huberts, L.C. and Fuller, R.J., Predictability bias in the U.S. equity market, *Financial Analysts Journal*, 12–28, March–April 1995.

Kahneman, D. and Tversky, A., On the psychology of prediction, *Psychological Review*, Vol. 80, No. 2, 273–251, June 1973.

Kahneman, D. and Tversky, A., Prospect theory: An analysis of decision under risk, *Econometrica*, Vol. 47, 263–271, March 1979.