GLOBAL EQUITY

Fundamental Factor Model

Authors: Ercument Cahan and Lei Ji

Date: August 2016

Bloomberg

TABLE OF CONTENTS

Chapter 1: Introduction

This document describes the Bloomberg Global Equity Fundamental Factor Model. The model employs a multiple factor modeling approach, which allows a responsive yet stable assessment of major risk factors in the Global equity markets. The broad universal coverage makes this model suitable for clients whose portfolios span multiple countries and regions and for those who prefer to take a global perspective on risk factors.

The main characteristics of the Bloomberg Global Model are:

- Coverage of over 100,000 Global equity securities with model start date 1999.
- Dynamic estimation universe updated weekly to ensure the model is responsive to the changing market environments; additionally, a gatekeeping system is designed to smooth out the estimation universe changes.
- 39 industry factors based on Bloomberg Industry Classification System (BICS).
- 10 style factors: Momentum, Value, Dividend Yield, Size, Trading Activity, Earnings Variability, Profitability, Volatility, Growth, and Leverage.
- 44 country / group factors covering over 100 countries
- Idiosyncratic risk modeling based on a separate structural factor model which incorporates additional variables useful for forecasting non-factor risk.

The rest of this document is organized as follows. Chapter 2 introduces factor models in terms of their general characteristics and how they are utilized to decompose and forecast portfolio risk. In Chapter 3, we go over the main features of the Global Equity Model: security universes, risk factors employed, the framework to predict factor covariance matrix, and the methodology to forecast idiosyncratic volatility. Finally, we present the model's performance statistics in Chapter 4. Definitions of style descriptors and factors as well as combination weights assigned to style descriptors are explained in Appendix A. The list of industry factors and the corresponding BICS codes are presented in Appendix B, whereas the list of country factors is shown in Appendix C.

Chapter 2: Factor Model Introduction

Since its introduction three decades ago, the factor model framework has been widely used in different areas of quantitative finance, including asset pricing, return forecasting, risk modeling, transaction cost analysis, and performance attribution. By decomposing asset returns into different components (also called factors), factor models allow a better understanding of the sources of portfolio return and risk, and have become an indispensable tool for portfolio analysis and risk management. In recent years, events such as the 2008 global financial crisis, the European debt crisis starting in 2010, and Britain's surprise exit from the European Union in 2016, have highlighted the importance of controlling portfolio risk and have greatly increased the interest of traditional and quantitative investors in factor models.

2.1 Structure of Factor Models

Factor models are based on the idea that security returns consist of two components: the part that is driven by a set of common factors, and the part unexplained by these factors hence idiosyncratic in nature. Mathematically:

$$r_{nt} = \sum_{k=1}^{K} X_{nkt} f_{kt} + \varepsilon_{nt} , \qquad (2.1)$$

where:

- r_{nt} is the return of security n over period t
- X_{nkt} is the exposure of security n to factor k at the beginning of period t
- f_{kt} is the factor return to factor k over period t
- ε_{nt} is the residual return of security n over period t
- There are N assets, K factors, and T time periods

In Equation (2.1), r_{nt} is known and ε_{nt} is estimated. How X_{nkt} and/or f_{kt} are estimated depends on the type of the factor model, which we discuss next.

2.2 Types of Factor Models

In practice, there are three common classes of factor models depending on the statistical and econometric techniques used to estimate factor returns and/or factor exposures. Before explaining each category in detail, we summarize in Table 2.1 the different characteristics of these models in terms of inputs, outputs, and estimation methods.

Table 2.1: Summary characteristics of the three types of factor models.

Model Type	Inputs	Outputs	Estimation Method
Explicit	Security returns; Factor returns	Factor exposures	Time-series regression
Implicit	Security returns; Factor exposures	Factor returns	Cross-sectional regression
Statistical	Security returns	Factor exposures; Factor returns	Principal Component Analysis

Explicit Factor Models start by specifying factor returns explicitly and then use econometric techniques such as regression analysis to estimate factor exposures. These models require the time series returns of each security and each factor. Sometimes these models are also called exogenous factor models (since factor returns are specified outside of the model, i.e. exogenously), time-series factor models (since each security's factor exposures are estimated by running time-series regressions), or macroeconomic factor models (since factors used are often macroeconomic variables). The advantage of explicit factor models is that they allow for inclusion of arbitrary factors, as long as factor time-series data is available. Besides, theoretically it's appealing to link the portfolio risk to macroeconomic fundamentals. The main disadvantage is that security's factor exposures tend to be non-intuitive. Moreover, since factor exposures are determined from historical time-series regressions, such models might have good fits in-sample but may have low explanatory power out of sample.

Implicit Factor Models treat factor exposures as observable, and then estimate factor returns with cross-sectional regressions of securities' returns on their factor exposures. Implicit factor models require returns and various exposures data for each security, and are thus the most data intensive among the three types. Sometimes these models are also called endogenous factor models (since factor returns are derived within the model), cross-sectional factor models (since cross-sectional regression is used), or fundamental factor models (since exposures are typically based on securities' fundamental attributes). One advantage of these models is that they are intuitive to understand. For example, assume the model includes size and momentum factors. Then a large-cap stock that has recently underperformed the market would have a positive exposure to the size factor and a negative exposure to the momentum factor. Another advantage is their responsiveness to changes in asset characteristics. For example, a sudden change in fundamentals, say due to corporate actions, can be instantly incorporated into the model when the exposure data is updated. The main disadvantage of these models is simply that they are much more data intensive than other types of models.

Statistical Factor Models use statistical techniques such as Principal Component Analysis (PCA) to simultaneously estimate factor returns and factor exposures from security returns. The advantages of such models are that they require minimal data (only the security-level price data) and are relatively easy to build. In addition, since the factors are derived from asset returns they might be able to capture new sources of risk. Their main disadvantage is the lack of interpretability. For example, it is not clear what a portfolio manager should do if the risk model ascribes a big portion of portfolio risk to the fourth principal component because there is usually no easy way to associate a clear economic meaning to this component.

We summarize the advantages and disadvantages of different model types in the table below.

Table 2.2: Advantages and disadvantages of factor model types.

Model Type	Advantages	Disadvantages
Explicit	▪ Accommodates any time series as factors	▪ Possible non-intuitive factor exposures
	▪ Linked to macroeconomic fundamentals	
Implicit	▪ Intuitive exposures linked to asset characteristics	▪ Data intensive
	▪ Incorporates security fundamental data	
	▪ Responsive to asset characteristics changes	
Statistical	▪ Requires only asset-level price data	▪ Lack of interpretability
	▪ Might capture new risk sources	
	▪ Relatively easy to build	

Bloomberg Equity Factor Models are based on the implicit factor approach. The regression in Equation (2.1) utilizes daily or weekly stock returns, and is run on a daily or weekly basis for a set of stocks that we call the Estimation Universe. To minimize the estimation error, securities in the Estimation Universe are assigned different weights in the regression process. Mathematically, a sensible weighting scheme would under (over) weight stocks with large (small) idiosyncratic volatility. It has been observed that larger firms tend to have lower idiosyncratic volatility than smaller firms. Empirically, the square root of stocks' market capitalization is roughly inversely proportional to their idiosyncratic variance. Therefore we use square root of the market-cap as the weighting scheme when we estimate the regression in Equation (2.1).

2.3 Risk Decomposition and Estimation with Factor Models

Once factor exposures are calculated, one can carry out risk decompositions and forecasts at both the individual security and the aggregate portfolio levels.

Let R_P denote the return of some portfolio P over period t, then we have:

$$R_P = \mathbf{w}'_P \mathbf{r} \ , \tag{2.2}$$

where $\mathbf{w}_P = \left(w_1, ..., w_N\right)'$ is the column vector representing portfolio weight[1] and $\mathbf{r} = \left(r_1, ..., r_N\right)'$ denotes the column vector of security returns.

Note that Equation (2.1) can be written in matrix form as follows (time subscript suppressed):

$$\mathbf{r} = \mathbf{X}\mathbf{f} + \boldsymbol{\varepsilon} \ , \tag{2.3}$$

where $\mathbf{r}$ is the column vector of security returns and $\mathbf{X}$ is the $N \times K$ matrix specifying each security's exposures to the model factors. Substituting Equation (2.3) into Equation (2.2), we obtain:

---

[1] Note that portfolio weight elements can be zero, positive, or negative. Portfolio weight can change through time. In the exposition above the time subscript is suppressed for simplicity.

$$R_p = \mathbf{w}'_P \mathbf{X}\mathbf{f} + \mathbf{w}'_P \boldsymbol{\varepsilon} \ . \tag{2.4}$$

To calculate the risk σ_P of the portfolio, we need to first estimate the $N \times N$ asset covariance matrix $\mathbf{\Omega}$. Direct calculation (i.e. without using a factor model) of this covariance matrix involves computing the pairwise sample covariance for each element in the matrix. Typically, the number of assets N is large relative to the number of time periods T of the return time series, which makes the estimated covariance matrix unstable and ill-conditioned.

The factor model framework allows us to reduce the dimensionality of this problem significantly. In this approach, we can express the asset covariance matrix in terms of factor model ingredients:

$$\mathbf{\Omega} = \mathbf{X}\mathbf{F}\mathbf{X}' + \mathbf{\Delta} \ , \tag{2.5}$$

where $\mathbf{X}$ is the factor exposure matrix, $\mathbf{F}$ is the $K \times K$ factor covariance matrix, and $\mathbf{\Delta}$ is the $N \times N$ idiosyncratic covariance matrix. In most applications, $\mathbf{\Delta}$ is assumed to be a diagonal matrix whose N diagonal elements represent each security's idiosyncratic variance[2].

Compared to estimating $N(N+1)/2$ parameters by calculating $\mathbf{\Omega}$ directly, we only need to estimate $K(K+1)/2+N$ parameters with the factor model framework. Given N (hundreds or thousands) is usually much larger than K (dozens), this is a quite significant reduction in dimensionality.

With $\mathbf{\Omega}$ estimated via the factor model, the variance of portfolio P is given by,

$$\sigma_P^2 = \mathbf{w}'_P \mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{w}_P + \mathbf{w}'_P \mathbf{\Delta} \mathbf{w}_P \ . \tag{2.6}$$

In the above equation, portfolio risk is decomposed into the factor component (the first term) and the idiosyncratic component (the second term). This allows us to build separate forecasting models for each component.

---

[2] Note that Δ ceases to be diagonal when there are securities that are connected to each other with a parent-child relation. For details, see the Parent-Child Securities section below.

2.4 Advantages of Factor Models

Through the lenses of the factor model framework, we can decompose asset and portfolio risk into systematic and idiosyncratic components, and we can further break down systematic risk into intuitive and easily interpretable systematic factors. This decomposition gives users a deep and intuitive understanding of sources of risk.

By imposing a parsimonious structure on the covariance matrix of a large number of assets, factor models significantly reduce the dimensionality of the problem, and this parsimony manifests itself in many ways. It tends to filter out the noise in the data and increases the signal-to-noise ratio, and as a result it leads to more robust covariance estimators for portfolio optimization. In addition, it allows very efficient numerical calculations for covariance matrix, value at risk (VaR), expected tail loss, and many other risk measures. All of these are extremely important for modern portfolio construction and risk management.

Chapter 3: Model Overview

3.1 Security Universes

There are two types of security universes in the model: the coverage universe and the estimation universe. The former refers to the universe of securities covered by the model, while the latter defines the set of securities used in the cross-sectional regression to estimate factor returns.

Coverage Universe

Coverage for the Global Equity Fundamental Factor model extends to all equity securities listed on major exchanges with security types one of the following: common stock, preferred stock, unit, REITs and receipts. In addition, the security must satisfy the following data availability requirements in order to be covered:

- Bloomberg price and market-cap data are available
- Industry and country membership information is available

The coverage universe is quite broad, with an average of roughly 100,000 securities over the model history.

Estimation Universe

The estimation universe defines the subset of stocks used in the cross-sectional regression in Equation (2.1), and is constructed as follows. All the stocks in the coverage universe are first sorted by decreasing market-cap within each of the 44 country groups as listed in Appendix B. Then those that cumulatively account for 98% of the total market-cap of the respective country group become candidates for the estimation universe. This process is repeated each week to incorporate the latest data. We take additional steps to ensure the universe's stability by making it harder for stocks to enter or leave the universe in the case of short-term fluctuations in their market-cap. Finally, if a security is a member of a major global equity index, it's included in the estimation universe. Figure 3.1 and Figure 3.2 below show how the number of securities and the total market-cap (in USD) of the Global estimation universe change over time, respectively.

Figure 3.1: Number of securities in the estimation universe.

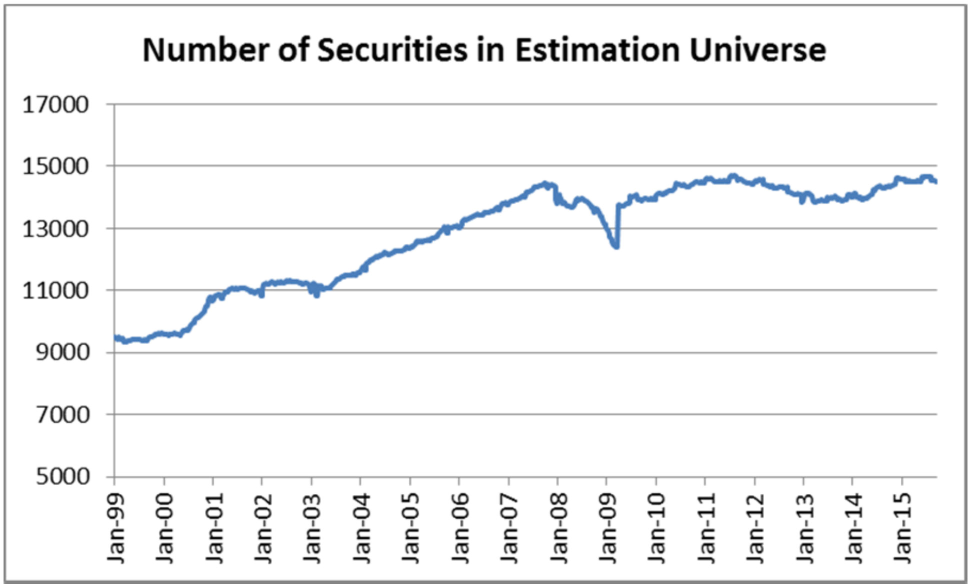
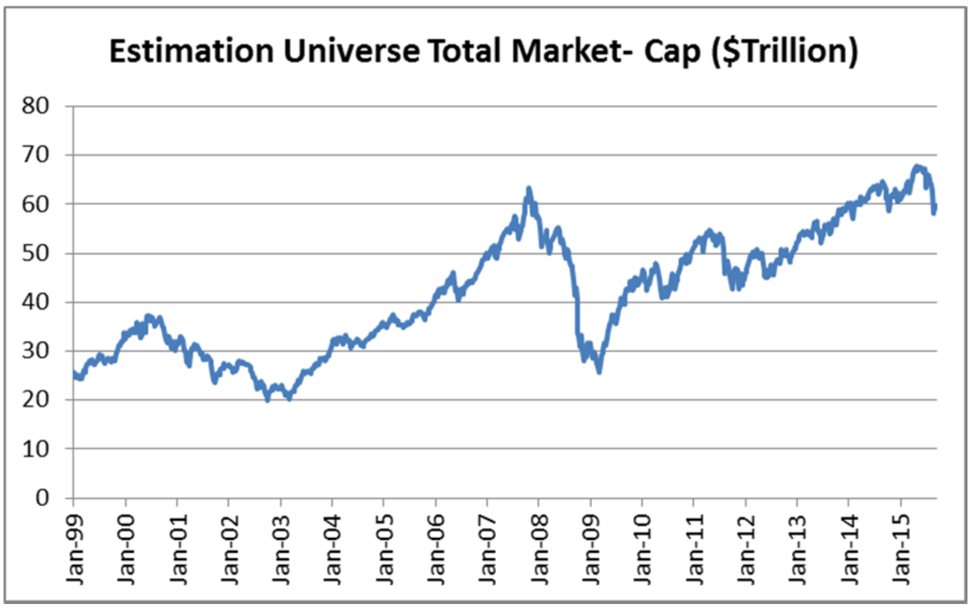


Figure 3.2: Total market-cap of the estimation universe (in Trillion USD).



3.2 Model Factors

The Bloomberg Global Equity Factor Model includes four categories of factors: Market, Country, Industry, and Style. In this section, we explain each category in turn, and then describe our approach to combine multiple descriptors to form a single factor exposure.

Market Factor

Every stock in the model has a unit exposure to the Global market factor. This factor is the dominant source of total risk for diversified long-only portfolios. However, this factor typically has little contribution to a portfolio's active risk (risk relative to a benchmark) unless the portfolio holds a sizable cash position. This factor can be thought of as the intercept term in Equation (2.1).

Industry Factors

The Global model employs a bespoke set of 39 industry factors based on a combination of BICS Level 1, 2, and 3 industry groupings. These industry factors give a detailed yet parsimonious representation of the industry structure of the global equity market. We list in Appendix B the 39 industry factors and the corresponding BICS codes.

If a security belongs to a given industry, its exposure to that industry is set to one, while its exposures to the other industries are set to zero. Every stock belongs to one and only one industry, and the 39 industry factors form a complete and disjoint classification of all the stocks.

By definition, the 39 industry exposures sum to one for each stock, which implies that the market factor exposure can be expressed as the summation of the 39 industry exposures. This introduces the perfect multicollinearity issue into the regression in Equation (2.1), and we need to impose certain linear restriction(s) to obtain a unique regression solution. In particular, the regression is run with the restriction that the cap-weighted industry factor returns sum to zero.

Country Factors

The Global Equity Model utilizes 44 country factors defined by 44 country groups. The list of these country groups and their corresponding constituent countries are reported in Appendix C.

Each equity is assigned unit exposure to its corresponding country factor, and zero exposures to other countries. As with industries, the country factor exposures are exactly collinear with the market factor. As a result, we must impose an additional constraint to obtain a unique regression solution. Similar to the case with industries, we set the cap-weighted sum of country factor returns equal to zero in each period.

Style Factors

Style factors are the type of factors that are related to securities' fundamental characteristics. Each style factor consists of one or multiple "atomic" descriptors. The advantages of using multiple descriptors are more robust calculations of factor exposure and better explanatory power. Precise definitions and formulas for style factors and descriptors as well as the descriptor combination weights can be found in Appendix A.

Bloomberg detects outliers in the descriptor data and apply rigorous checks to the underlying data to uncover data errors. To further guarantee the model estimation is not adversely influenced by outliers, we apply "winsorization limits" specific to each descriptor. For example, dividend yield is only allowed to vary between 0 and 100%.

Because different raw descriptors have different units and scales, we standardize them before combining them into a single style factor. To standardize a given descriptor, we first subtract the cap-weighted mean within each country / country group, and divide by the equal-weighted standard deviation across the entire estimation universe. Standardized values outside of the -3.0/+3.0 bounds are capped at -3.0 or +3.0. This process is repeated multiple times until the mean and the standard deviation converge to zeros and one numerically. Note that using the cap-weighted mean guarantees that the market portfolio within a given country will have zero exposures to the style factors.

Once descriptor values are calculated and standardized, a weighted combination of these standardized descriptors within the same category forms the exposure to the factor in question. For example, to form the Size factor, we combine three standardized descriptors: logarithm of market-cap, logarithm of sales, and logarithm of total assets. We explain in the next section how the combination weights assigned to descriptors are determined.

Style factor exposures are designed to be relatively stable through time, as can be seen in Table 3.1, which reports the average Spearman autocorrelations for each style factor. Most factor exposures, except momentum, maintain relatively high correlations with their past values even after one year.

Table 3.1: Spearman autocorrelation of style factor exposures.

Factor	1-week	4-week	52-week
Momentum	0.97	0.91	0.01
Value	0.99	0.96	0.65
Dividend Yield	0.99	0.98	0.80
Size	0.99	0.99	0.97
Trading Activity	0.99	0.99	0.90
Earning Variability	0.99	0.98	0.87
Profitability	0.99	0.97	0.68
Volatility	0.99	0.95	0.44
Growth	0.99	0.97	0.76
Leverage	0.99	0.99	0.90

Descriptor Weighting

To combine atomic descriptors into style factors, Bloomberg has developed a descriptor-weighting algorithm. The main idea is to find a common dimension among the descriptors within a given style category. For example, the growth factor consists of five atomic descriptors. Three are based on historical growth of sales, earnings, and total assets, while the other two are based on forward-looking growth of sales and earnings which are inferred from fundamental analysts' estimates. How should we combine them? Equal weighting would be the simplest solution but not necessarily the optimal approach since it ignores the correlation structure between descriptors. Bloomberg uses the following intuitive approach designed to capture the most common information contained in descriptors relationships.

First, we calculate the cross-sectional Spearman rank correlation matrix of descriptors[3]. Then we run principal component analysis (PCA) on this correlation matrix and extract the first principal component (PC1) since it explains the most descriptor variability. The loadings of PC1 on the descriptors are then normalized to sum up to 100% and these normalized PC1 loadings are the final weights used to combine the descriptors. To illustrate, we have the following five atomic descriptors within the Growth style factor:

- SG: Sales growth
- EG: Earnings Growth
- TAG: Total Asset Growth
- SFG: near-term forecast sales growth based on analyst estimates
- EFG: near-term forecast earnings growth based on analyst estimates

In Table 3.2, the second column reports PC1's loadings on the five descriptors, and the third column presents the normalized weights:

Table 3.2: Weights assigned to Growth descriptors based on PCA.

Descriptor	Descriptor loading on PC1	Descriptor Weights
SG	0.55	25%
TAG	0.51	23%
SFG	0.44	20%
EG	0.40	18%
EFG	0.31	14%

Note that SG has the highest loading on PC1 (0.55) which indicates it points closer to the direction of the overall growth characteristic (i.e. the direction pointed by PC1) than the other descriptors. As a result, SG gets the highest weight (25%) in the combined growth factor.

The combination weights for all the style descriptors can be found in Appendix A.

[3] Compared to the usual Pearson correlation, Spearman correlation is less sensitive to outliers.

3.3 Factor Covariance Matrix

Bloomberg follows a systematic approach to estimate the covariance matrix of factor returns estimated from Equation (2.1). First we decompose the factor covariance matrix into a correlation matrix and a diagonal matrix, whose diagonal elements are factor volatilities, and estimate the two matrices separately:

$$\mathbf{F} = \mathbf{VCV} \, , \tag{3.1}$$

where $\mathbf{F}$ is the factor covariance matrix, $\mathbf{C}$ the factor correlation matrix, and $\mathbf{V}$ the diagonal factor volatility matrix.

To estimate $\mathbf{V}$, we use exponentially weighted moving average (EWMA) with a 26-week half-life. The correlation matrix estimator $\mathbf{C}$ is constructed by blending the sample correlation matrix $\mathbf{C}_0$ and the PCA correlation matrix $\mathbf{C}_p$:

$$\mathbf{C} = 0.8\mathbf{C}_0 + 0.2\mathbf{C}_p \, , \tag{3.2}$$

where $\mathbf{C}_0$ is computed using EWMA with a 52-week half-life, and $\mathbf{C}_p$ is estimated using the J principal components with the largest J eigenvalues. The number of components J is determined as 10% of the number of factors in the model (rounded down to the nearest integer). For the Global Equity model, J equals to nine since there are 94 factors.

Bloomberg research has shown that this blending methodology significantly improves the out-of-sample performance of the model compared to other approaches commonly used in the industry (such as simple EWMA or RMT cleaning). For detailed information about the methodology of the factor covariance matrix forecasting, please refer to the Bloomberg “Multi-Asset Class Risk Model (MAC2)” white paper available in the PORT help page.

3.4 Non-Factor Risk

Non-factor or idiosyncratic risk is the risk of idiosyncratic returns, the return component unexplained by the factors in the risk model. Bloomberg’s non-factor risk forecast is based on a structural factor model. We use style, industry, and country factors and supplement with variables that are specifically useful to forecast the idiosyncratic risk.

The non-factor risk model decomposes non-factor risk into the market-wide component (which is common to all stocks) and the stock-relative component (which is stock specific). Both components evolve over time. By modeling these two components separately and then combining them, Bloomberg is able to forecast non-factor risk in a systematic approach.

As a start, we collect the non-factor returns from the regression in Equation (2.1), and calculate their absolute values $|\varepsilon_{nt}|$ for each stock. Let M_t denote the cap-weighted average of $|\varepsilon_{nt}|$ across all the securities in period t. We express $|\varepsilon_{nt}|$ in the following form:

$$|\varepsilon_{nt}| = M_t(1+S_{nt}) \ , \tag{3.3}$$

where S_{nt} is the stock-relative part of $|\varepsilon_{nt}|$, given by:

$$S_{nt} = \frac{|\varepsilon_{nt}| - M_t}{M_t} \ . \tag{3.4}$$

Equation (3.3) decomposes each security's non-factor risk into the market-wide part M_t and an asset specific part S_{nt}. This allows us to model the two pieces of the puzzle separately. First, we forecast M_t by applying EWMA to its historical time series with a 26-week half-life. To forecast the stock-relative part, we run the following pooled regression of historical S_{nt} on a set of forecasting variables:

$$S_{nt} = \sum_{k=1}^{K} X_{nkt} G_{k\tau} + Y_{nt} B_{\tau} + \varphi_{nt} \ , \tag{3.5}$$

where the first K exposures X_{nkt} are the same as the K factor exposures in Equation (2.1) , and the variable Y_{nt} is constructed using previously realized values of S_{nt} to capture the persistence effect. This extra variable is calculated as the EWMA (with a 26-week half-life) of the past stock-relative component values before time τ, the period being forecast. The coefficients are estimated by running the above regression on weekly data pooled over the most recent two years. Note that the coefficients $G_{k\tau}$ and B_τ don't have the time t subscript because this is a regression pooled over time. However, since Equation (3.5) is re-estimated every week, coefficients $G_{k\tau}$ and B_τ depend on time τ and evolve slowly over time.

It is important to emphasize the major differences between the regressions in Equation (2.1) and in Equation (3.5). First, the former is run on returns, whereas the latter is run on absolute returns. Therefore the regression coefficients have very different meanings even though they load on the same factors. For example, Bloomberg finds that the coefficient on the Size style factor is consistently negative in Equation (3.5), meaning smaller stocks on average have higher idiosyncratic volatility, all else being equal. On the other hand, the Size coefficient estimated from Equation (2.1) changes signs frequently, depending on whether small stocks outperform large stocks.

Another important difference is that Equation (2.1) is estimated cross-sectionally using data in one period, whereas the regression in Equation (3.5) is run using data pooled over two years. The power of pooling ensures that the relationships uncovered between absolute idiosyncratic returns and stock characteristics are robust, leading to more stable idiosyncratic risk forecasts. To make the pooled regression model more responsive to changing market conditions, we assigned exponential weight to the pooled data (52-week half-life).

Running the regression in Equation (3.5) we obtain the forecast of $S_{n\tau}$ as follows:

$$\hat{S}_{n\tau} = \sum_{k=1}^{K} X_{nk\tau}\hat{G}_{k\tau} + Y_{n\tau}\hat{B}_\tau \ , \tag{3.6}$$

where $\hat{G}_{k\tau}$ and $\hat{B}_\tau$ are estimated coefficients as of time τ. Once we have the component forecasts $\hat{M}_\tau$ and $\hat{S}_{n\tau}$, we combine them to obtain the forecast for the absolute magnitude of the idiosyncratic return. Finally, we convert the absolute magnitude forecast to the standard deviation forecast by applying the $\sqrt{\pi/2}$ conversion multiplier[4]. Mathematically:

$$\hat{\sigma}_{\varepsilon,n\tau} = \hat{M}_{\tau}\left(1+\hat{S}_{n\tau}\right)\sqrt{\pi/2}\ . \tag{3.7}$$

For more detailed information about the non-factor volatility forecast model, please refer to "The Equity Portfolio Factor Model: Non-Factor Risk Forecasting" white paper in the PORT help page.

3.5 Treatment of Parent-Child Securities

The coverage universe contains parent-child securities such as depositary receipts and closely related securities (e.g. different share classes of the same security). For these securities, Bloomberg determines which security is the parent and which one is the child (e.g. depositary receipts are taken as the children of their parent companies). In general, parent securities have more reliable fundamental and market data compared to their children. Therefore, when calculating the standardized factor exposures for a child security, Bloomberg uses the raw fundamental and market data of the parent.

Note that the statistical relation between parent-child securities cannot be explained by model factors alone. To account for their residual relations, we set their non-factor correlation to be 0.9. This makes the relevant off-diagonal elements of the non-factor covariance matrix Δ in Equation (2.5) non-zero. Therefore, Δ is no longer strictly diagonal when there are parent-child securities present in the universe.

3.6 Missing Data for Factor Exposures

Despite our best data collection efforts, sometimes some stocks don't have the underlying data available to calculate certain factor exposures. Rather than exclude such stocks from our coverage universe, we fill the missing factor exposure data with a

---

[4] Suppose the random variable ε has a normal distribution with mean 0 and standard deviation σ_{ε}. Then it can be shown that the mean of $|\varepsilon|$ is equal to $\sigma_{\varepsilon}\sqrt{2/\pi}$. Therefore, one can obtain σ_{ε} by multiplying $E\left(|\varepsilon|\right)$ by $\sqrt{\pi/2}$. By analogy, the sample estimate $\hat{\sigma}_{\varepsilon}$ can be obtained by multiplying the sample counterpart of $E\left(|\varepsilon|\right)$ by $\sqrt{\pi/2}$.

special algorithm. This procedure infers the stock's missing descriptor values based on its other available attributes, such as industry membership, firm size, and country membership. For example, if a momentum factor is missing for a stock due to a short data history or recent IPO, we fill that stock's missing momentum exposure with the average momentum exposure of stocks in the same industry and of similar size.

Chapter 4: Model Performance

This chapter presents a high-level summary of empirical results on the model's performance along the following dimensions:

- Overall model explanatory power
- Factor behavior and significance
- Bias tests on different testing portfolios

The main takeaways are: the risk factors included in the model are mostly significant in different market environments, the model's risk forecasts do not exhibit large biases for a variety of testing portfolios, and overall the model demonstrates strong forecasting capability even during financial crisis periods.

4.1 Overall Explanatory Power

Figure 4.1 shows the adjusted R^2 (smoothed using EWMA with a six-month half-life) of the cross-sectional regressions over time. As can be seen, the model's explanatory power is generally high throughout the history, and it peaked during the quite volatile 2008-2009 period in the wake of the 2008 global financial crisis.

Figure 4.1: Smoothed adjusted R-squared.

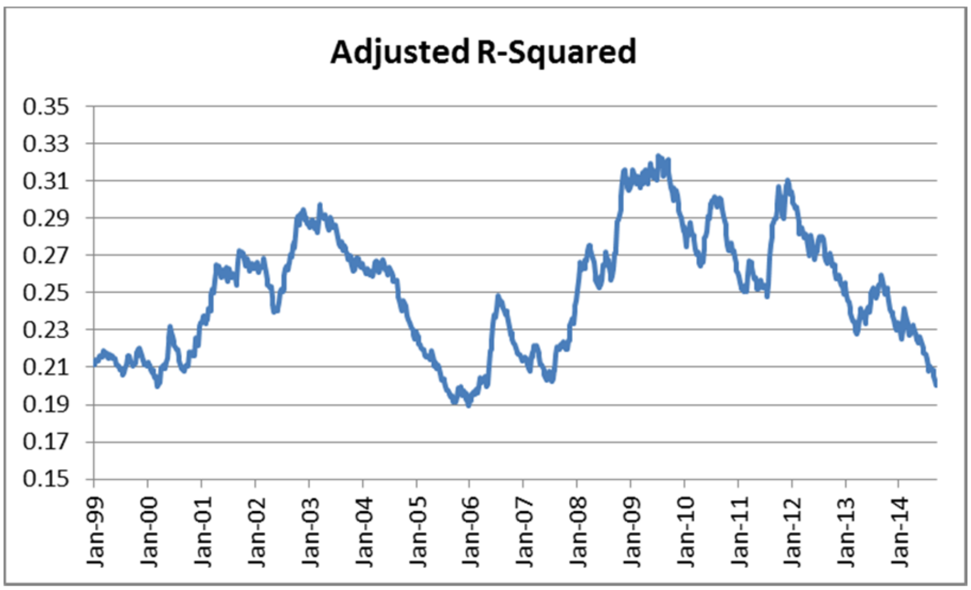
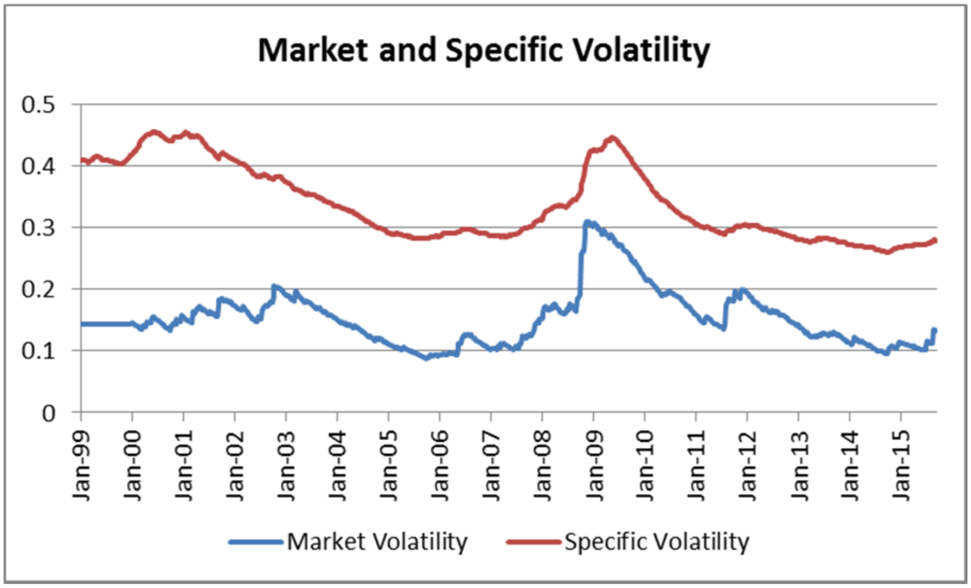


Figure 4.2 shows the estimated volatility of the market factor and the equally-weighted average specific volatility of the estimation universe. As expected, during the 2008 crisis both volatilities increased sharply and they stayed high until 2010.

Figure 4.2: Market factor volatility and average residual volatility.



4.2 Factor Behavior and Significance

Factor Returns over Time: A desirable property of estimated factor returns is their ability to convey both the magnitude and directional effects of systematic risk sources. Figure 4.3 and Figure 4.4 below show the cumulative performance of the ten style factors starting from 1999, with a unit initial investment. As we can see, the Value, Dividend Yield, and Momentum factors have had relatively consistent performance over time, although the Momentum factor had a significant drawdown in 2009.

To illustrate how well the model forecasts the volatility of the risk factors, we depict in Figure 4.5 the realized returns of a representative factor (the market factor) along with the 95% (two standard deviations) confidence interval based on the model's volatility forecast. The 95% bound is breached about 5.2% of the time, which is very close to the theoretical value under the normal distribution assumption. Some major periods of breach include Fall 2008 when the Lehman bankruptcy rocked the global financial markets and August 2011 when S&P's abrupt downgrading of US debt and amounting fears of the European sovereign debt crisis triggered a sharp selloff in global markets.

Figure 4.3: Cumulative performance of style factors (part I).

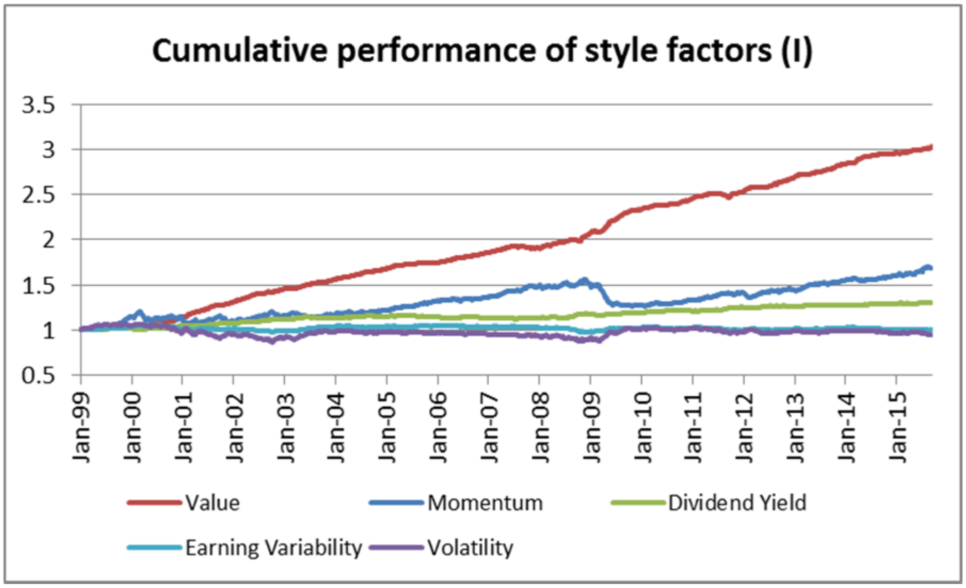


Figure 4.4: Cumulative performance of style factors (part II).

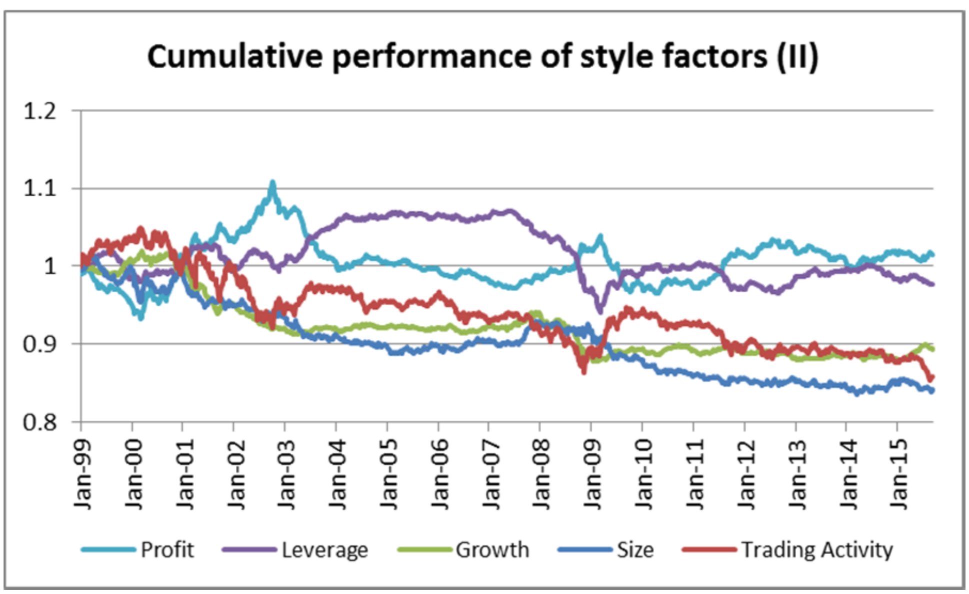
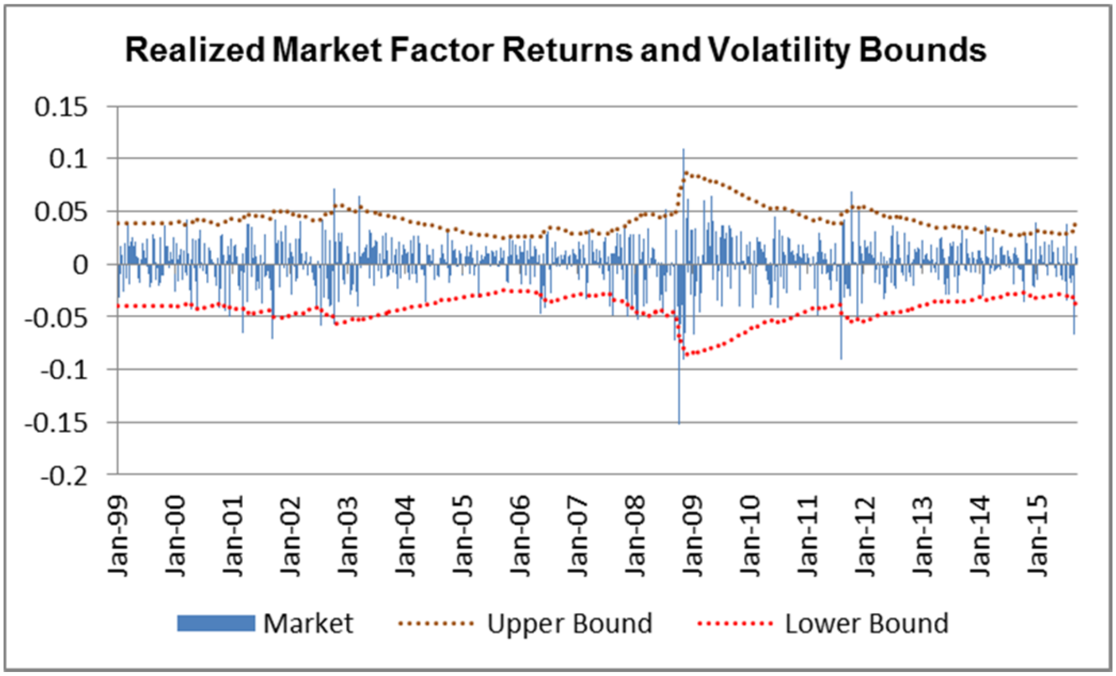


Figure 4.5: Market factor's realized weekly returns with +/- 2 standard deviation bounds.



Statistical Significance of Factors: Examining the statistical significance of factors helps to judge whether the model has chosen the relevant factors. For this, we evaluate the t-statistics associated with the estimated factor returns from the cross-sectional regressions. Under standard regression assumptions, a large t-statistic value leads to rejecting the null hypothesis of regression coefficient being zero. The t-statistic critical value we use is 2, corresponding to the commonly used 95% significance level.

The t-statistics for the 1999-2015 period are summarized in the first two columns of Table 4.1 and Table 4.2. Table 4.1 reports the results for the market factor and for the average style and industry factors, whereas Table 4.2 studies these characteristics for each individual style factor. The first column shows the percentage of time the t-statistic has absolute value exceeding 2 (hence significant at 95% confidence level), while the second column represents the average magnitude of the t-statistic over time.

As we can see, the market factor is significant 95.5% of the time, while typical style, industry, and country factors are significant 61.7%, 46.1%, and 59.1% of the time, respectively. In line with these numbers, the market factor demonstrates an average absolute t-statistic of 27.78 versus 4.03, 2.53, and 4.09 for typical style, industry, and country factors, respectively.

Table 4.1: Statistical properties of market, and average style / industry / country factors (1999-2015).

Factor	Freq (|t|>2)	Mean(|t|)	Factor Volatility	Market Correlation	Sharpe Ratio
Market	95.5%	27.78	15.9%	1.00	0.49
Average Style	61.7%	4.03	2.2%	0.26	0.50
Average Industry	46.1%	2.53	7.6%	-0.05	-0.01
Average Country	59.1%	4.09	15.2%	-0.07	0.11

Table 4.2: Statistical properties of individual style factors (1999-2015).

Factor	Freq (|t|>2)	Mean(|t|)	Factor Volatility	Market Correlation	Sharpe Ratio
Dividend Yield	43.4%	2.12	1.1%	-0.36	1.47
Earning Variability	51.9%	2.53	1.6%	0.59	0.01
Growth	42.4%	2.07	1.3%	0.43	-0.51
Leverage	49.7%	2.69	1.5%	0.43	-0.09
Momentum	80.2%	6.50	3.9%	-0.17	0.82
Profitability	62.2%	3.33	2.0%	-0.61	0.06
Size	69.9%	4.45	2.0%	0.46	-0.51
Trading Activity	78.0%	5.90	3.2%	0.73	-0.27
Value	59.7%	3.01	1.6%	0.23	4.08
Volatility	79.8%	7.72	4.2%	0.83	-0.05

Additional columns in Table 4.1 and Table 4.2 describe other aspects of the factor returns. The "*Factor Volatility*" column shows the annualized volatility of each factor return series. In general, the market and industry factors have much higher volatility than style factors. The "*Market Correlation*" column displays the historical sample correlation between each factor's return series and the market factor return. Finally, the "*Sharpe Ratio*" column presents the annualized Sharpe Ratio (annualized mean divided by annualized volatility) for each factor. As can be seen, Value, Dividend Yield, and Momentum factors had higher Sharpe ratios than the other factors, in line with the evidence in Figure 4.3.

4.3 Portfolio Bias Tests

The bias statistic is a useful metric for measuring the accuracy of risk forecasts. It's calculated as the standard deviation of realized returns normalized by risk forecast. Conceptually, the bias statistic represents the ratio of realized risk to forecast risk.

We have extensively evaluated the model's performance using bias tests for a broad variety of testing portfolios. All the portfolios are formed with stocks in the estimation universe. The following three groups of portfolios are considered:

a) Factor-Mimicking Portfolios: Pure factor portfolios generated by restricted cross-sectional regressions[5]. There are 94 portfolios in this category.

b) Specific-Return Portfolios: Portfolios created with the specific returns of 100 randomly selected stocks.

c) Factor-Tilted Portfolios: There are the three sub-groups for this portfolio group,

i. Market Portfolio: the market-cap weighted portfolio of all the stocks in the estimation universe.
ii. Industry-Tilt Portfolios: portfolios that long a cap-weighted industry and short the cap-weighted market portfolio. These portfolios typically have incidental exposures to style factors. There are 39 such portfolios.
iii. Long-Short Style Portfolios: portfolios that long the top quintile of stocks sorted on a given style factor exposure, and short the bottom quintile of stocks. Stocks within the long (short) side are market-cap weighted. These portfolios might have incidental industry exposures. There are 10 such portfolios.

For each of these test portfolios, we first calculate its bias statistic each week using a rolling 26-week look-back window. To summarize the distribution of bias statistics in a given week, we compute the 10th, 50th and 90th percentile values of these statistics across all the testing portfolios within the portfolio group. This process is then repeated each week to generate three weekly time series corresponding to the three percentiles.

[5] For details on factor mimicking portfolios, please see: Menchero, Jose. 2010. "The Characteristics of Factor Portfolios." The Journal of Performance Measurement, Fall, 52-62.

Observing the plots of these three time series lets us understand how the model's performance changes over time and across different portfolios.

It's commonly known that if we assume risk forecasts are correct and the returns are normally distributed, the bias statistic is normally distributed with mean 1 and standard deviation $1/\sqrt{2H}$, where H is the number of time periods in the rolling estimation window. Given $H = 26$, it's straightforward to see that the theoretical 10th, 50th and 90th percentiles for the bias statistic are 0.82, 1 and 1.18, respectively.

Figure 4.6 shows these bias statistic percentiles for the Factor-Mimicking portfolios. Similarly, Figure 4.7 and Figure 4.8 demonstrate how the percentiles change over time for the Specific-Return and Factor-Tilted portfolio groups respectively.

Figure 4.6: Bias statistics summary for the factor-mimicking portfolios.

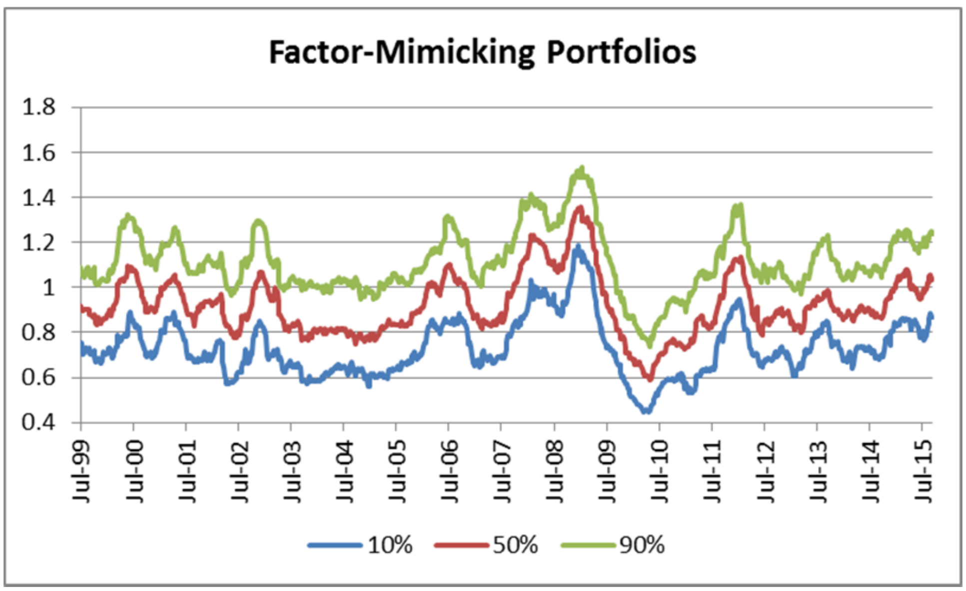


Figure 4.7: Bias statistics summary for the specific-return portfolios.

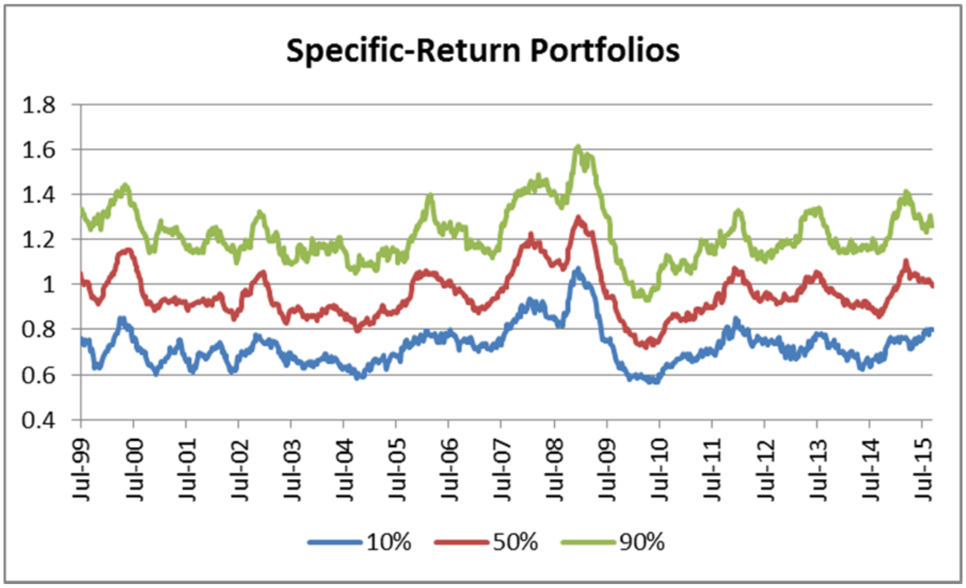
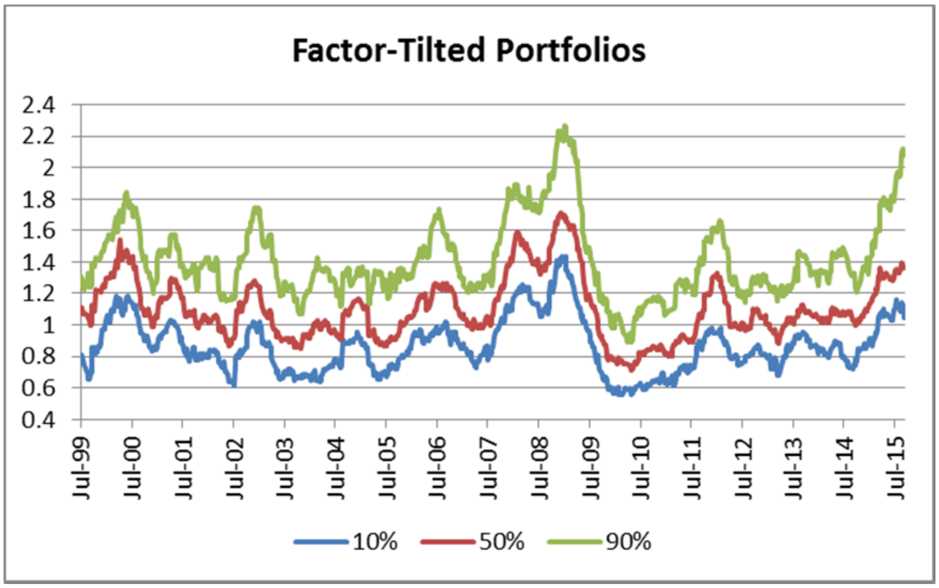


Figure 4.8: Bias statistics summary for the factor-tilted portfolios.



As can be seen, bias statistic percentile values generally move around their theoretical values and the model does not demonstrate very large biases in different market environments. We also show the average of each percentile time series plotted above in Table 4.3 below.

Table 4.3: Time-series average of bias statistics percentiles.

Portfolio Group	10%	50%	90%
Factor-Mimicking	0.73	0.92	1.12
Specific-Return	0.73	0.96	1.22
Factor-Tilted	0.87	1.1	1.41

Note that the values in Table 4.3 are slightly different from their theoretical counterparts obtained from the distribution of the bias statistics in Equation (4.1). There are two explanations for this discrepancy. First, the values in Table 4.3 are sample estimates hence subject to sampling errors. Second, the theoretical values are derived from the normal distribution in Equation (4.1), but normality is often violated by the real financial data series considered.

Appendix A: Style Factor Definitions

Factor	Calculation	Bloomberg Data
Momentum separates stocks that have outperformed over the past year and those that have underperformed.	Cumulative return over one year (averaged), skipping the most recent two weeks to mitigate the price reversal effect: $$Momentum_t = \frac{1}{53} \sum_{s=54\ weeks}^{s=2\ weeks} r_{n,t-s} \quad (100\%)$$	PX_LAST
Value is a composite metric that differentiates “rich” and “cheap” stocks. Bloomberg combines fundamental and analyst consensus data to calculate this factor.	Combination of the following descriptors: • Book to Price (13%) • Earnings to Price (19%) • Cash Flow to Price (18%) • Sales / EV (10%) • EBITDA / EV (21%) • Forecast Earnings to Price (19%) Note1: EV (Enterprise Value) is given by: EV=Market Cap + LT Debt + max(ST Debt–Cash,0), where LT (ST) stands for long (short) term Note2: Forecast Earnings are calculated from Bloomberg earnings consensus estimates data. We focus on the 1-year and 2-year forward earnings because data coverage drops off for longer horizons.	BEST_EPS BS_LT_BORROW BS_ST_BORROW BS_TOT_ASSET CF_CASH_FROM_OPER EBITDA EQY_SH_OUT IS_INC_BEF_XO_ITEM PX_LAST SALES_REV_TURN
Dividend Yield is another dimension of value, but distinct enough to be a standalone factor.	Most recently announced net dividend (annualized) divided by the current market price (100%)	EQY_DVD_YLD_IND_NET PX_LAST

Factor	Calculation	Bloomberg Data
Size is a composite metric distinguishing between large and small stocks.	Combination of the following descriptors: • log(Market Capitalization) (31%) • log(Sales) (34%) • log(Total Assets) (35%)	BS_TOT_ASSET EQY_SH_OUT PX_LAST SALES_REV_TURN
Trading Activity is a turnover based measure. Bloomberg focuses on turnover which is trading volume normalized by shares outstanding. This indirectly controls for the Size effect.	The exponential weighted average (EWMA) of the ratio of shares traded to shares outstanding: $TradingActivity_t = \sum_{s=504\ days}^{s=1\ day} \exp\left(-\gamma * (t-s)\right) * \frac{V_{t-s}}{O_{t-s}}$, where the average is over 2 years (504 trading days) with a half-life of 180 days ($\gamma = log\left(2\right)/180$), and V_s and O_s are shares traded and shares outstanding on day s , respectively. (100%) In addition, to mitigate the impacts of those sharp short-lived spikes in trading volume, Bloomberg winsorizes the data: first daily trading volume data is compared to the long-term EWMA volume (180 day half-life), then the data is capped at 3 standard deviations away from the EWMA average.	EQY_SH_OUT PX_VOLUME
Earnings Variability gauges how consistent earnings, cash flows, and sales have been in recent years.	Combination of the following descriptors: • Earnings Volatility to Total Assets (34%) $\frac{\text{Earnings Volatility over the last 5 years}}{\text{Median Total Assets over the last 5 years}}$ • Cash Flow Volatility to Total Assets (35%) $\frac{\text{Cash Flow Volatility over the last 5 years}}{\text{Median Total Assets over the last 5 years}}$ • Sales Volatility to Total Assets (31%) $\frac{\text{Sales Volatility over the last 5 years}}{\text{Median Total Assets over the last 5 years}}$	BS_TOT_ASSET CF_CASH_FROM_OPER IS_INC_BEF_XO_ITEM SALES_REV_TURN

<table>
<tr><th>Factor</th><th>Calculation</th><th>Bloomberg Data</th></tr>
<tr><td>Profitability studies firms' profit margins to differentiate between money makers and money losers.</td><td>Combination of the following descriptors:

• Return on Equity (26%)
$\frac{\text{Net Income}}{\text{Book Value}}$,

• Return on Assets (28%)
$\frac{\text{Net Income}}{\text{Total Assets}}$,

• Return on Capital Employed (28%)
$\frac{\text{Net Income}}{\text{Capital Employed}}$,

• EBITDA Margin (18%)
$\frac{\text{EBITDA}}{\text{Sales}}$.

Note: Net Income is that before extraordinary items.</td><td>BS_ACCT_PAYABLE

BS_TOT_ASSET

CAPITAL_EMPLOYED

CASH_AND_MARKETABLE_SECURITIES

EBITDA

IS_INC_BEF_XO_ITEM

SALES_REV_TURN

TOT_COMMON_EQY</td></tr>
<tr><td>Volatility differentiates more volatile stocks and less volatile ones by quantifying "volatile" from several different angles.</td><td>Combination of the following descriptors:

• Rolling Volatility = Return volatility over latest 252 trading days (29%)

• Rolling CAPM Beta = Regression coefficient from the rolling window regression of stock returns on local index returns (18%)

• Historical Sigma = Residual volatility from the rolling window regression of stock returns on local index returns (28%)

• Cumulative Range = The ratio of maximum and minimum stock price over the previous year (25%)

Note: Bloomberg adjusts the volatility factor exposure by regressing it on the rest of factor exposures, and standardizes the regression residual the usual way. This does not change the explanatory power of the model, but it makes the volatility factor more distinct by significantly reducing its correlation with other factors.</td><td>Index Returns

PX_LAST</td></tr>
</table>

Factor	Calculation	Bloomberg Data
Growth aims to capture the difference between high and low growers by using historical fundamental and forward-looking analyst data.	Combination of the following descriptors: • Total Asset Growth (23%) $\frac{\text{5-year average growth in Total Assets}}{\text{Average Total Assets over the last 5 years}}$, • Sales Growth (25%) $\frac{\text{5-year average growth in Sales}}{\text{Average Total Assets over the last 5 years}}$, • Earnings Growth (18%) $\frac{\text{5-year average growth in Earnings}}{\text{Average Total Assets over the last 5 years}}$, • Forecast of Earnings Growth (14%) $\frac{\text{2-year forecast EPS}}{\text{1-year forecast EPS}}$, • Forecast of Sales Growth (20%) $\frac{\text{2-year forecast Sales}}{\text{1-year forecast Sales}}$.	BEST_EPS BEST_SALES BS_TOT_ASSET IS_INC_BEF_XO_ITEM SALES_REV_TURN
Leverage is a composite metric to gauge a firm's level of leverage.	Combination of the following descriptors: • Book Leverage (34%) $\frac{\text{LT Debt} + \max(\text{ST Debt} - \text{Cash}, 0)}{\text{Book Value} + \text{LT Debt} + \max(\text{ST Debt} - \text{Cash}, 0)}$, • Market Leverage (33%) $\frac{\text{LT Debt} + \max(\text{ST Debt} - \text{Cash}, 0)}{\text{Market Cap} + \text{LT Debt} + \max(\text{ST Debt} - \text{Cash}, 0)}$, • Debt to Total Assets (33%) $\frac{\text{LT Debt} + \max(\text{ST Debt} - \text{Cash}, 0)}{\text{Total Assets}}$. Note: LT (ST) stands for long (short) term	BS_LT_BORROW BS_TOT_ASSET CASH_AND_MARKETABLE_ SECURITIES EQY_SH_OUT PX_LAST TOT_COMMON_EQY

Appendix B: List of Industry Factors

Industry Factor	BICS Codes
Media	1010
Telecom	1011
Automotive	1111
Office and Leisure Products	1112, 1113
Commercial and Consumer Services	1114, 1115
Gaming & Lodging	111610, 111611, 111612
Restaurants	111613
Passenger Transportation	1117
Recreation Facilities	1118
Apparel & Textiles	1110, 1119, 1120
Consumer Products	1210
Retail & Distributors	1211, 1212
Exploration & Production & Integrated Oil & Services	131010, 131011, 131016, 131015
Midstream Oil and Gas	131012, 131013
Oil Refining & Marketing	131014
Investment Companies & Diversified Banks	141010, 141110, 141111
Multiple Finance	141011, 141013, 141210, 141214, 1413, 141012, 141213
Mortgage Finance	141212
Consumer Finance	141211
Life Insurance	141410
Insurance Excluding Life Insurance	141411, 141412, 141413, 141414, 141415
Real Estate Owners & Services	141510, 141512
REIT	141511
Biotech & Pharma	1510
Health Care Facilities and Services	151110, 151111, 151112
Managed Care	151113
Medical Equipment	1512
Aerospace & Defense	1610
Electrical Equipment & Machinery	1611, 1612, 1613
Transportation	1614, 1617
Engineering, Industrial, and Waste Services	1615, 1616, 1618
Chemicals	1710
Construction Materials	1711
Containers & Packaging Materials	1712, 1713
Iron & Steel	1714
Metals & Mining	1715
Hardware & Semiconductors	1810, 1811, 1813, 1311
Software & Technology Services	1812, 1814
Utilities	1910

Appendix C: List of Country Factors

Country Factor	Country Factor	Country Factor
Argentina	Eastern Europe Developed	Emerging Latin America
Australia	Hungary	Bolivia
Austria	Poland	Colombia
Brazil	Czech Republic	Ecuador
Belgium	Emerging Europe	Jamaica
Canada	Albania	Peru
Chile	Belarus	Trinidad / Tobago
China	Bosnia	France
Denmark	Herzegovina	France
Finland	Bulgaria	Luxembourg
Germany	Croatia	Monaco
Greece	Cyprus	Frontier
Hong Kong	Estonia	Bangladesh
India	Latvia	Kazakhstan
Indonesia	Lithuania	Pakistan
Ireland	Macedonia	Sri Lanka
Israel	Romania	Vietnam
Italy	Serbia	Great Britain
Japan	Slovakia	Great Britain
Korea	Slovenia	Gibraltar
Malaysia	Ukraine	Guernsey
Mexico	Emerging Middle East	Isle of Man
Netherlands	Egypt	Jersey
New Zealand	Bahrain	British Virgin Islands
Norway	Jordan	Switzerland
Philippines	Morocco	Switzerland
Portugal	Tunisia	Liechtenstein
Russia	Bahrain	US
Singapore	Kuwait	US
Spain	Lebanon	Bermuda
South Africa	Oman	Bahamas
Sweden	Qatar	Cayman Islands
Turkey	Saudi Arabia	
Taiwan		
Thailand		

Note: Bloomberg whitepapers may be found by typing "PORT HELP <GO>" to access the PORT help page.

Contact Us

To learn more about Bloomberg's Portfolio & Risk Analytics, contact your Bloomberg account representative or press the <HELP> key twice on the Bloomberg Professional® service.

>>>

BEIJING
+86 10 6649 7500

FRANKFURT
+49 69 9204 1210

LONDON
+44 20 7330 7500

NEW YORK
+1 212 318 2000

SÃO PAULO
+55 11 2395 9000

SYDNEY
+61 2 9777 8600

DUBAI
+971 4 364 1000

HONG KONG
+852 2977 6000

MUMBAI
+91 22 6120 3600

SAN FRANCISCO
+1 415 912 2960

SINGAPORE
+65 6212 1000

TOKYO
+81 3 3201 8900

bloomberg.com