

Quantopian Risk Model

Abstract

Risk modeling is a powerful tool that can be used to understand and manage sources of risk in investment portfolios. In this paper we lay out the logic and the implementation of the Quantopian Risk Model (QRM), an equity risk factor model developed by Quantopian to decompose and attribute risk exposures taken on by arbitrary equity investment strategies. By defining sources of risk, it is possible to consider the residual or remainder term as a strategy's "alpha", or the portion of an investment strategy's return that is derived from skill. Combined with some other tools, the QRM is used by Quantopian to evaluate quantitative trading strategies on the basis of observing that strategy's portfolio holdings over time.

Introduction

Risk management considers how we can consciously determine how much risk we are willing to take in order to attain a future gain. This process involves:

1. Identifying sources of risk
2. Quantifying exposure to those sources of risk
3. Determining the effects of those risk exposures
4. Developing a mitigation strategy
5. Observing consequent performance and amending the mitigation strategy as needed

Risk comes from uncertainty regarding a portfolio's future losses and gains and different individuals and institutions will have differing amounts of tolerance to risk. We quantify the total risk of a portfolio over T time periods as the standard deviation of that portfolio's returns:

$$\sigma = \sqrt{\frac{1}{T} \sum_{l=1}^T (r_l - \bar{r})^2},$$

Where:

- r_l is the return on time l
- \bar{r} is the average portfolio return over T time periods

This is a common definition of risk that captures sufficient information for our purposes. It treats gains and losses symmetrically and can be used to evaluate each level of the portfolio, from individual assets to the portfolio itself. Standard deviation, also called volatility, gives us an idea of how close we should expect values to fall relative to the mean. A common rule of thumb is that around 68% of values lie within one standard deviation of the mean, 95% of values lie within two standard deviations of the mean, and 99% of values lie within three standard deviations of the mean. A population of observations with a low standard deviation will contain individual observations largely clustered around the population mean value, while a population

Quantopian Risk Model

of observations with a higher standard deviation can be expected to contain a larger number of extreme values, both for gains and losses. This fits with common intuition about financial returns. More extreme values go hand in hand with more volatile assets. They bring with them greater chance of both profit and ruin.

Evaluating risk is not only about evaluating the amount of potential loss. It allows us to set reasonable expectations for returns and make well-informed decisions about potential investments. Quantifying the sources of risk associated with a portfolio can reveal to what extent the portfolio is actually accomplishing a stated investment goal. If an investment strategy is described as targeting market and sector neutrality, for example, the underlying portfolio should not be achieving significant portions of its returns from a persistent long exposure to the technology sector. While this strategy may show profit over a given timeframe, understanding that those profits are earned on the basis of unintended bets on a single sector may lead the investor to make a different decision about whether and how much capital to allocate. Quantifying risk exposures allows investors and managers to create risk management strategies and refine their portfolio.

Developing a risk model allows for a clear distinction between **common risk** and **specific risk**. Common risk is defined here as risk attributable to common factors which drive returns within the equity stock market. These factors can be composed of either fundamental or statistical information about the underlying investment assets that make up the market. Fundamental factors are often observable fundamental ratios reported by companies that issue stock, such as the ratio of book value to share price, or earnings per share. These factors are typically derived from financial and macroeconomic sources of data. Statistical factors use mathematical models to explain the correlations between asset returns time-series without consideration of company-specific fundamental data (**Axioma, Inc. 2011**).

Some commonly-cited risk factors are the influence of an overall market index, as in the Capital Asset Pricing Model (CAPM) (**Sharpe 1964**), risk attributable to investing within individual sectors, which give an idea of the space a company works within, as in the BARRA risk model (**BARRA, Inc. 1998**), or style factors, which mimic investment styles such as investing in “small cap” companies or “high growth” companies, as in the Fama-French 3-factor model (**Fama and French 1993**).

Specific risk is defined here as risk that is unexplainable by the common risk factors included in a risk model. Typically, this is represented as a residual component left over after accounting for common risk (**Axioma, Inc. 2011**). When we consider risk management in the context of quantitative trading, our understanding of risk is used in large part to clarify our definition of “alpha”. This residual after accounting for the common factor risk of a portfolio can be thought of as a proxy for or estimate of the alpha of the portfolio.

Quantopian Risk Model

Factor Models

A common first introduction to factor modeling and the notion of common factor risk is the Capital Asset Pricing Model (CAPM). In the CAPM, we define an equilibrium relationship between returns and common factor risk, using only a single common factor risk, the returns of the market itself. The CAPM expresses the returns of any individual asset like so:

$$E[r_i] = r_f + \beta_i^M (r_M - r_f)$$

Where:

- r_f is the risk-free rate
- r_M is the return of the market
- $\beta_i^M = \frac{COV[r_i, r_M]}{VAR[r_M]}$ is the influence the market on the excess return of the i -th asset over the market

The CAPM decomposes the returns of individual assets into the portion of the returns explained by the market at large and the portion of returns expected from a risk-free asset. This is a simplistic view of the market that does not hold up in empirical tests. Many improvements upon the CAPM have been developed since its inception, such as the Fama-French three-factor model (**Fama and French 1993**), which extends from a single source of risk to include two additional sources of risk, company size and company value. The Fama-French three factor model can be expressed as:

$$r_i = r_f + \beta_i^M (r_M - r_f) + \beta_i^{SMB} r_{SMB} + \beta_i^{HML} r_{HML} + \alpha_i$$

Where:

- r_{SMB} is the return from the risk premium of small market cap stocks over large market cap stocks (SMB)
- r_{HML} is the return from the risk premium of high book-to-market ratio stocks over low book-to-market ratio stocks (HML)
- β_i^{SMB} is the risk exposure of r_i to SMB
- β_i^{HML} is the risk exposure of r_i to HML
- α_i is the unexplained return of asset i

As a final example, Arbitrage Pricing Theory (APT) (**Ross 1976**) is a generalization of the CAPM and similar models, which allows us to measure the influence of more than one factor when considering the forces that drive returns.

Quantopian Risk Model

APT expresses the returns of individual assets using a multiple linear regression, a linear factor model, like so:

$$r_i = \alpha_i + \beta_{i,0}F_0 + \beta_{i,1}F_1 + \dots + \beta_{i,m}F_m + \epsilon_i$$

Where:

- $F_j, j \in \{0, m\}$ is the set of return streams of the common factors in our model
- $\beta_{i,j}, j \in \{0, m\}$ is the set of risk exposures of r_i to each common factor risk
- α_i is the unexplained return of asset i
- ϵ_i the idiosyncratic shock of asset i

The factors in APT are return streams that are entirely dominated by single characteristics. In the Fama-French model, our ways to explain returns are limited to the market, SMB, and HML, while with APT we can add as many factors as we want in order to account for the various common factors that are relevant to us. APT forms the basis of the Quantopian Risk Model.

Implementation

The QRM is a multi-factor risk model which seeks to decompose each asset's returns across a suite of 16 individual fundamental factors. The 16 factors in the model are comprised of 11 sector factors and 5 style factors. The QRM does not explicitly model a standalone market factor. The factors included as common risk factors were chosen for their degree of independence from each other while seeking to explain the returns of the largest number of assets in the market possible.

The QRM design is by definition designed to model historical and current risk as opposed to serving as a risk forecasting tool.

This section begins the technical implementation of the QRM.

Notation Guide

Summary of basic notational conventions:

- Boldface lowercase letter indicates vector, e.g. \mathbf{a}
- Boldface uppercase letter indicates matrix, e.g. \mathbf{A}
- Boldface uppercase letter with subscript \cdot, j indicates the j^{th} column of matrix \mathbf{A} , e.g. $\mathbf{A}_{\cdot,j}$
- Boldface uppercase letter with subscript i, \cdot : indicates the i^{th} row of matrix \mathbf{A} , e.g. $\mathbf{A}_{i,\cdot}$

Quantopian Risk Model

Mathematical Model

Mathematically, the QRM has the following form,

$$r_{i,t} = \sum_{j=1}^n \beta_{i,j,t}^{sect} f_{j,t}^{sect} + \sum_{k=1}^m \beta_{i,k,t}^{style} f_{k,t}^{style} + \varepsilon_{i,t}, \quad (1)$$

where:

- $r_{i,t}$ is the return of asset i on day t ,
- n is the number of sector factors,
- m is the number of style factors,
- $\beta_{i,j,t}^{sect}$ is the j^{th} sector factor exposure¹ of asset i on day t ,
- $f_{j,t}^{sect}$ is the return of j^{th} sector factor on day t ,
- $\beta_{i,k,t}^{style}$ is the k^{th} style factor exposure of asset i on day t ,
- $f_{k,t}^{style}$ is the return of k^{th} style factor on day t .
- $\varepsilon_{i,t}$ is the residual term for asset i on day t in model (1)

The mathematical model (1) is derived from sub-model (1a)

$$r_{i,t} = \sum_{j=1}^n \beta_{i,j,t}^{sect} f_{j,t}^{sect} + \varepsilon_{i,t}^{sect}, \quad (1a)$$

and sub-model (1b)

$$\varepsilon_{i,t}^{sect} = \sum_{k=1}^m \beta_{i,k,t}^{style} f_{k,t}^{style} + \varepsilon_{i,t}, \quad (1b)$$

where:

- $\varepsilon_{i,t}^{sect}$ is the residual term for asset i on day t in sub-model (1a).

¹ Factor exposure is also called a factor loading. It measures the relationship between the dependent variable and the underlying factor.

² For stock i , the $\beta_{i,j,t}^{sect}$ is zero if stock i does not belong to the j^{th} sector.

Quantopian Risk Model

Sector Factors

Sector factors are used to represent the influence of different sectors. The QRM defines sectors using sector classifications as defined by Morningstar (Morningstar, Inc. n.d.). Furthermore, the QRM uses sector ETF returns to represent the corresponding sector factor returns. The following table maps each sector to its index in the mathematical model (1), corresponding ETF, Morningstar sector code, Quantopian security identifier (SID), and variable name used in the Quantopian API, respectively.

Sector	Sector index j	ETF	Morningstar code	SID	Variable name in the Quantopian API
Materials sector	1	XLB	101 - Basic Materials	19654	materials
Consumer Discretionary SPDR	2	XLY	102 - Consumer Cyclical	19662	consumer_discretionary
Financials	3	XLF	103 - Financial Services	19656	financials
Real Estate	4	IYR	104 - Real Estate	21652	real_estate
Consumer Staples	5	XLP	205 - Consumer Defensive	19659	consumer_staples
Health Care	6	XLV	206 - Healthcare	19661	health_care
Utilities	7	XLU	207 - Utilities	19660	utilities
Telecom	8	IYZ	308 - Communication Services	21525	telecom
Energy	9	XLE	309 - Energy	19655	energy
Industrials	10	XLI	310 - Industrials	19657	industrials
Technology	11	XLK	311 - Technology	19658	technology

Each sector factor's returns are known and every asset in Quantopian's database is mapped to at most one single sector. Therefore, only the sector factor exposures need to be estimated.

Quantopian Risk Model

Style Factors

The QRM includes 5 style factors: momentum, size, value, short term reversal, and volatility. Each style factor is designed to replicate a traditional investment strategy. The following table maps each style factor to its index in the mathematical model (1) and variable name in the Quantopian API.

Style factor name	Style index k	Variable name in the Quantopian API
momentum	1	momentum
size	2	size
value	3	value
short term reversal	4	reversal_short_term
volatility	5	volatility

Style Factor Definitions

Momentum

The momentum factor captures the difference in returns between stocks on an upswing (winner stocks) and stocks on a downswing (loser stocks) over a trailing 11-month period.

Size

The size factor captures the difference in returns between large-cap stocks and small-cap stocks.

Value

The value factor captures the difference in returns between inexpensive stocks and expensive stocks.

Volatility

The volatility factor captures the difference in returns between high volatility stocks and low volatility stocks.

Short-term reversal

The short-term reversal factor captures the difference in returns between stocks with strong recent losses theoretically primed to reverse (recent loser stocks) and stocks with strong recent gains theoretically primed to reverse (recent winner stocks) in a short time period.

Quantopian Risk Model

Style Factor Metric Formulas

The style factor metrics are used to describe the style factors. Below, we provide the mathematical definition for each style factor metric in the QRM.

Momentum

The momentum metric of an asset i , MOMENTUM, on day t is computed by calculating the 11-month cumulative return from 12 months ago to 1 month ago³. The formula is

$$\text{MOMENTUM} = \prod_{l=-c*12+t-1}^{-c+t-1} (1 + r_{i,l}),$$

Where:

- $r_{i,l}$ is return of asset i on day l
- c is the number of trading days in one month⁴

Size

The size metric of asset i , SIZE, on day t is computed by calculating the \log of its company's market capitalization. The formula is

$$\text{SIZE} = \log(M_{i,t-1})$$

Where:

- $M_{i,t-1}$ is the market capitalization of asset i on day $t - 1$ ⁵

³ To avoid look-ahead bias, all the style factor metrics are lagged by one day.

⁴ Here, the constant c is set as 21.

⁵ The companies' financial data used on Quantopian is Morningstar fundamental data accessed through the Pipeline API.

Quantopian Risk Model

Value

The value metric of asset i , VALUE, on day t is computed by calculating the ratio of the company's stockholders' equity and market capitalization. The formula is

$$\text{VALUE} = \frac{S_{i,t-1}}{M_{i,t-1}}$$

Where:

- $S_{i,t-1}$ is the company's stockholder's equity of asset i on day $t - 1$

Short-term reversal

The short-term reversal metric of asset i , STR, on day t is computed by calculating the negative relative strength index (RSI). The formula is

$$\text{STR} = -1 * \text{RSI}_{t-1}$$

Where:

- RSI_{t-1} is relative strength index on a 14-day time frame from day $t - 1$ to $t - 15$.

Volatility

The volatility of asset i , VOL, on day t is computed by calculating the trailing 6 month return volatility. The formula is

$$\text{VOL} = \sqrt{\frac{1}{6c} \sum_{l=t-1}^{-6c-1+t} (r_{i,l} - \bar{r}_i)^2}$$

Where:

- c is number of trading days in one month⁴
- \bar{r}_i is the mean return of asset i in time period $(-6c - 1 + t, t - 1)$

Quantopian Risk Model

Methodology

As introduced, the QRM consists of two sub-models. Sub-model (1a) estimates the sector factor exposures of all stocks using linear regression and passes the residual returns $\boldsymbol{\varepsilon}_{i,t}^{sect}$ to sub-model (1b). Then, sub-model (1b) uses $\boldsymbol{\varepsilon}_{i,t}^{sect}$ as input to estimate the style factor returns associated with style factor exposures.

Sector Factor Calculation

The QRM estimates the sector factor exposure of each asset using a trailing 2-year window of stock returns and its respective sector factor returns. The procedure is as follows.

for each stock i on day t :

1. Find the Morningstar sector code of stock i
2. Choose the sector ETF that matches the Morningstar sector code of stock i
3. Compute the 2-year trailing historical returns of stock i on day t , and format them in a vector column \mathbf{r}_i
4. Compute the 2-year trailing historical returns of selected ETF from step 2 on day t , and format them in a vector column \mathbf{f}
5. Regress the vector column \mathbf{r}_i on the vector column \mathbf{f}
6. Obtain the regression coefficient, β , and set it as the respective sector factor exposure
7. Set other sector factor exposures as zeros
8. Compute the 2-year trailing historical sector residual returns, $\boldsymbol{\varepsilon}_i^{sect}$, by subtracting $\beta\mathbf{f}$ from \mathbf{r}_i

For example, let t be 2013-JAN-02, stock i be AAPL, then

- the selected ETF would be XLK
- the vector column \mathbf{f} would be the daily returns of XLK from 2010-DEC-31 to 2013-JAN-02
- the vector column \mathbf{r}_i would be the daily returns of AAPL from 2010-DEC-31 to 2013-JAN-02
- the β would be the technology sector factor exposure of AAPL
- the vector column, $\boldsymbol{\varepsilon}_i^{sect}$, that equals to $\mathbf{r}_i - \beta\mathbf{f}$ would be the trailing historical sector returns from 2010-DEC-31 to 2013-JAN-02

Quantopian Risk Model

If we plug this example into the sub-model (1a), it can be written as

$$r_{i,t} = \sum_{j=1}^{n-1} \beta_{i,j,t}^{sect} f_{j,t}^{sect} + \beta f_t + \varepsilon_{i,t}^{sect},$$

where

- f_t is the last entry of f
- $\varepsilon_{i,t}^{sect}$ is the last entry of ε_i^{sect}
- the term $\sum_{j=1}^{n-1} \beta_{i,j,t}^{sect} f_{j,t}^{sect}$ equals to 0

Style Factor Calculation

To estimate the returns of the style factor, it is not proper to use all stocks in the market. We need to define a universe, the estimation universe, which can represent the market while excluding ‘problematic’ assets such as REITs, ADRs, illiquid stocks, etc. Choosing the stocks in an estimation universe is subjective. The estimation universe in the QRM has about 2100 stocks. The selection criteria include:

- being a common stock
- having enough data to compute style factor metrics
- being in the top 3000 most liquid stocks

The stocks outside the estimation universe are called complementary stocks. The universe that includes both the stocks in the estimation universe and the complementary stocks is called the coverage universe. We will demonstrate how to compute the style factor exposures of stocks in the estimation universe, how to estimate style factor returns, and how to compute the style factor exposures of complementary stocks.

Style factor exposures of stocks in estimation universe

The style factor exposures of the stocks in the estimation universe on day t are calculated by z-scoring the style factor metrics of the stocks on day t . They are standardized (z-scored) with respect to the estimation universe.

Estimating style factor returns

The style factor returns are estimated day-by-day for two years using a cross-sectional regression.

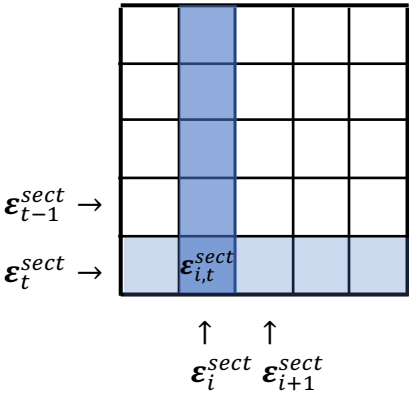
Quantopian Risk Model

For each day t in trailing two years:

1. Calculate the 5 style factor exposures of the stocks in the estimation universe, and store them in the columns of a matrix B .
2. Collect day t 's sector residuals of stocks in the estimation universe and form them in a column vector $\boldsymbol{\varepsilon}_t^{sect}$
3. Regress the column vector $\boldsymbol{\varepsilon}_t^{sect}$ on the matrix B
4. Obtain 5 style factor returns from the coefficients of the regression, $f_{1,t}^{style}, f_{2,t}^{style}, \dots$, and $f_{5,t}^{style}$
5. Generate 5 vector columns, $f_k^{style} (k = 1, 2, \dots, 5)$, and store $f_{k,t}^{style}$ in f_k^{style}
6. Collect the residual returns of the stocks in the estimation universe on day t by subtracting $\sum_{k=1}^5 B_{:k} f_{k,t}^{style}$ from $\boldsymbol{\varepsilon}_t^{sect}$

Figure 1 shows the relationship among $\boldsymbol{\varepsilon}_t^{sect}$, $\boldsymbol{\varepsilon}_i^{sect}$, and $\boldsymbol{\varepsilon}_{i,t}^{sect}$ in a matrix form.

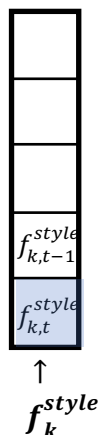
Figure 1



Quantopian Risk Model

Figure 2 shows the relationship between f_k^{style} and f_{kt}^{style} .

Figure 2



Style factor exposures of complementary stocks

Style factor exposures for complementary stocks are calculated by solving a time-series multi-linear regression of 5 style factor returns with the sector residuals. Here, we use a 2-year window of return series of the style factor returns and sector residuals. The procedure is as follows.

For each complementary stock i on day t :

1. Collect the style factor returns f_k^{style} , $k = 1, 2, \dots, 5$
2. Collect 2 years tailing historical sector residual returns, ϵ_i^{sect}
3. Run multi-linear regression with dependent variable ϵ_i^{sect} and independent variables, f_k^{style} , $k = 1, 2, \dots, 5$
4. Obtain the regression coefficients, $\beta_{k,t}^{style}$ ($k = 1, 2, \dots, 5$), and set them as corresponding style factor exposure
5. Compute 2 years tailing historical residual returns, ϵ_i , by subtracting $\sum_{k=1}^5 \beta_{i,k,t}^{style} f_k^{style}$ from ϵ_i^{sect}

Quantopian Risk Model

Risk Calculation

The risk of an asset i over T time periods is defined as

$$\sigma_i = \sqrt{\frac{1}{T} \sum_{l=1}^T (r_{i,l} - \bar{r}_i)^2} \quad (2)$$

where

- $r_{i,l}$ is the return of asset i on time l
- \bar{r}_i is the average return of asset i over T time periods

The risk of each factor return can be calculated directly by equation (2). For example, the risk of k^{th} style factor over T time periods is $\sqrt{\frac{1}{T} \sum_{l=1}^T (f_{k,l}^{style} - \bar{f}_k^{style})^2}$.

Similarly, the risk of each exposure weighed factor returns also can be calculated by equation (2). For example, the risk of k^{th} exposure weighed style factor over T time periods is

$$\sqrt{\frac{1}{T} \sum_{l=1}^T (\beta_{i,k,t}^{style} f_{k,l}^{style} - \beta_k^{style} f_k^{style})^2}$$

Summary and Conclusions

The Quantopian Risk Model is a 16-factor risk model built to aid our users and investment in the research and evaluation of high-quality trading algorithms. We use classical techniques in finance to compute risk exposures to each relevant factor for US equities. The risk model factors loadings and factor returns are fully available, for free, within the Quantopian Research environment and backtester on the Quantopian website. Further research is to come on common factors that could be useful in international markets.

Quantopian Risk Model

References

- Axioma, Inc. 2011. *Axioma Robust Risk Model Handbook*. Axioma, Inc.
- BARRA, Inc. 1998. *United States Equity*. BARRA, Inc.
- Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* (Journal of Financial Economics) 3-56.
- Morningstar, Inc. n.d. *Morningstar® Data for Equities*.
- Ross, Stephen A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 341-360.
- Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance* 425-442.

Quantopian Risk Model

Appendix

A. Portfolio turnover assumptions.

The periodicity of the data used for QRM is daily, which results in the underlying assumption that the minimum holding period for each asset is at least 1 day. An investment strategy with significant intraday trading or a very high turnover rate would not be appropriate to analyze with the current QRM.

B. Instrument coverage

The QRM covers about 4000 stocks in US stock market, but it does not all of the assets. If a portfolio has a sizeable weight invested in assets outside of the coverage universe, it is not proper to analyze it with QRM.

The current QRM does not cover the ETFs except the sector ETFs and a few pre-selected ETFs (that are used for testing).

C. Summary of calculations

Factor type	<i>Stock type?</i>	Factor exposures	Factor returns
Sector	<i>Stocks in coverage universe</i>	Time-series linear regression	Given sector ETFs
Style	<i>Stocks in estimation universe</i>	Normalized risk metrics	Cross-sectional regression
	<i>Complementary stocks</i>	Time-series linear regression	Time-series linear regression