The Influence of Exogenous Capital on Asset Return

Victor Kolev & Stefan Hadzhistoykov

Mentor:

Rafael Rafailov

Abstract

The principal aim of our research is to analyze the exogenous factor and its effect on asset return. It is a financial factor, which accounts for assets that are currently not tradeable on the market. These assets, which we call exogenous capital, do have an, albeit indirect, effect on the market, since they dictate investor decisions. An example of exogenous capital are restricted shares, often awarded as employee compensation. Notwithstanding their influence, they are ignored by the vast majority of asset pricing models and that motivated our investigation into their properties.

A mathematical model is constructed, which describes a stock market system. Utilizing behavioral mathematics to simulate investor decisions, the optimal portfolio for each agent is derived, thus cognizing the market demand, which dictates the market equilibrium prices, given the total supply. Based on the Capital Asset Pricing Model, we linearly relate the exogenous factor, contained within the system described above, to asset return. Through calculus we arrive at the hypothesis that the exogenous factor is a source of additional market return, as it increases yield.

We continue with an empirical study to confirm this postulation by statistical analysis – linear regression models on panel data with corrections for heteroscedasticity. The data is comprised of prevalent economic factors about 3000 companies across a period of 10 years, including the exogenous factor from multiple time moments. By using generalized least squares to regress the daily return of each company's stock, we estimate the beta coefficient of each factor. It is proven that every beta represents the profit of its respective factor-mimicking portfolio. Analysis of the return of this portfolio for the exogenous factor results in clearly positive yield with marginal variance, meaning that the factor truly is a statistically significant source of additional asset return, hence confirming our hypothesis.

The results from this research into exogenous capital, namely its ability to improve yield, could be employed by quantitative funds, which rely on factor investing to construct their portfolios the and would therefore have the potential to increase profit.

Keywords: Mathematical Finance, Risk Factor, Multi-Factor Model, Factor-mimicking portfolio, Restricted shares

Contents

1 Introduction

The principal objective of our research is to analyze the economic effects of exogenous capital and elucidate upon its impact on capital asset return. Let us start with the definition of exogenous capital: it refers to all shares that are currently no tradeable on the market. The most prevalent form of this type of assets are the restricted shares that company executives often receive as bonuses.

The intuition that provoked this research paper is the following: when investors have exogenous capital, it limits the amount of demand they can exert on the market. Therefore, there is a shortage of demand, which implies a drop in prices. Subsequently, for investors without exogenous capital, it may become viable to take up the additional risk and invest further in return for higher expected yield. In this manner we suppose that exogenous capital increases profits on portfolios, all else equal.

In our investigation of the exogenous factor (which is a measurement of exogenous capital), we begin by mathematically modeling the stock market. Behavioral mathematics are utilized to cognize the optimal portfolio of each investor, considering their tolerance for risk (risk aversion). We derive a condition for market equilibrium, under which we prove a formula, which linearly relates the return of a particular asset to the market return, whilst accounting for the exogenous factor (based on the Capital Asset Pricing Model). Through the mathematical model we derive that the exogenous factor is a distinct source of additional asset return.

This theoretical hypothesis is challenged by empirical means. We perform statistical analysis on the effect of exogenous capital on the US stock market for the last 10 years. Linear regression models on panel data for 3000 publicly-traded companies are performed with corrections for heteroscedasticity. It is proven that from this data, a time series of returns of a so called factor-mimicking portfolio can be derived. The portfolio, influenced solely by the exogenous factor, is analyzed and shown to be a profitable investment, therefore reinforcing our hypothesis that exogenous capital is a statistically significant anomaly, distinct from other established risk factors and, moreover, it is a source of additional asset return.

The findings from this paper could be exploited by quantitative funds, which utilize such economic factors to construct their portfolios, thereby increasing their profit and improving the yield of their customers.

2 Definitions

We shall begin by introducing some of the core terminology:

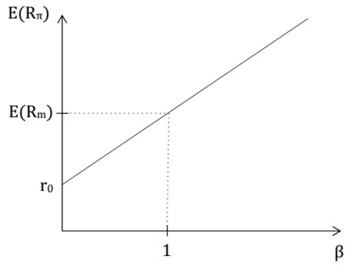
- Return on a portfolio – the profit, derived from selling stocks, divided by the investment; exogenous capital is not considered an investment, rather it is endowed

- Exogenous capital – capital that cannot be traded, which the agent possesses prior to time 0; it is considered endowment; an example of exogenous capital are restricted shares, which are often given as employee compensation

- Risk factor - a common financial factor, which affects stock returns; ex. market return, size premium, value premium, momentum, industrial production

- CAPM - Capital asset pricing model [1]; CAPM is widely utilized for forecasting asset return. It describes a linear relationship between the return on a given asset and the market return, the paramount source of yield for capital assets:

$$E(R_j) = r_0 + \beta E(R_m - r_0)$$

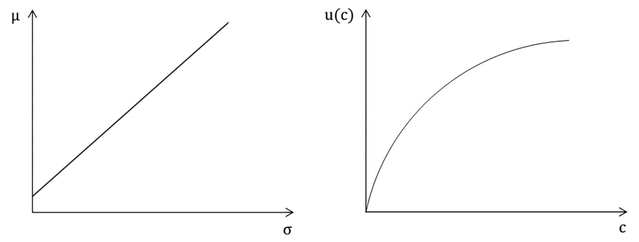


where j is the particular asset, $E[R_m]$ is the market return, r_0 represents a risk-free asset (defined later in the paper), and β is a coefficient, specific to the asset, also known as a risk premium (estimated through Fama-Macbeth regressions).
The underlying intuition behind the beta coefficient is that it is representative of the so called unsystematic risk of the asset compared to the systematic risk of the market; since the market is the primary source of an asset's return, the relative volatility determiners the relationship between the asset's and the market's payouts.

- utility function - a function that denotes the relative utility (satisfaction, preference) of the agent from the argument, in this case that is certain wealth, subject to risk

- Constant absolute risk aversion utility (CARA utility) [2]- it is defined as

$$u_i(c) = 1 - e^{-ac}$$



a - coefficient of risk aversion
c - commodity

The coefficient of absolute risk aversion indicates the demanded additional return for the agent to accept a further unit of risk. It is defined through the following differential equation:

$$A(c) = -\frac{u_i''(c)}{u_i'(c)}$$

We utilize constant absolute risk aversion, therefore we have:

$$A(c) = a, \qquad a - const$$

The solutions of this equation are the function class described above, as we can see here:

$$u_i'(c) = (-a)e^{-ac}$$
$$u_i''(c) = a^2 e^{-ac}$$
$$\Rightarrow -\frac{u_i''(c)}{u_i'(c)} = a$$

- Nash equilibrium - a state of a system of agents, which is unique in that no agent can benefit by a change of strategy, assuming the other's remain unaltered
- systematic risk (undiversifiable risk) - risk that affects the overall market and is therefore unavoidable and unpredictable; common causes include, but are not limited to, recessions, natural disasters, law alterations and inflation
- unsystematic risk - risk, inherent in a unique stock and therefore able to be mitigated through portfolio diversification (investing in a variety of assets)

3 Mathematical Model

3.1 System Description

We model the stock market as follows:

- m agents (investors), each with wealth W_i^0 at time 0 and coefficient of risk aversion γ_i, who wish to invest in the stock market
- each agent optimizes a CARA utility
- n assets with prices $p_0 \in \mathbb{R}^n$ at time 0 and a total supply of shares $\Theta \in \mathbb{R}^n$
- at time 1 the assets will pay out $x \sim \mathcal{N}(p_1, \Sigma)$, $x, p_1 \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$; this distribution is known to all agents
- $g_i \in \mathbb{R}^n$ - endowment vector of shares for agent i (exogenous capital); $\theta_i \in \mathbb{R}^n$ - portfolio of the agent i
- r_0 - risk-free asset, at which the agents can borrow to invest or deposit safely; $R_f := r_0 + 100\%$; most commonly used are either interest rates of bank deposits or the yield of treasury bonds

3.2 Portfolio Management

The wealth of each agent at time 1 is:

$$W_i^1 = (g_i + \theta_i)^T x + R_f(W_i^0 - p_0^T \theta_i)$$

Given that every agent seeks maximum utility with consideration for his specific risk aversion coefficient, we have $\max E[u(W_i^1)]$ or

$$\max E[-e^{-\gamma_i W_i^1}]$$

Upon substituting W_i^1 with its partitioned value, we get:

$$\begin{aligned}
\max E[-e^{-\gamma_i W_i^1}] &\iff \max E[-e^{-\gamma_i((g_i+\theta_i)^T x + R_f(W_i^0 - p_0^T\theta_i))}] \\
&\iff \max -e^{-\gamma_i R_f(W_i^0 - p_0^T\theta_i)} . e^{-\gamma_i (g_i+\theta_i)^T p_1 + \frac{\gamma_i^2}{2}(g_i+\theta_i)^T \Sigma (g_i+\theta_i)} \\
&\iff \max -e^{-\gamma_i R_f W_i^0 + \gamma_i R_f p_0^T \theta_i} . e^{-\gamma_i (g_i+\theta_i)^T p_1 + \frac{\gamma_i^2}{2}(g_i+\theta_i)^T \Sigma (g_i+\theta_i)} \\
&\iff \max -e^{-\gamma_i (g_i+\theta_i)^T (p_1 - R_f p_0) + \frac{\gamma_i^2}{2}(g_i+\theta_i)^T \Sigma (g_i+\theta_i)} \\
&\iff \max \gamma_i (g_i+\theta_i)^T (p_1 - R_f p_0) - \frac{\gamma_i^2}{2}(g_i+\theta_i)^T \Sigma (g_i+\theta_i) \\
&\iff \max \frac{1}{\gamma_i} (g_i+\theta_i)^T (p_1 - R_f p_0) - \frac{1}{2}(g_i+\theta_i)^T \Sigma (g_i+\theta_i)
\end{aligned}$$

We have, in this manner, obtained the optimization function for our agent, which we shall rewrite as:

$$\min_{\theta_i} : \frac{1}{2}(g_i+\theta_i)^T \Sigma (g_i+\theta_i) - (g_i+\theta_i)^T \frac{p_1 - R_f p_0}{\gamma_i}$$

the function in question is quadratic with respect to the vector $(g_i + \theta_i)$. Let $Q := (g_i + \theta_i)$.

$$\Rightarrow \min \frac{1}{2} Q^T \Sigma Q + c^T Q$$

Where $c := -\frac{p_1 - R_f p_0}{\gamma_i}$. Solving this through quadratic optimization techniques, we get that the following condition is necessary and sufficient for reaching optimum:

$$\begin{aligned}
\Sigma Q^* + c = 0 &\iff \Sigma Q^* = -c \\
&\iff Q^* = \Sigma^{-1}(-c) \\
&\iff g_i + \theta_i = \Sigma^{-1} \frac{p_1 - R_f p_0}{\gamma_i}
\end{aligned}$$

Thus we have derived the optimal portfolio for each agent

$$\theta_i = \Sigma^{-1} \frac{p_1 - R_f p_0}{\gamma_i} - g_i$$

In this manner we derived the optimal portfolio for each agent.

3.3 Market Equilibrium

We are now primed to obtain the equilibrium prices, meaning those, under which the market is in equilibrium - demand is met by the supply, in mathematical terms: $\sum_{i=1}^m \theta_i = \Theta$; this is also a Nash equilibrium, for every agent holds an optimal portfolio.

$$\begin{aligned}
\sum_{i=1}^m \theta_i = \Theta &\iff \sum_{i=1}^m \Sigma^{-1} \frac{p_1 - R_f p_0^*}{\gamma_i} - g_i = \Theta \\
&\iff \Sigma^{-1}(p_1 - R_f p_0^*).(\sum_{i=1}^m \frac{1}{\gamma_i}) - \sum_{i=1}^m g_i = \Theta \\
&\iff \Sigma^{-1}(p_1 - R_f p_0^*).(\sum_{i=1}^m \frac{1}{\gamma_i}) = \Theta + \sum_{i=1}^m g_i \\
&\iff \Sigma^{-1} R_f p_0^*.(\sum_{i=1}^m \frac{1}{\gamma_i}) = \Sigma^{-1} p_1.(\sum_{i=1}^m \frac{1}{\gamma_i}) - (\Theta + \sum_{i=1}^m g_i) \\
&\iff p_0^* = \frac{p_1(\sum_{i=1}^m \frac{1}{\gamma_i}) - \Sigma(\Theta + \sum_{i=1}^m g_i)}{R_f. \sum_{i=1}^m \frac{1}{\gamma_i}}
\end{aligned}$$

We denote $\gamma := \sum_{i=1}^m \frac{1}{\gamma_i}$ and $G := \sum_{i=1}^m g_i$

$$\Rightarrow p_0^* = \frac{p_1}{R_f} - \frac{\Sigma(\Theta + G)}{\gamma R_f}$$

Under this set of prices the market is in equilibrium and we are ready to move further with the model. From now on, p_0 is equated to p_0^*.

3.4 Formula for Return Forecasting

We shall prove the following linear relation between the return on a given asset and the market return:

$$E[R_j] = \frac{\mathrm{Cov}(R_j, R_m)}{\mathrm{Var}(R_m)} \left(E[R_m] - \frac{r_0 G^T p_0 + G^T p_0}{\Theta^T p_0} - r_0 \right) + \frac{e_j^T G + r_0 e_j^T G}{e_j^T \Theta} + r_0$$

After simplification by denoting the variables:

- $\alpha_j := \dfrac{e_j^T G + r_0 e_j^T (\Theta + G)}{e_j^T \Theta}$
- $\beta_j := \dfrac{\mathrm{Cov}(R_j, R_m)}{\mathrm{Var}(R_m)}$
- $k := E[R_m] - \dfrac{r_0 (\Theta + G)^T p_0 + G^T p_0}{\Theta^T p_0}$

the formula becomes:

$$E[R_j] = \beta_j . k + \alpha_j$$

Let us first examine $E[R_j]$.

$$\begin{aligned}
E[R_j] &= \frac{e_j^T (\Theta + G) e_j^T p_1}{e_j^T \Theta e_j^T p_0} - 1 \\
\Longrightarrow E[R_j] - \alpha_j &= E[R_j] - \frac{e_j^T G + r_0 e_j^T (\Theta + G)}{e_j^T \Theta} = \\
&= \frac{e_j^T (\Theta + G) e_j^T p_1 - e_j^T \Theta e_j^T p_0}{e_j^T \Theta e_j^T p_0} - \frac{e_j^T G e_j^T p_0 + r_0 e_j^T (\Theta + G) e_j^T p_0}{e_j^T \Theta e_j^T p_0} = \\
&= \frac{e_j^T (\Theta + G) e_j^T p_1 - e_j^T \Theta e_j^T p_0 - e_j^T G e_j^T p_0 - r_0 e_j^T (\Theta + G) e_j^T p_0}{e_j^T \Theta e_j^T p_0} = \\
\Longrightarrow E[R_j] - \alpha_j &= \frac{e_j^T (p_1 - R_f p_0) e_j^T (\Theta + G)}{e_j^T \Theta e_j^T p_0}
\end{aligned}$$

If we show that $\beta_j k$ can be transformed to an identical expression, the formula will be proved.

$$\begin{aligned}
\mathrm{Var}\left(R_m\right) &= \mathrm{Var}\left(\frac{(\Theta + G)^T x}{\Theta^T p_0} \right) = \\
&= \left(\frac{1}{\Theta^T p_0} \right)^2 \mathrm{Var}\left((\Theta + G)^T x \right) = \\
&= \left(\frac{1}{\Theta^T p_0} \right)^2 (\Theta + G)^T \mathrm{Var}(x) (\Theta + G) = \\
&= \frac{(\Theta + G)^T}{\Theta^T p_0} \Sigma \frac{\Theta + G}{\Theta^T p_0}
\end{aligned}$$

$$\mathrm{Cov}(R_j, R_m) = \mathrm{Cov}\left(\frac{e_j^T x e_j^T(\Theta+G)}{e_j^T p_0 e_j^T \Theta}, \frac{(\Theta+G)^T x}{\Theta^T p_0}\right) =$$

$$= \frac{e_j^T(\Theta+G)}{e_j^T p_0 e_j^T \Theta} . \frac{1}{\Theta^T p_0} \mathrm{Cov}\left(e_j^T x, (\Theta+G)^T x\right) =$$

$$= \frac{e_j^T(\Theta+G)}{e_j^T \Theta} . \frac{e_j^T}{e_j^T p_0} \mathrm{Cov}(x,x) \frac{\Theta+G}{\Theta^T p_0} =$$

$$= \frac{e_j^T(\Theta+G)}{e_j^T \Theta} . \frac{e_j^T}{e_j^T p_0} \Sigma \frac{\Theta+G}{\Theta^T p_0}$$

Thus we have arranged both the denominator and numerator of β_j in an easy-to-work-with form and are ready to establish its value.

$$\beta_j = \frac{\mathrm{Cov}(R_j, R_m)}{\mathrm{Var}(R_m)} = \frac{\frac{e_j^T(\Theta+G)}{e_j^T \Theta} . \frac{e_j^T}{e_j^T p_0} \Sigma \frac{\Theta+G}{\Theta^T p_0}}{\frac{(\Theta+G)^T}{\Theta^T p_0} \Sigma \frac{\Theta+G}{\Theta^T p_0}} =$$

$$= \frac{e_j^T \Sigma(\Theta+G)}{(\Theta+G)^T \Sigma (\Theta+G)} . \frac{e_j^T(\Theta+G)\Theta^T p_0}{e_j^T p_0 e_j^T \Theta}$$

Only k remains unknown.

$$k = E[R_m] - \frac{r_0(\Theta+G)^T p_0 + G^T p_0}{\Theta^T p_0} = \frac{(\Theta+G)^T p_1}{\Theta^T p_0} - 1 - \frac{r_0(\Theta+G)^T p_0 + G^T p_0}{\Theta^T p_0} =$$

$$= \frac{(\Theta+G)^T(p_1 - R_f p_0)}{\Theta^T p_0}$$

Now all multipliers of the right hand side are cognized and its value can be determined

$$\beta_j k = \frac{e_j^T \Sigma(\Theta+G)}{(\Theta+G)^T \Sigma(\Theta+G)} . \frac{e_j^T(\Theta+G)\Theta^T p_0}{e_j^T p_0 e_j^T \Theta} . \frac{(\Theta+G)^T(p_1 - R_f p_0)}{\Theta^T p_0} =$$

$$= \frac{e_j^T \Sigma(\Theta+G)}{(\Theta+G)^T \Sigma(\Theta+G)} . \frac{e_j^T(\Theta+G)(\Theta+G)^T(p_1 - R_f p_0)}{e_j^T p_0 e_j^T \Theta}$$

Given that the market is in equilibrium, $\sum_{i=1}^m \theta_i = \Theta$ or:

$$\Theta = \Sigma^{-1}(p_1 - R_f p_0).\gamma - G \iff \Theta + G = \Sigma^{-1}(p_1 - R_f p_0).\gamma$$

We substitute this value of $\Theta + G$ in the expression.

$$\beta_j k = \frac{e_j^T \Sigma(\Theta+G)}{(\Theta+G)^T \Sigma(\Theta+G)} . \frac{e_j^T(\Theta+G)(\Theta+G)^T(p_1 - R_f p_0)}{e_j^T p_0 e_j^T \Theta} =$$

$$= \frac{e_j^T \Sigma \Sigma^{-1}(p_1 - R_f p_0)\gamma}{(\Theta+G)^T \Sigma \Sigma^{-1}(p_1 - R_f p_0)\gamma} . \frac{e_j^T(\Theta+G)(\Theta+G)^T(p_1 - R_f p_0)}{e_j^T p_0 e_j^T \Theta} =$$

$$= \frac{e_j^T(p_1 - R_f p_0)}{(\Theta+G)^T(p_1 - R_f p_0)} . \frac{e_j^T(\Theta+G)(\Theta+G)^T(p_1 - R_f p_0)}{e_j^T p_0 e_j^T \Theta} =$$

$$= \frac{e_j^T(p_1 - R_f p_0) e_j^T(\Theta+G)}{e_j^T p_0 e_j^T \Theta} = E[R_j] - \alpha_j$$

Q.E.D.

Actually the proven formula is equivalent to CAPM in the case of $G = 0$:

$$E[R_j] = \frac{\text{Cov}(R_j, R_m)}{\text{Var}(R_m)}(E[R_m] - r_0) + r_0$$

Exogenous capital is not traded on the market, thereby it is disregarded by most asset pricing models, including CAPM, despite having a substantial impact on individual portfolio management. The model we developed accounts for this factor and therefore improves upon the estimation of CAPM, all whilst maintaining its wide applicability, insofar as it uses solely elements known in the initial moment, videlicet Θ, G, p_0.

3.5 Effect of the Exogenous Factor

We analyze more deeply how the exogenous factor influences the market and its components.

Due to its inherent nature, exogenous capital restricts the demand investors are able to exert on the market and, therefore, it alters the market equilibrium. We examine how the exogenous capital introduces changes in the equilibrium prices and the expected return of assets.

$$p_0^* = \frac{p_1}{R_f} - \frac{\Sigma(\Theta + G)}{\gamma R_f}$$

From this we can conclude that, depending on the covariance matrix Σ, prices could not only be driven down by exogenous capital, but also increased. This reflects the fact that when exogenous factor is taken into account, it adds a further level of complexity to the system. Due to it, some assets may now become more viable investments and so their prices would be higher, because of the higher interest in them.

Let us now examine how the return of an asset changes.

$$E[R_j] = \frac{e_j^T(\Theta + G)e_j^T p_1}{e_j^T \Theta e_j^T p_0*} - 1$$

$$\frac{\partial E[R_j]}{\partial G_j} = \frac{e_j^T p_1 \left(e_j^T p_0^* + \frac{e_j^T(\Theta+G)\Sigma_{j,j}}{\gamma R_f} \right)}{e_j^T \Theta (e_j^T p_0^*)^2} > 0$$

$$\frac{\partial E[R_j]}{\partial G_i} = \frac{e_j^T p_1 e_j^T (\Theta + G)\Sigma_{j,i}}{\gamma R_f e_j^T \Theta (e_j^T p_0^*)^2}$$

Subsequently of the positivity of the partial derivative, we can conclude that increasing the exogenous factor leads to improved return of the asset, as well as other assets, strongly correlated with it (for example those within the same industry). We therefore draw the hypothesis that the exogenous factor is a distinct source of additional asset return for investors, which we now challenge.

4 Empirical Analysis

We continue with an empirical analysis of the exogenous factor. We have so far postulated that it is a source of asset return, which we now seek to confirm. To accomplish this task we perform linear generalized least squares regression on the daily returns of 3000 US stocks, modeling them as a linear function of risk factors (a multi-factor model). In this manner, we derive the beta of the exogenous factor, signifying its effect on asset return. These β-coefficients actually represent the return of a portfolio, which is hedged against all other market factors; in other words, it is dependent solely on the exogenous factor. It is shown that this portfolio is a worthwhile investment, consequently the exogenous factor is a statistically significant source of additional asset return, which is in unison which the hypothesis of the theoretical model.

4.1 Terminology

Here we define the factors, used in the regression models, as well any other terminology.

- size premium (SMB) [3]– a factor that measures the effect of market capitalization on the return; it accounts for the outperformance in returns of small stocks (companies with small market capitalization) over big stocks
- value premium (HML) [3]– it accounts for the better performance of the so called value stocks over growth stocks (this means that, in general, companies that are underpriced tend to generate higher return in the long run)
- momentum effect (MOM) [4]– it accounts for the anomaly that assets tend to adhere to their historical trend
- quality factor [5]– it measures the effect of stability and rigidness of the company on its stock return; companies with consistent growth and low debt tend to generate higher returns
- portfolio hedging – a financial operation that intends to reduce exposure to unsystematic risk; it is akin to insurance in the sense that it provides a safety net that prevents insufferable losses
- generalized least squares regression (GLS) - given a linear regression model:

  $$Y = X\beta + \varepsilon$$

  $Y \in \mathbb{R}^n$ - regressed variable; $X \in \mathbb{R}^{n \times k}$ - design matrix, in which the values of the observed parameters (factors for the companies' stocks) are stored; $\beta \in \mathbb{R}^k$ - beta coefficients of the regression, $\varepsilon \in \mathbb{R}^n$ - error terms.

GLS differs from standard linear regression in that it allows for heteroscedasticity, meaning that the error terms are correlated with one another - $Cov(\varepsilon_i, \varepsilon_j) = \Omega_{i,j}$ or

$$Var(\varepsilon) = \Omega$$

The β-coefficient is the solution to the following optimization problem:

$$min_\beta : \sqrt{\varepsilon^T \Omega^{-1} \varepsilon}$$

which is (see the Appendix for proof)

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y$$

4.2 Theoretical Background

We shall prove that each element of $\hat{\beta}$ represents the return of a so called factor-mimicking portfolio. Those portfolios are characterized by their quality that they are fully hedged against all factors, but the one in question.
We begin with a standard generalized least squares regression.

$$Y = X\beta + \varepsilon \qquad Var(\varepsilon) = \Omega$$

$Y \in \mathbb{R}^n$ is a vector of returns of all assets in the system, $X \in \mathbb{R}^{n \times k}$ is a matrix, the columns of which store the values of a macroeconomic factors for each asset, so the i-th column represents the i-th factor. Using GLS we estimate the β-coefficient, namely

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$$

It is sought to prove that $e_i^T \hat{\beta}$ is a return of an optimal portfolio, which is dependent only on the i-th factor. Let it have wights w.

$$\Rightarrow R_{portfolio} = R_w = w^T Y$$

Since the portfolio is optimal, it satisfies a minimum variance condition.

$$Var(R_w) = Var(w^T Y) = w^T Var(X\beta + \varepsilon) w = w^T Var(\varepsilon) w = w^T \Omega w$$

In addition to optimality, the portfolio is independent of other factors, therefore it has 0 exposure to columns of X other than i. Thusly we conclude that w solves the following optimization problem:

$$min_w : \frac{1}{2} w^T \Omega w$$
$$s.t : X^T w = e_i$$

Utilizing Lagrange multipliers, we get:

$$
\begin{aligned}
\mathcal{L}(w, \lambda) &= \tfrac{1}{2} w^T \Omega w + \lambda^T (X^T w - e_i) \\
\frac{\partial \mathcal{L}}{\partial w} &= w^T \Omega + \lambda^T X^T = \mathbf{0} \\
w^T \Omega &= -\lambda^T X^T \\
w^T &= -\lambda^T X^T \Omega^{-1} \\
e_i^T &= w^T X = -\lambda^T X^T \Omega^{-1} X \\
-\lambda^T &= e_i^T (X^T \Omega^{-1} X)^{-1} \\
\Rightarrow w^T &= e_i^T (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} \\
\Rightarrow R_w = w^T Y &= e_i^T (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y = e_i^T \hat{\beta}
\end{aligned}
$$

We derived that the β-coefficient of each factor represents the return of a portfolio with unit exposure to the factor it mimics.
It therefore follows that the β-coefficients of the exogenous factor from each cross-section regression form a time series of returns of the exogenous factor-mimicking portfolio. If it exhibits statistically significant growth and positive return, said otherwise, if it is a worthwhile investment, we would have confirmed our hypothesis.

4.3 Raw Data

The dataset we utilize to construct the panel data for the regressions is the Core US Equities Bundle by Sharadar via Quandl under an academic license. It includes 3 tables we use - Core US Fundamentals, Core US Insiders and Equity Prices.

- The dataset Core US Fundamentals is comprised of quarterly data on 3000 US companies across a period of 10 years, from Q1 2009 to Q1 2019. It details a variety of financial metrics, gathered from the companies' quarterly statements.
- Core US Insiders is the most essential part of the empirical analysis. It catalogues all transactions of assets to insider agents (those who are employees of the company) for each company. By taking solely the transactions of shares and summing them by quarter, we get the total amount of novel exogenous capital for the quarter, which, after dividing by the total amount of shares, provides us with a measure for the exogenous factor.
- The data table Equity Prices records the price of each ticker for the aforementioned period, thereby enabling us to easily calculate the daily return for each asset.

In addition to this, we use the industry returns of the 49 industries, defined by E. Fama and K. French [6]

4.4 Regression Dataset

In order to obtain a beta coefficient of the exogenous factor for each date, the use of panel data is necessary. Its cross section incorporates the following data for each stock:

- daily return
- shares issued to insiders during the quarter as a percentage of the basic shares; measure of the exogenous factor
- market capitalization, used to construct the SMB factor; recorded daily
- price-to-book ratio, used to construct the HML factor; on a daily basis
- return of the previous quarter, used to estimate the momentum effect
- quarterly debt/equity ratio, used to construct the quality factor
- industry beta - β-coefficient of the regression that models the time-series of daily returns of the companies' stock as a linear function of the total growth of its respective industry (49 industries in total); Bayesian shrinkage [7] is applied, for the length of the time-series is short; recalculated for each quarter

The data that is on a quarterly basis remains the same for all dates in the quarter.
Apart from the effect of the exogenous factor, we are further interested in its longevity. In order to cognize it, the exogenous factor from multiple quarters before the date of the regression is used in X_{exo}.

4.5 Regression Models

With the general structure of the cross section clarified, we are ready to proceed with the factor model itself.

$$Y = X\beta + \alpha_{market}$$

$$X = \begin{bmatrix} X_{exo} & X_{SMB} & X_{HML} & X_{MOM} & X_{quality} & X_{ind} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_{exo} \\ \beta_{SMB} \\ \beta_{HML} \\ \beta_{MOM} \\ \beta_{quality} \\ \beta_{ind} \end{bmatrix}$$

The design matrix X consists of the horizontally stacked column vectors for each factor, in which the values for each company are held. The regressant variable Y is a vector, containing all the daily yields for each company in the cross section. For each factor corresponds a β-coefficient, which is a best linear unbiased estimator for the regression. There is also a common alpha that together with the industry beta form the market risk premium. The reason the industry premium is incorporated is that it offers higher accuracy, as different industries have vastly dissimilar growth in this period.

4.6 Data Pre-processing

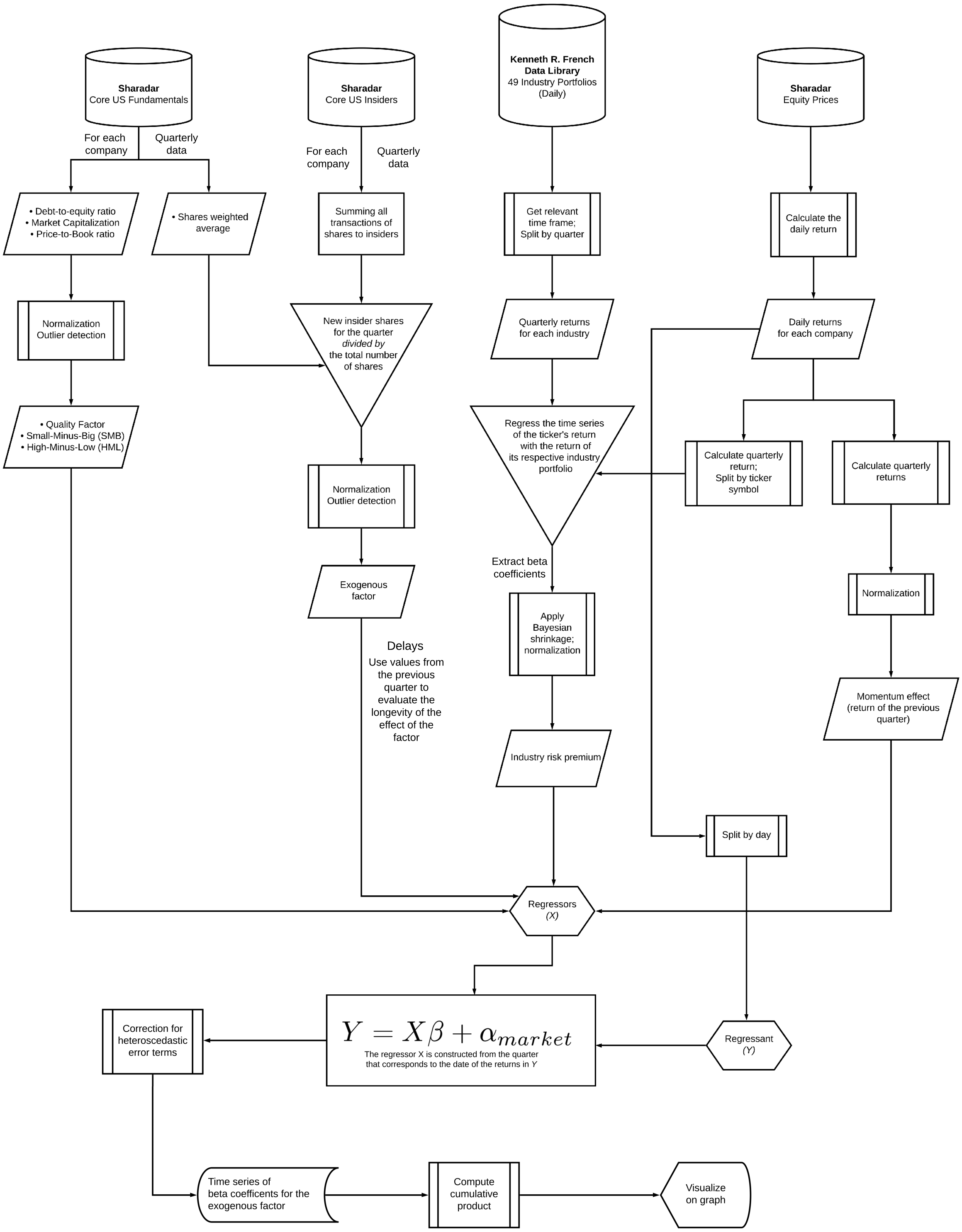
All the data goes through a normalization routine with mean $\mu = 0$ and standart deviation $\sigma = 1$. This ensures that the β-coefficients are unambiguous and noise-free. When appropriate, outliers are removed as the uppermost and lowermost 2.5%, equating them to the new greatest and smallest value respectively.

4.7 Data Post-processing

The error terms in the regression are typically not uncorrelated, therefore corrections for heteroscedasticity are applied using a robust covariance matrix (sandwich) estimator [8] [9]. After the corrected betas are computed, outliers are removed and the cumulative product is calculated, so that the time-series is primed for visualization.

4.8 Data Processing Pipeline

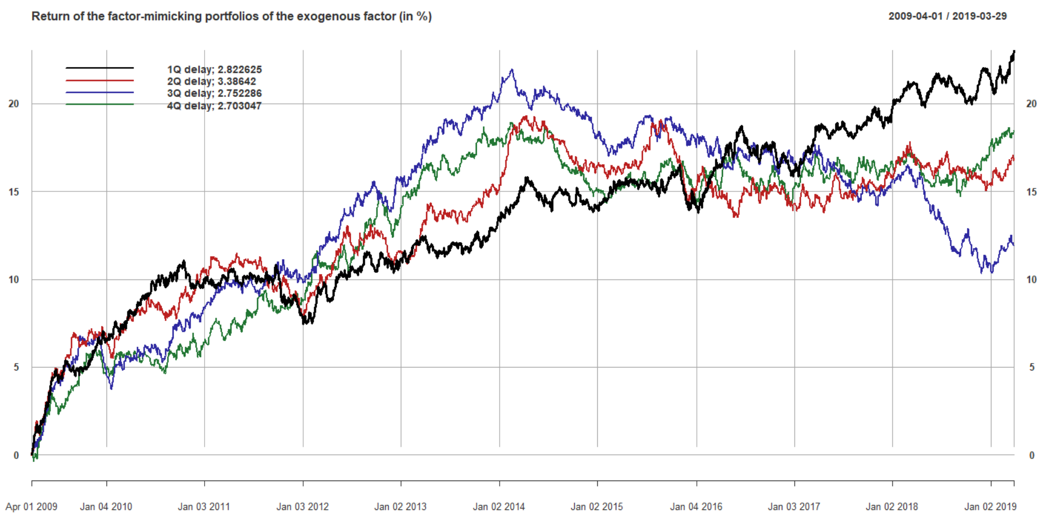
On the info graphic below is expounded the end-to-end process how the results are obtained.



Sharadar
Core US Fundamentals
For each company
Quarterly data
• Debt-to-equity ratio
• Market Capitalization
• Price-to-Book ratio
• Shares weighted average
Normalization
Outlier detection
• Quality Factor
• Small-Minus-Big (SMB)
• High-Minus-Low (HML)
Sharadar
Core US Insiders
For each company
Quarterly data
Summing all transactions of shares to insiders
New insider shares for the quarter divided by the total number of shares
Normalization
Outlier detection
Exogenous factor
Delays
Use values from the previous quarter to evaluate the longevity of the effect of the factor
Kenneth R. French
Data Library
49 Industry Portfolios
(Daily)
Get relevant time frame;
Split by quarter
Quarterly returns for each industry
Regress the time series of the ticker's return with the return of its respective industry portfolio
Extract beta coefficients
Apply Bayesian shrinkage; normalization
Industry risk premium
Sharadar
Equity Prices
Calculate the daily return
Daily returns for each company
Calculate quarterly return;
Split by ticker symbol
Calculate quarterly returns
Normalization
Momentum effect
(return of the previous quarter)
Split by day
Regressors
(X)
Regressant
(Y)
Y = Xβ + α_market
The regressor X is constructed from the quarter that corresponds to the date of the returns in Y
Correction for heteroscedastic error terms
Time series of beta coefficents for the exogenous factor
Compute cumulative product
Visualize on graph

4.9 Results

Owing to what we proved in the theoretical background section, the effect of exogenous factor incarnates in its β-coefficient, which signifies the return of its factor-mimicking portfolio. Given that we have a cross-section regression model for each working day, we can extract a time series of β-coefficients, i.e. returns, which we visualize in the graph below. Although some exhibiting



Sharpe Ratios are shown in the legend

volatility, the trend of all cumulative returns is one of growth. Even with exogenous capital issued one year ago, the portfolio still delivers steadily increasing returns. This shows that the effect of the exogenous factors is lasting, rather than temporary and short-term.

We further examine the Sharpe ratio for each time series - in essence they measure the effectiveness of investing in the portfolio in question, with values from 1 to 2 ranging from acceptable to good and above 2 - very good. The values we get are displayed next to the legend. All of them are around 2.7 or higher, for that reason we can claim that the exogenous factor-mimicking portfolio is a worthwhile investment and, therefore, the exogenous factor is a statistically significant source of asset return.

5 Conclusion

Through the mathematical model we postulate that the exogenous factor is a source of asset return. We analyze the statement of this hypothesis by empirical means and arrive at the conclusion that not only does it increase yield, but also it does so by a statistically significant margin. These findings do in fact possess a practical implication - they could be put into use by quantitative funds, thereby improving portfolio management and generating additional return, resulting in better profits for the funds and higher yields for their customers. Further development of the project could be an implementation of other capital assets, for example options, as well as an examination of a longer delay interval, in order to outline the time range of the effect of the exogenous factor.

6 Appendix

Here we detail the proofs of theorems used in the paper.

Quadratic optimization proof

We have a general quadratic optimization problem:

$$min : f(Q) = Q^T \Sigma Q + c^T Q$$

Optimum is reached at:

$$\Sigma Q^* + c = 0 \iff Q^* = -\Sigma^{-1} c$$

Proof:

Let Q^* satisfy the condition $\Sigma Q^* + c = 0$

$$\begin{aligned} f(Q_0) =& f(Q^* + (Q_0 - Q^*)) = \frac{1}{2}(Q^* + (Q_0 - Q^*))^T \Sigma (Q^* + (Q_0 - Q^*)) + c^T (Q^* + (Q_0 - Q^*)) = \\ =& \frac{1}{2} Q^{*T} \Sigma Q^* + (Q_0 - Q^*)^T \Sigma Q^* + \frac{1}{2}(Q_0 - Q^*)^T \Sigma (Q_0 - Q^*) + c^T Q^* + c^T (Q_0 - Q^*) = \\ =& \frac{1}{2} Q^{*T} \Sigma Q^* + c^T Q^* + (Q_0 - Q^*)^T (\Sigma Q^* + c) + \frac{1}{2}(Q_0 - Q^*)^T \Sigma (Q_0 - Q^*) = \\ =& f(Q^*) + \frac{1}{2}(Q_0 - Q^*)^T \Sigma (Q_0 - Q^*) \geq f(Q^*) \end{aligned}$$

Now let optimum be achieved at Q^* and $\Sigma Q^* + c = d$ and let us suppose that $d \neq 0$

$$\begin{aligned} \Rightarrow f(Q^* + ad) &= \frac{1}{2}(Q^* + ad)^T \Sigma (Q^* + ad) + c^T (Q^* + ad) = \\ &= \frac{1}{2} Q^{*T} \Sigma Q^* + ad^T \Sigma Q^* + \frac{1}{2} a^2 d^T \Sigma d + c^T Q^* + c^T ad = \\ &= f(Q^*) + ad^T (\Sigma Q^* + c) + \frac{1}{2} a^2 d^T \Sigma d = \\ &= f(Q^*) + ad^T d + \frac{1}{2} a^2 d^T \Sigma d \end{aligned}$$

For $a < 0$, but sufficiently close to 0

$$\Rightarrow ad^T d + \frac{1}{2} a^2 d^T \Sigma d < 0 \Rightarrow f(Q^* + ad) < f(Q^*)$$

which contradicts the assumption

Q.E.D.

Estimation of $\hat{\beta}$ from a Generalized Least Squares Regression (GLS)

We have the following linear regression model:

$$Y = X\beta + \varepsilon$$
$$E[\varepsilon|X] = 0$$
$$Var(\varepsilon|X) = \Omega$$

$\hat{\beta}$ is a best estimator if and only if it solves the following optimization problem:

$$\begin{aligned}
&min_\beta : \sqrt{\varepsilon^T \Omega^{-1} \varepsilon} \\
\iff &\min_\beta : (Y - X\beta)^T \Omega^{-1} (Y - X\beta) \\
\iff &\min : S(\beta) = Y^T \Omega^{-1} Y - 2Y^T \Omega^{-1} X\beta + \beta^T X^T \Omega^{-1} X\beta \\
\Longrightarrow &\frac{\partial S}{\partial \beta} = -2X^T \Omega^{-1} Y + 2X^T \Omega^{-1} X\beta = 0 \\
\Longrightarrow &\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y
\end{aligned}$$

Gauss-Markov theorem

The Gauss-Markov theorem states that the above estimation $\hat{\beta}$ is the Best Linear Unbiased Estimator (BLUE).
Let us begin by proving that $\hat{\beta}$ is unbiased, in other terms:

$$\begin{aligned}
&E[\hat{\beta}|X] = \beta \\
&E[\hat{\beta}|X] = E[(X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y|X] = \\
&E[(X^T \Omega^{-1} X)^{-1} (X^T \Omega^{-1} X)\beta|X] = \beta
\end{aligned}$$

All that remains unproven is the statement that $\hat{\beta}$ is the best of all linear unbiased estimators, meaning that:

$$Var(\hat{\beta}|X) \leq Var(\beta^*|X)$$

$$\text{for some other unbiased estimator } \beta^*$$

$$\begin{aligned}
Var(\hat{\beta}|X) &= Var((X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y|X) = \\
&= (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Var(Y|X) \Omega^{-1} X (X^T \Omega^{-1} X)^{-1} = \\
&= (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} \Omega \Omega^{-1} X (X^T \Omega^{-1} X)^{-1} = \\
&= (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} X (X^T \Omega^{-1} X)^{-1} = \\
&= (X^T \Omega^{-1} X)^{-1}
\end{aligned}$$

Taking into account that β^* is an unbiased estimator, it can be formulated as

$$\beta^* = QY, \quad Q = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} + A, A \in \mathbb{R}^{k \times n}$$

β^* is unbiased, therefore:

$$E[\beta^*|X] = \beta$$

$$\begin{aligned}
E[\beta^*|X] &= E[((X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} + A)(X\beta + \varepsilon)|X] = \\
&= E[\hat{\beta} + AY|X] = E[\hat{\beta}|X] + E[AX\beta + A\varepsilon|X] = \\
&= \beta + E[AX\beta|X] \\
&\Rightarrow AX = 0
\end{aligned}$$

We shall now examine $Var(\beta^*|X)$

$$
\begin{aligned}
Var(\beta^*|X) &= Var(QY|X) = Q^T \Omega Q = \\
&= ((X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1} + A)\Omega(\Omega^{-1}X(X^T\Omega^{-1}X)^{-1} + A^T) = \\
&= Var(\hat{\beta}|X) + (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}\Omega A + \\
&+ A\Omega\Omega^{-1}X(X^T\Omega^{-1}X)^{-1} + A\Omega A^T = \\
&= Var(\hat{\beta}|X) + A\Omega A^T \geq Var(\hat{\beta}|X)
\end{aligned}
$$

Q.E.D.

References

[1] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, vol. 19, no. 3, pp. 425–442, 1964.

[2] J. W. Pratt, "Risk aversion in the small and in the large," *Econometrica*, vol. 32, pp. 122–136, 1964.

[3] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, vol. 33, no. 1, pp. 3–56, 1993.

[4] M. M. Carhart, "On persistence in mutual fund performance," *The Journal of finance*, vol. 52, no. 1, pp. 57–82, 1997.

[5] L. C. Bhandari, "Debt/equity ratio and expected common stock returns: Empirical evidence," *The journal of finance*, 1988.

[6] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research`.

[7] O. A. Vasicek, "A note on using cross-sectional information in bayesian estimation of security betas," *The Journal of Finance*, 1973.

[8] H. White *et al.*, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, vol. 48, no. 4, pp. 817–838, 1980.

[9] A. Zeileis, "Object-oriented computation of sandwich estimators," 2006.

[10] Data from Sharadar via Quandl.

[11] J. H. Cochrane, *Asset pricing: Revised edition.* Princeton university press, 2009.

[12] B. E. Hansen, *Econometrics.* University of Wisconsin, 2019.