Economics Bulletin

Volume 41, Issue 3

The three-factor model without a linear return generating process

Claude Bergeron
School of Business Administration, Teluq University

Abstract

From a theoretical point of view, the Fama and French three-factor model requires the following implicit assumptions: (i) the excess return of an asset is correlated with market, size, and book-to-market factors, and (ii) the return generating process is linear. In this note, we demonstrate that the linearity assumption of the return generating process can be relaxed. This suggests that assumption (i) alone is sufficient for the three-factor model.

Citation: Claude Bergeron, (2021) "The three-factor model without a linear return generating process", *Economics Bulletin* Vol. 41 No. 3 pp. 1763-1772
Contact: Claude Bergeron - claude.bergeron@teluq.ca.
Submitted: March 03, 2021. **Published:** September 17, 2021.

1. Introduction

According to Cox and Britten (2019), the Fama and French three-factor model explains stock returns better than the classic capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). The three-factor model of Fama and French (1993) was motivated by previous observations that demonstrated an empirical correlation between stock returns, size, and book-to-market equity. More specifically, the model assumes that an asset's excess return can be explained by a time-series linear regression that integrates the following factors: (1) the excess return on a broad market portfolio, (2) the spread in returns between *small* and *big* firms, and (3) the spread in returns between *value* and *growth* stocks.

Following Fama and French's (1993) initial observations, many other studies have been proposed. For example, Fama and French (1995) showed that there are *market*, *size*, and *book-to-market* factors in earnings like those in returns. Carhart (1997) proposed a four-factor model that introduced a momentum factor. Barry et al. (2002) tested for the size and value effects and found strong evidence of a value premium only. Basiewicz and Auret (2009) revisited the three-factor model, taking liquidity into account. Fama and French (2015 and 2017) added the factors of profitability and investment to their original three-factor model to create a five-factor version (see also Foye, 2018).

However, all the aforementioned studies suppose that the return generating process is linear. In this note, we demonstrate that the main theoretical prediction of the three-factor model can be obtained without assuming the linearity of the return generating process.

As noted by Kwon (1985), there have been several attempts to relax some restrictive assumptions that underlie the CAPM. For instance, Fama (1971) and Ross (1978) showed that the normality assumption is not necessary for the model. Kwon (1985) derived a model similar to the CAPM without normality or quadratic preference. Nielsen (1990) presented a general equilibrium version of the CAPM without riskless assets (see also Berk, 1997, or Shalit and Yitzhaki, 2009). In this vein, our goal is to relax a restrictive assumption that underlies the three-factor model. Our motivation comes from the following observations: (1) the importance of the three-factor model in the field of asset pricing, (2) the high level of influence the three-factor model has on investors and portfolio managers, and (3) the natural tendency in asset pricing theory to relax the number of restrictive assumptions.

Following Connor and Linton (2007), Connor et al. (2012) also adopt a nonlinear version of the Fama-French three-factor model. Nevertheless, in their setup, they assume that: (1) assets returns are generated by a specific weighted additive nonparametric regression model (see equation 1, on page 716); (2) the semiparametric model use different *characteristic-beta functions* (g), where each function is time-invariant; (3) the security characteristic variables (X) are also time-invariant; and (4) the standard error term (ε) has a mean of zero. Our model makes none of these assumptions. In this regard, compared to these papers, our manuscript contributes to the asset pricing literature by initiating a general approach without any specific restrictive assumption related to the return generating process.

The remainder of the paper will proceed as follows. Section 2 presents the three-factor model (in theory), Section 3 proposes an extension model without linearity, and Section 4 provides the paper's conclusion.

2. The three-factor model (in theory)

Let $\tilde{R}_{i,t+1}$ be the random return of asset i, at time $t+1$, $R_{F,t+1}$ be the return of the riskless asset F, at time $t+1$, and $\tilde{r}_{i,t+1}$ be the excess return of asset i, at time $t+1$ ($\tilde{r}_{i,t+1} \equiv \tilde{R}_{i,t+1} - R_{F,t+1}$).[1] In theory, given the available information at time t, the three-factor model supposes that the return generating process of an asset can be expressed in this manner (for $i = 1, 2, \dots, N$):

$$\tilde{r}_{i,t+1} = \beta_{Mit}\tilde{r}_{M,t+1} + \beta_{Sit}\tilde{S}_{t+1} + \beta_{Hit}\tilde{H}_{t+1} + \tilde{e}_{i,t+1}, \tag{1}$$

with:

$$0 = E_t\left[\tilde{e}_{i,t+1}\right] = Cov_t\left[\tilde{e}_{i,t+1}, \tilde{r}_{M,t+1}\right] = Cov_t\left[\tilde{e}_{i,t+1}, \tilde{S}_{t+1}\right] = Cov_t\left[\tilde{e}_{i,t+1}, \tilde{H}_{t+1}\right],$$

where $\tilde{r}_{M,t+1}$ is the excess return of the market portfolio M, at time $t+1$ ($\tilde{r}_{M,t+1} \equiv \tilde{R}_{M,t+1} - R_{F,t+1}$), $\tilde{S}_{t+1}$ is the difference, at time $t+1$, between the return on a portfolio of small stocks and the return on a portfolio of large stocks (*small* minus *big*), and $\tilde{H}_{t+1}$ is the difference, at time $t+1$, between the return on a portfolio of high book-to-market stocks and the return on a portfolio of low book-to-market stocks (*high* minus *low*). Considering the information at time t, parameters β_{Mit}, β_{Sit}, and β_{Hit} represent, respectively, for the asset i, the return sensitivity to factors $\tilde{r}_{M,t+1}$, $\tilde{S}_{t+1}$, and $\tilde{H}_{t+1}$, while $\tilde{e}_{i,t+1}$ corresponds to the usual random term, for asset i, at time $t+1$.

The three-factor model's approach is in the spirit of Merton's (1973) intertemporal capital asset pricing model (ICAPM) and Ross's (1976) arbitrage pricing theory (APT).[2] In Eq. (1), the three factors $\tilde{r}_{M,t+1}$, $\tilde{S}_{t+1}$, and $\tilde{H}_{t+1}$, affect the returns on more than one asset. The parameter β_{Mit}, β_{Sit}, or β_{Hit}, is unique to each asset and represents an attribute of the asset that is considered a factor loading. Taking the expected value on both sides of Eq. (1) gives the main prediction of the three-factor model, that is:

$$E_t\left[\tilde{r}_{i,t+1}\right] = \beta_{Mit}E_t\left[\tilde{r}_{M,t+1}\right] + \beta_{Sit}E_t\left[\tilde{S}_{t+1}\right] + \beta_{Hit}E_t\left[\tilde{H}_{t+1}\right], \tag{2a}$$

$$E_t\left[\tilde{r}_{i,t+1}\right] = \lambda_{Mt}\beta_{Mit} + \lambda_{St}\beta_{Sit} + \lambda_{Ht}\beta_{Hit}, \tag{2b}$$

[1] In this manuscript, the tilde (~) indicates a random variable. Operators E_t, V_t, and Cov_t refer respectively to mathematical expectations, variance and covariance, where index t implies that we consider the available information at time t.

[2] See Fama and French (2004, page 38).

where $\lambda_{Mt} \equiv E_t[\tilde{r}_{M,t+1}]$, $\lambda_{St} \equiv E_t[\tilde{S}_{t+1}]$, and $\lambda_{Ht} \equiv E_t[\tilde{H}_{t+1}]$. Eq. (2a) gives the expected returns of an asset when returns are generated by a linear three-index model. In Eq. (2b), parameters λ_{Mt}, λ_{St}, and λ_{Ht} indicate the price of risk for the corresponding factor sensitivities β_{Mit}, β_{Sit}, and β_{Hit}. Therefore, in theory, the three-factor model requires the following implicit assumptions: (i) the excess return of an asset is correlated with *market-return*, *size*, and *book-to-market* factors, as expressed by variables $\tilde{r}_{M,t+1}$, $\tilde{S}_{t+1}$, and $\tilde{H}_{t+1}$, and (ii) the return generating process defined by Eq. (1) is linear. In the next section, we will demonstrate that the linearity assumption of the return generating process can be relaxed. This suggests that assumption (i) alone is sufficient to obtain a risk-return relationship equivalent to Eq. (2a) or (2b).

3. The extension model without linearity restriction

In this section, we extended the three-factor model without assuming that the return generating process is linear. We first derived a direct relationship with one general risk measure, and then with three risk measures.

3.1. One general risk measure

Without any specific assumption related to linearity, arbitrage or equilibrium, mathematically, from the covariance definition, we can write:

$$Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}] = E_t[\tilde{Y}_{t+1}\tilde{r}_{i,t+1}] - E_t[\tilde{Y}_{t+1}]E_t[\tilde{r}_{i,t+1}], \tag{3}$$

with $\tilde{Y}_{t+1} \equiv (1 + \tilde{X}_{t+1})^{-1}$, where $\tilde{X}_{t+1}$ corresponds to a general random variable positively correlated with asset returns. Rearranging Eq. (3), we get:

$$E_t[\tilde{Y}_{t+1}\tilde{r}_{i,t+1}] = Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}] + E_t[\tilde{Y}_{t+1}]E_t[\tilde{r}_{i,t+1}], \tag{4a}$$

or, to simplify the notation:

$$E_t[\tilde{Y}_{t+1}\tilde{r}_{i,t+1}] = \emptyset_{it}, \tag{4b}$$

with $\emptyset_{it} \equiv Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}] + E_t[\tilde{Y}_{t+1}]E_t[\tilde{r}_{i,t+1}]$, where $\emptyset_{it}$ is not necessarily equal to zero. Dividing on each side by $\emptyset_{it}$ allows us to show a particular form of the familiar Euler equation in which central random variables are driven by the asset's return and a general variable, that is:

$$E_t[\tilde{Y}_{t+1}\tilde{r}_{i,t+1}/\emptyset_{it},] = E_t[\tilde{Y}_{t+1}\tilde{Z}_{i,t+1}] = 1, \tag{5}$$

where $\tilde{Z}_{i,t+1} \equiv \frac{\tilde{r}_{i,t+1}}{Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}] + E_t[\tilde{Y}_{t+1}]E_t[\tilde{r}_{i,t+1}]}$. In the same manner, for the portfolio p, whose returns are perfectly correlated with $\tilde{X}_{t+1}$, we have:

$$E_t\left[\tilde{Y}_{t+1}\tilde{Z}_{p,t+1}\right] = 1, \tag{6}$$

where $\tilde{Z}_{p,t+1} \equiv \frac{\tilde{r}_{p,t+1}}{Cov_t[\tilde{Y}_{t+1},\tilde{r}_{p,t+1}]+E_t[\tilde{Y}_{t+1}]E_t[\tilde{r}_{p,t+1}]}$. Thus, Eq. (5) minus Eq. (6) gives:

$$E_t\left[\tilde{Y}_{t+1}(\tilde{Z}_{i,t+1} - \tilde{Z}_{p,t+1})\right] = 0, \tag{7}$$

and the mathematical definition of covariance implies that:

$$Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{i,t+1} - \tilde{Z}_{p,t+1}\right] = -E_t[\tilde{Y}_{t+1}]E_t\left[\tilde{Z}_{i,t+1} - \tilde{Z}_{p,t+1}\right], \tag{8}$$

or, after simple manipulations:

$$Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{i,t+1} - \tilde{Z}_{p,t+1}\right] = E_t[\tilde{Y}_{t+1}]E_t\left[\tilde{Z}_{p,t+1}\right] - E_t[\tilde{Y}_{t+1}]E_t\left[\tilde{Z}_{i,t+1}\right]. \tag{9}$$

Isolating the expected value of variable $\tilde{Z}_{i,t+1}$ indicates that:

$$E_t\left[\tilde{Z}_{i,t+1}\right] = E_t\left[\tilde{Z}_{p,t+1}\right] - Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{i,t+1} - \tilde{Z}_{p,t+1}\right]/E_t[\tilde{Y}_{t+1}], \tag{10}$$

or, using the basic properties of mathematical covariance, that:

$$E_t\left[\tilde{Z}_{i,t+1}\right] = E_t\left[\tilde{Z}_{p,t+1}\right] + E_t^{-1}[\tilde{Y}_{t+1}]Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}\right] - E_t^{-1}[\tilde{Y}_{t+1}]Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{i,t+1}\right], \tag{11}$$

with $E_t^{-1}[\tilde{Y}_{t+1}] \equiv 1/E_t[\tilde{Y}_{t+1}]$. Multiplying on each side by the dominator of variable $\tilde{Z}_{i,t+1}$, we can write:

$$E_t\left[\tilde{r}_{i,t+1}\right] = E_t\left[\tilde{Z}_{p,t+1}\right](Cov_t\left[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}\right] + E_t[\tilde{Y}_{t+1}]E_t\left[\tilde{r}_{i,t+1}\right]) + E_t^{-1}[\tilde{Y}_{t+1}]Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}\right](Cov_t\left[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}\right] + E_t[\tilde{Y}_{t+1}]E_t\left[\tilde{r}_{i,t+1}\right]) - E_t^{-1}[\tilde{Y}_{t+1}]Cov_t\left[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}\right]. \tag{12}$$

Developing, we can also write:

$$E_t\left[\tilde{r}_{i,t+1}\right] = E_t\left[\tilde{Z}_{p,t+1}\right]Cov_t\left[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}\right] + E_t\left[\tilde{Z}_{p,t+1}\right]E_t[\tilde{Y}_{t+1}]E_t\left[\tilde{r}_{i,t+1}\right] + E_t^{-1}[\tilde{Y}_{t+1}]Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}\right]Cov_t\left[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}\right] + Cov_t\left[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}\right]E_t\left[\tilde{r}_{i,t+1}\right] - E_t^{-1}[\tilde{Y}_{t+1}]Cov_t\left[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}\right]. \tag{13}$$

Regrouping the elements, Eq. (13) becomes:

$$E_t[\tilde{r}_{i,t+1}] = E_t[\tilde{r}_{i,t+1}](E_t[\tilde{Z}_{p,t+1}]E_t[\tilde{Y}_{t+1}] + Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}]) + (E_t[\tilde{Z}_{p,t+1}]+E_t^{-1}[\tilde{Y}_{t+1}]Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}]-E_t^{-1}[\tilde{Y}_{t+1}])Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]. \tag{14}$$

Isolating the expected excess return of the asset, we have:

$$E_t[\tilde{r}_{i,t+1}] = \frac{E_t^{-1}[\tilde{Y}_{t+1}]\left(Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}] - 1\right) + E_t[\tilde{Z}_{p,t+1}]}{1 - E_t[\tilde{Z}_{p,t+1}]E_t[\tilde{Y}_{t+1}] - Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}]} \; Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]\,. \tag{15}$$

Multiplying by $E_t[\tilde{Y}_{t+1}]$ on each side of Eq. (15) allows us to express:

$$E_t[\tilde{r}_{i,t+1}] = Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]x \; \frac{E_t^{-1}[\tilde{Y}_{t+1}]\left(Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}] - 1\right) + E_t[\tilde{Z}_{p,t+1}]}{1 - E_t[\tilde{Z}_{p,t+1}]E_t[\tilde{Y}_{t+1}] - Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}]} \; \frac{E_t[\tilde{Y}_{t+1}]}{E_t[\tilde{Y}_{t+1}]}. \tag{16}$$

Taking the product of the two numerators indicates that:

$$E_t[\tilde{r}_{i,t+1}] = Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]x \; \frac{Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}] + E_t[\tilde{Z}_{p,t+1}]E_t[\tilde{Y}_{t+1}] - 1}{1 - E_t[\tilde{Z}_{p,t+1}]E_t[\tilde{Y}_{t+1}] - Cov_t[\tilde{Y}_{t+1}, \tilde{Z}_{p,t+1}]} \; \frac{1}{E_t[\tilde{Y}_{t+1}]}\,. \tag{17}$$

Thereby, after simplification, we have:

$$E_t[\tilde{r}_{i,t+1}] = -E_t^{-1}[\tilde{Y}_{t+1}]Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]. \tag{18}$$

Our Eq. (18) is analogous, in its form, to equation (4.39) in Campbell (2018, page 94) or equation (1.12) in Cochrane (2005, page 17). Multiplying by $V_t[\tilde{Y}_{t+1}]$ on each side gives:

$$E_t[\tilde{r}_{i,t+1}] = -\frac{V_t[\tilde{Y}_{t+1}]}{E_t[\tilde{Y}_{t+1}]} \, \frac{Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]}{V_t[\tilde{Y}_{t+1}]}\,, \tag{19a}$$

and we can rewrite this equation as:

$$E_t[\tilde{r}_{i,t+1}] = \lambda_t \beta_{it}, \tag{19b}$$

or, alternatively, as:

$$E_t[\tilde{R}_{i,t+1}] = R_{F,t+1} + \lambda_t \beta_{it}, \tag{19c}$$

where $\lambda_t \equiv V_t[\tilde{Y}_{t+1}]/E_t[\tilde{Y}_{t+1}]$, and $\beta_{it} \equiv -Cov_t[\tilde{Y}_{t+1}, \tilde{r}_{i,t+1}]/V_t[\tilde{Y}_{t+1}]$.

Eq. (19) represents our first (theoretical) result. This equation reveals that the expected return of an asset is equal to the risk-free rate, plus a risk premium directly proportional to a general beta, obtained from the covariance between the excess return of the asset and a function of a general random variable ($\tilde{X}_{t+1}$) positively correlated with asset returns.

Here, Eq. (19) is similar to equation (1.15) in Cochrane (2005, page 19) and equation (4.40) in Campbell (2018, page 94). Like these two familiar equations, Eq. (19) suggests that all assets have an expected return equal to the return of the riskless asset, plus a risk premium. In Eq. (19), the risk premium is equivalent to $\lambda_t\beta_{it}$, and is defined by our variable $\tilde{Y}_{t+1}$. The coefficient λ_t is the same for all assets, while the β_{it} varies from asset to asset, and we interpret λ_t as the *price of risk* and the *general beta* (β_{it}) as the *quantity of risk*.

Starting from the covariance definition, the linearity of the risk-return relationship predicted by Eq. (19) represents a mathematical result obtained from simple algebraic manipulations, only. Our Eq. (19) is not based on a pure *Arrow-Debreu security* and on the corresponding *stochastic discount factor*, as defined by Campbell (2018, Chapter 4). In this sense, our linearity prediction does not come from any equilibrium or arbitrage conditions, just as the three-factor model (recall that $\emptyset_{it}$, in Eq. (4b), does not have to be equal to 0).

Besides, in our theoretical framework, we implicitly assume that *investors prefer more to less, are risk averse, and demand a premium in the form of higher expected returns for the risks they assume*. Thus, in accordance with this fundamental postulate, our first result, expressed by Eq. (19), must imply that the *general beta* represents a rightful measure of risk (as we proposed above). This theoretical measure captures the sensitivity of the asset's returns to any general random variable or macroeconomic factor that influences the variability of the asset's returns. In that respect, it captures the instability of the asset that comes from the covariance between its returns and any general random variable (as initially proposed in our Eq. 3). Moreover, this *general beta* is similar to the risk measure predicted by the APT of Ross (1976) in its single-factor formulation, where the only factor is (a priori) unknown. Recall that in the APT, the factors that influence returns are not determined by a specific economic model. The potential random variable defining the general beta could be any factor typically used in a standard multifactor model. For example, we could define the general beta using the market portfolio return, aggregate consumption growth, inflation rate, industrial production growth, or other analogous variables. In the next subsection, we simply assume that the general beta can be obtained using a market factor (F_M), size factor (F_S), or book-to-market factor (F_H).

To compute this general beta, we can adopt the following procedure: (1) choose the general variable (X); (2) transform the variable in this manner: $Y = 1/(1+X)$; and (3) regress the asset's return on the variable Y, with a simple linear regression approach. If, for example, we establish that the transformed variable Y corresponds to a market factor (F_M), size factor (F_S), or book-to-market factor (F_H), then any market data employed in standard empirical studies related to the Fama-French three-factor model could be used.

3.2. Three risk measures

From a general point of view, if we assume that y is correlated with x, z, and w, then, from the covariance definition, we can see that the expected value of y can be obtained (or calculated) in three different ways, using three different covariance values. In the same manner, if we assume that asset returns are correlated with *market*, *size*, and *book-to-market* variables (or factors), then, from Eq. (18), we can write:

$$E_t[\tilde{r}_{i,t+1}] = -E_t^{-1}[\tilde{F}_{M,t+1}]Cov_t[\tilde{F}_{M,t+1}, \tilde{r}_{i,t+1}], \tag{20a}$$

$$E_t[\tilde{r}_{i,t+1}] = -E_t^{-1}[\tilde{F}_{S,t+1}]Cov_t[\tilde{F}_{S,t+1}, \tilde{r}_{i,t+1}], \tag{20b}$$

$$E_t[\tilde{r}_{i,t+1}] = -E_t^{-1}[\tilde{F}_{H,t+1}]Cov_t[\tilde{F}_{H,t+1}, \tilde{r}_{i,t+1}], \tag{20c}$$

where $\tilde{F}_{M,t+1} \equiv 1/(1+\tilde{r}_{M,t+1})$ represents the *market* factor at time $t+1$, $\tilde{F}_{S,t+1} \equiv 1/(1+\tilde{S}_{t+1})$ represents the *size* factor at time $t+1$, and $\tilde{F}_{H,t+1} \equiv 1/(1+\widetilde{H}_{t+1})$ represents the *book-to-market* factor (*high* minus *low*) at time $t+1$.

Multiplying by $V_t[\tilde{F}_{M,t+1}]$, $V_t[\tilde{F}_{S,t+1}]$, and $V_t[\tilde{F}_{H,t+1}]$ on each side of Eq. (20a), (20b), and (20c), respectively, gives:

$$E_t[\tilde{r}_{i,t+1}] = -\frac{V_t[\tilde{F}_{M,t+1}]}{E_t[\tilde{F}_{M,t+1}]}\frac{Cov_t[\tilde{F}_{M,t+1}, \tilde{r}_{i,t+1}]}{V_t[\tilde{F}_{M,t+1}]}, \tag{21a}$$

with:

$$E_t[\tilde{r}_{i,t+1}] = -\frac{V_t[\tilde{F}_{S,t+1}]}{E_t[\tilde{F}_{S,t+1}]}\frac{Cov_t[\tilde{F}_{S,t+1}, \tilde{r}_{i,t+1}]}{V_t[\tilde{F}_{S,t+1}]}, \tag{21b}$$

and:

$$E_t[\tilde{r}_{i,t+1}] = -\frac{V_t[\tilde{F}_{H,t+1}]}{E_t[\tilde{F}_{H,t+1}]}\frac{Cov_t[\tilde{F}_{H,t+1}, \tilde{r}_{i,t+1}]}{V_t[\tilde{F}_{H,t+1}]}. \tag{21c}$$

Using a compact notation, we can write:

$$E_t[\tilde{r}_{i,t+1}] = \Lambda_{Mt}\mathrm{B}_{Mit}, \tag{22a}$$

$$E_t[\tilde{r}_{i,t+1}] = \Lambda_{St}\mathrm{B}_{Sit}, \tag{22b}$$

$$E_t[\tilde{r}_{i,t+1}] = \Lambda_{Ht}\mathrm{B}_{Hit}, \tag{22c}$$

$$\Lambda_{Mt} \equiv V_t[\tilde{F}_{M,t+1}]/E_t[\tilde{F}_{M,t+1}], \text{ and } \mathrm{B}_{Mit} \equiv -Cov_t[\tilde{F}_{M,t+1}, \tilde{r}_{i,t+1}]/V_t[\tilde{F}_{M,t+1}];$$

$$\Lambda_{st} \equiv V_t[\tilde{F}_{S,t+1}]/E_t[\tilde{F}_{S,t+1}], \text{ and } \mathrm{B}_{Sit} \equiv -Cov_t[\tilde{F}_{S,t+1}, \tilde{r}_{i,t+1}]/V_t[\tilde{F}_{S,t+1}];$$

$$\Lambda_{Ht} \equiv V_t[\tilde{F}_{H,t+1}]/E_t[\tilde{F}_{H,t+1}], \text{ and } \mathrm{B}_{Hit} \equiv -Cov_t[\tilde{F}_{H,t+1}, \tilde{r}_{i,t+1}]/V_t[\tilde{F}_{H,t+1}].$$

Eq. (22a) plus Eq. (22b) and Eq. (22c) shows:

$$3E_t[\tilde{r}_{i,t+1}] = \Lambda_{Mt}\mathrm{B}_{Mit} + \Lambda_{St}\mathrm{B}_{Sit} + \Lambda_{Ht}\mathrm{B}_{Hit}. \quad (23)$$

Dividing by 3 on each side of Eq. (23) gives:

$$E_t[\tilde{r}_{i,t+1}] = \lambda^*_{Mt}\mathrm{B}_{Mit} + \lambda^*_{St}\mathrm{B}_{Sit} + \lambda^*_{Ht}\mathrm{B}_{Hit}. \quad (24)$$

where $\lambda^*_{Mt} \equiv \Lambda_{Mt}/3$, $\lambda^*_{St} \equiv \Lambda_{St}/3$, and $\lambda^*_{Ht} \equiv \Lambda_{Ht}/3$.

Eq. (24) indicates that the expected returns of an asset can be described by an *N*-dimensional hyper plane (with $N = 3$). This equation is very close to the main prediction of the standard three-factor model as expressed by Eq. (2b).

Examining the portfolio m, that has a B_{Mmt} of *one*, with B_{Smt} and β_{Hmt} equal to *zero*, implies that $E_t[\tilde{r}_{m,t+1}] = \lambda^*_{Mt}$, while the portfolio s, that has a B_{Sst} of *one*, with B_{Mst} and β_{Hst} equal to *zero*, implies that $E_t[\tilde{r}_{s,t+1}] = \lambda^*_{St}$, and the portfolio, h that has a B_{Hht} of *one*, with B_{Mht} and β_{Sht} equal to *zero*, implies that $E_t[\tilde{r}_{h,t+1}] = \lambda^*_{Ht}$. Thus, we get:

$$E_t[\tilde{r}_{i,t+1}] = \mathrm{B}_{Mit}E_t[\tilde{r}_{m,t+1}] + \mathrm{B}_{Sit}E_t[\tilde{r}_{s,t+1}] + \mathrm{B}_{Hit}E_t[\tilde{r}_{h,t+1}]. \quad (25)$$

Eq. (25) is now quite similar to the main prediction of the standard three-factor model as expressed by Eq. (2a).

Eq. (24) or (25) represents our second result. It says that the expected excess return of an asset is linearly related to three factor sensitivities (*betas*) associated with *market-returns*, *size*, and *book-to-market*. Here, the linearity of the risk-return relation predicted by Eq. (24) or (25) is not merely a direct implication of a (predetermined) assumption that supposes a linear return generating process. It is a mathematical result obtained from the covariance definition, using basic algebraic manipulations. In other words, the linearity of our risk-return prediction is not predetermined, or imposed, by a subjective restrictive assumption.

In relaxing the linearity assumption, the extension model improves the robustness of the initial model from a theoretical point of view. It reveals that the main prediction of the Fama-French three factor model (as expressed by Eq. 2) cannot be attacked based on a restrictive assumption that arbitrarily supposes (a priori) that the return generating process (as expressed by Eq. 1) is necessarily linear. In our understanding, in science or economics, if we can relax a model's restrictive assumption, without adding a new assumption, and without modifying the main results, then we improve its theoretical solidity.

4. Conclusion

In this note, we showed that the linearity assumption of the return generating process can be relaxed in the three-factor model.

The main contributions of this paper can be summarized as follows. First, this paper indicates that a risk-return relationship can be expressed with a *general beta* obtained from a general variable positively correlated with returns (see Eq. (19)). Second, it demonstrates that this prediction can be obtained without assuming the existence of an unrealistic security, such as the *Arrow-Debreu security*, and without specific equilibrium or arbitrage restrictions. Third, it reveals that the restrictive linearity assumption of the return generating process is not necessary to express a risk-return relationship with three risk measures, associated to *market*, *size* and *book-to-market* factors (see Eq. (24) or Eq. (25)).

Overall, this paper contributes to improve the robustness of the Fama-French model, from a theoretical point of view.

References

Barry, C.B., Goldreyer, E., Lockwood, L. and M. Rodriguez (2002) "Robustness of size and value effects in emerging equity markets, 1895-2000" *Emerging Markets Review* **3**, 1-30.

Basiewicz, P.G. and C.J. Auret (2009) "Another look at the cross-section of average returns on the JSE" *Financial Analysts Journal* **38**, 23-38.

Berk, J.B. (1997) "Necessary conditions for the CAPM " *Journal of Economic Theory* **73**, 245-257.

Campbell, J.Y. (2018) *Financial decisions and markets: A course in asset pricing*, Princeton University Press, Princeton.

Carhart, M. (1997) "On persistence in mutual fund performance" *The Journal of Finance* **52**, 57-82.

Cochrane, J.H (2005) *Asset pricing*, Princeton University Press, Princeton.

Connor, G., Hagmann, M., Lockwood, L. and O. Linton (2012) "Efficient semiparametric estimation of the Fama-French model and extensions" *Econometrica* **80**, 713-754.

Connor, G. and O. Linton (2007) "Semiparametric estimation of a characteristic-based factor model of common stock returns" *Journal of Empirical Finance* **14**, 694-717.

Cox, S. and J. Britten (2019) "The Fama-French five-factor model: Evidence from the Johannesburg Stock Exchange" *Investment Analysts Journal* **48**, 240-261.

Fama, E.F. (1971) "Risk, return, and equilibrium" *Journal of Political Economy* **379**, 30-55.

Fama, E.F. and K. French (1993) “Common risk factors in the returns on stocks and bonds” *Journal of Financial Economics* **33**, 3-56.

Fama, E.F. and K. French (1995) “Size and book-to-market factors in earnings and returns” *The Journal of Finance* **50**, 131-155.

Fama, E.F. and K. French (2004) “The capital asset pricing mode: Theory and evidence” *Journal of Economic Perspectives* **18**, 25-46.

Fama, E.F. and K. French (2015) “A five-factor asset pricing model” *Journal of Financial Economics* **116**, 1-22.

Fama, E.F. and K. French (2017) “International tests of a five-factor asset pricing model” *Journal of Financial Economics* **123**, 441-463.

Foye, J. (2018) “A comprehensive test of the Fama-French five-factor model in emerging markets” *Emerging Markets Review* **37**, 199-222.

Kwon, Y.K. (1985) “Derivation of the capital asset pricing model without normality or quadratic preference: A note” *The Journal of Finance* **40**, 1505-1509.

Lintner, J. (1965) “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets” *Review of Economics and Statistics* **47**, 13-37.

Merton, R.C. (1973) “An intertemporal capital asset pricing model” *Econometrica* **41**, 867-887.

Nielsen, L.T. (1990) “Equilibrium in CAPM without a riskless asset” *The Review of Economic Studies* **10**, 349-365.

Ross, S. (1976) “The arbitrage theory of asset pricing” *Journal of Economic Theory* **13**, 341-360.

Ross, S. (1978) “Mutual fund separation in financial theory – The separating distributions” *Journal of Economic Theory* **17**, 254-286.

Shalit, H. and S. Yitzhaki (2009) “Capital market equilibrium with heterogeneous investors” *Quantitative Finance* **9**, 757-766.

Sharpe, W. F. (1964) “Capital asset prices: A theory of market equilibrium under conditions of risk” *The Journal of Finance* **19**, 425-442.