A Statistical Factor Asset Pricing Model Versus the 4-Factor Model

(Modelo Estatístico de Apreçamento de Ativos Versus o Modelo de Quatro Fatores)

Verônica de Fátima Santana*
Alex Augusto Timm Rathke**

Abstract

This research aims to compare the performance of a statistical factor asset pricing model with the Fama-French-Carhart 4-factor model. We perform a Principal Component Analysis (PCA) to extract latent risk factors using data of stocks listed on B3 from 2001 to 2015. We test the abilities of the two models to explain assets' returns both in the time-series and in the cross-section dimension. We found that the statistical factor models generates statistically significant abnormal returns in the time-series analysis while the 4-factor model does not. In the cross section dimension, neither model generates significant abnormal returns but they also are not able to generate positive risk premia. Similar results are found if we consider different sets of time and assets. Therefore, although the 4-factor model performs slightly better in the set of tests, neither of the models can be considered fully adequate to explain expected returns of assets in the Brazilian stock market.

Keywords: 4-Factor Model; Principal Component Analysis; Risk Factors; Abnormal Returns.

Códigos JEL: G12.

Resumo

Esta pesquisa tem como objetivo comparar a performance de um modelo estatístico de apreçamento de ativos com o modelo de quatro fatores de Fama-French-Carhart. Aplicamos uma Análise de Componentes Principais (PCA) para extrair fatores de risco latentes usando dados de ações listadas

Submetido em 3 de abril de 2017. Reformulado em 11 de setembro de 2018. Aceito em 7 de novembro de 2018. Publicado on-line em 18 de janeiro de 2018. O artigo foi avaliado segundo o processo de duplo anonimato além de ser avaliado pelo editor. Editor responsável: Márcio Laurini.

*Universidade de São Paulo, São Paulo, SP, Brazil. E-mail: veronica.santana@usp.br

**Universidade de São Paulo, São Paulo, SP, Brazil. E-mail: rathke444@yahoo.com.br.

Rev. Bras. Finanças (Online), Rio de Janeiro, 16, No. 4, Oct. 2018, pp. 545-572
ISSN 1679-0731, ISSN online 1984-5146

©2018 Sociedade Brasileira de Finanças, under a Creative Commons Attribution 3.0 license - <http://creativecommons.org/licenses/by/3.0>

na B3 de 2001 a 2015. Testamos a habilidade dos dois modelos em explicar os retornos dos ativos tanto na dimensão das séries de tempo quanto na cross-section. Encontramos que o modelo estatístico gera retornos anormais estatisticamente significantes na série de tempo enquanto o modelo de quatro fatores não. Na dimensão da cross-section, nenhum dos modelos gera retornos anormais, porém, eles também não geram prêmios de risco positivos. Resultados semelhantes são encontrados se consideramos diferentes ativos e períodos. Portanto, apesar de o modelo de quatro fatores performar um pouco melhor nos testes, nenhum dos modelos pode ser completamente adequado para explicar os retornos dos ativos no mercado de ações brasileiro.

Palavras-Chave: Modelo de Quatro Fatores; Análise de Componentes Principais; Fatores de Risco; Retornos Anormais.

1. Introduction

This research aims to analyze how well a statistical factor model prices Brazilian stock returns compared to the 4-factor model as developed by Fama and French (1993, 1996), and Carhart (1997). To do so, we employ a Principal Component Analysis in a set of stocks listed on the São Paulo Stock Exchange (B3), extracting latent common components which are then considered as risk factors in a pricing model. We test the models' performance for pricing industry portfolios, analysing the implications of they being true in both the time series dimension, that is, investigating whether all intercepts are zero — risk factors can explain assets' returns over time — and in the cross-section dimension, analyzing whether the cross-section intercepts are zero — risk factors can explain the differences in returns among different assets — and whether the risk premia are positive according to the Fama-Macbeth Fama and MacBeth (1973) procedure.

For several decades, the theoretical and empirical research in Finance has been seeking to understand what drives investment decisions and, consequently, to derive the laws that govern market prices. Starting with the mean-variance analysis of Markowitz (1952) and the development of the CAPM by Sharpe (1964), Lintner (1969) and Black (1972), the asset pricing literature comes from a long way. While in the decades of 1960 and 1970 the CAPM was empirically validated, the next decades brought the documentation of several

anomalies, that is, several patterns of expected returns the CAPM framework was not able to explain. These anomalies gave rise to a line of research on multi-factor asset pricing models, which are mostly empirically motivated; notwithstanding, they can be fitted into the more sophisticated theoretical framework of the Intertemporal CAPM Merton (1973) and the Arbitrage Pricing Theory (APT) Ross (1976), that allows for multidimensional sources of risk.

The CAPM is an attractive model to explain how market risk is related with expected returns and investment risk, however early empirical analyses provided rather poor estimates. In order to advance in the subject, Fama and French (1993, 1996) propose a model with three factors to explain market returns: the original CAPM market return in addition to book-to-market ratio (BTM) and size. The authors explain that these two additional factors are able to capture the major variances that are not explained solely by the market return. Later, Jegadeesh and Titman (1993) observe that past returns of market assets also bear significant trends that are able to explain further returns, i.e., they find that shares with higher past returns are more likely to provide abnormal positive returns, in comparison with shares with lower past returns. From the findings on this additional factor, so called momentum, Carhart (1997) develops a deeper analysis of this effect on empirical predictions, so to propose its inclusion as a fourth factor on the Fama and French (1993, 1996) 3-factor model, yielding the well-known 4-factor asset pricing model. It has been widely applied in several studies, especially on the investigation of additional variables affecting asset prices, and on the assessment of most powerful versions of the model.

While the empirically chosen factors is the dominant approach to identify relevant risk components, other researchers have proposed different approaches, such as choosing macroeconomics variables as risk factors (Chen et al. (1986)) and even statistically-chosen factors. Arguing on the advantage of statistical techniques to find underlying, or latent, variables correlated with the movement of returns, some suggest applying Principal Component Analysis (PCA) for identifying common risk factors, such as Roll and Ross (1980) and Connor and Korajczyk (1988). In a PCA, existing n variables are reduced through orthogonal transformation into a set of $m \leq n$ linear uncorrelated factors bearing a relevant share of the variance of the original

variables. Studies show that this approach improves the estimation of large covariance matrices (Alexander (2000)), and provides portfolios with lower risk, in comparison with empirically-identified factors (Engle et al. (1990)).

Despite the simplicity advantage of the use of PCA for portfolio composition, its application on the analysis of asset pricing in Brazil is lacking, for the predominant approach is still the 4-factor model. Studies show that the variances on asset returns are explained by the 4 factors on the baseline model, however the significance of each factor is varying across studies. Rogers and Securato (2009) analyze the Brazilian market and compare three models: the original CAPM, Fama-French modification with 3 factors, and the Reward Beta model from Bornholt (2007). They find that the Fama-French version of CAPM carries higher explanatory power to drive assets returns, however the BTM factor is not significant. Matos and Rocha (2009) analyze the fit of the baseline 4-factor pricing model on the returns of mutual funds in the Brazilian market. Their results indicate that the performance of Brazilian mutual funds regarding pricing and returns forecasts depends not only on the model factors, but also on particular features of each fund. In special, Matos and Rocha (2009) find that the 4-factor model provides significant estimates for investment funds with higher stockholder equity and larger deviation from the market risk parameter, while it does not provide appropriate estimates for other types of mutual funds.

Chague (2007) analyzes the application of the original CAPM and Fama-French modification with 3 factors in Brazilian stock market for the period of 1999-2007, and compares the outcomes with U.S. stock market for the same period, as a benchmark comparison. The results indicate the Fama-French approach has stronger explanatory power in Brazil, however it provides loose estimates in comparison with results in U.S. market. Under the analysis, this outcome is attributed to the absence of clear-cut anomalies in Brazilian firms regarding firms size. Mussa et al. (2012) analyze the application of the 4-factor model for the returns on Brazilian stock market from 1995 to 2006, considering twelve portfolios. Results indicate the traditional market factor is always significant, however it does not explain comprehensibly the variation in stock returns. In this line, they explain the four factors have complementary explanatory power in the Brazilian context.

When comparing a statistical factor model and the 4-factor model, we found, in general, that the latter is better able to price assets in the time-series dimension. The GRS statistic (Gibbons et al. (1989), Campbell et al. (1997)) for the 4-factor model is smaller than the statistical factor model's (1.062 *versus* 2.794), so that it does not reject the null hypothesis of jointly zero time-series intercepts, while the statistical factor model does. Regarding the ability to explain why different assets have different levels of returns, both models generate alphas statistically indistinguishable from zero, indicating that no returns are left unexplained; however, none of the risk premia estimates are positive for both models. Although, again, the 4-factor model seems closer to attend the cross-section implications than the statistical model, all t statistics are outside the rejection area. However, it is necessary to point the presence of survivorship bias in the data as a limitation of this research. For the PCA analysis we need continuous series so we could only include active stocks in the analysis, so we can only conclude about the models' abilities to price surviving assets.

This paper provides two main contributions to the literature of asset price modeling in Brazil. First, it refers to a distinct analysis when comparing to the predominant approach of the 4-factor model. We show that returns can also be estimated by means of statistical factors, rather than empirically-supported factors. Although this approach deviates from the traditional CAPM-empirical rationale, it is consistent with the abstract theoretical view of the factor model construction, for it shows that the movement of variables is explained by a few factors bearing a relevant portion of variance. Second, we complement the literature on asset pricing in Brazil, for we provide new results that are comparative to the investigation of which empirical factors are relevant to explain investment returns in the Brazilian market.

The remaining of the paper is structured as follows. Section 2 reviews the asset pricing literature from the CAPM, passing through the time-series and cross-section tests of the model, to the multifactor models. Section 3 describes the data and the multifactor tests, while section 4 compares the time-series and cross-section performance of the models. Finally, section 5 gives some concluding remarks.

2. Factor Pricing Models

Suppose $i = 1, 2, \dots, I$ investors and $j = 1, 2, \dots, N$ assets, the decision to maximize the terminal wealth $\tilde{w}_{i,1}$ investing both in risky assets and in a risk-free asset R_f , given the initial wealth $w_{i,0}$ as the budget restriction, the investor's first order condition is

$$\text{Cov}[u'_i(\tilde{w}_{i,1}), \tilde{R}_j] = -E[u'_i(\tilde{w}_{i,1})][E(\tilde{R}_j) - R_f], \quad (1)$$

Assuming either joint normality or quadratic utility, equation (1) yields the Capital Asset Pricing Model (CAPM) with the aggregate wealth portfolio as the single risk component for any risky asset j :

$$E[\tilde{R}_j] = R_f + \frac{\text{Cov}[\tilde{R}_m, \tilde{R}_j]}{\text{Var}[\tilde{R}_m]} E[\tilde{R}_m] \quad (2)$$

which is credited to Sharpe (1964) and Lintner (1969).

In the following decades, the CAPM went through a battery of tests to evaluate its empirical validity. Despite the critique of Roll (1977) that the CAPM's theory cannot be empirically tested, several authors developed strategies and statistics to test the relationship between risk and return given by linear factor models.

Fama and MacBeth (1973) propose a two-step procedure to test the implications of the CAPM in a cross-sectional approach. In the first step, the authors run time-series regressions to obtain $\beta_{jm} = \text{Cov}[\tilde{R}_m, \tilde{R}_j] / \text{Var}[\tilde{R}_m]$ for each j . In the second step, for each t they run the cross-sectional regression of each asset's excess returns, denoted by Z_j , against the β_{jm} obtained in the first step:

$$Z_{jt} = \gamma_{0t} + \gamma_{1t}\beta_{jm} + \eta_{jt}, \quad (3)$$

forming a time-series of the coefficients $\{\hat{\gamma}_{0t}, \hat{\gamma}_{1t}\}_{t=1}^T$. Defining $\gamma_0 = E[\gamma_{0t}]$ and $\gamma_1 = E[\gamma_{1t}]$, the CAPM can be tested using:

$$t(\gamma_k) = \frac{\gamma_k}{\sigma_{\gamma_k}}, \quad (4)$$

where $\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^T \hat{\gamma}_{kt}$ and $\hat{\sigma}_{\gamma_k}^2 = \frac{1}{T(T-1)} \sum_{t=1}^T (\hat{\gamma}_{kt} - \hat{\gamma}_k)^2$.

The empirical validity of the CAPM implies that $\gamma_0 = 0$ and that $\gamma_1 > 0$, which is a positive risk premium. In order to the Fama-MacBeth procedure to be valid, the time series of γ must be iid

with mean γ_k . However, since the betas used in the second step are random (estimates from the first step), the independence assumption is violated. Although the measurement error is lower the larger is T so that the second step is T -consistent, it cannot be neglected even in large samples, as Shanken (1992) argues. The author then proposes an adjustment for calculating the standard errors of $\hat{\gamma}$:

$$\tilde{\sigma}_{\gamma_k}^2 = \left(1 + \frac{(\hat{\mu}_m - \hat{\gamma}_0)^2}{\hat{\sigma}_m^2}\right) \hat{\sigma}_{\gamma_k}^2. \quad (5)$$

Recalling the critique of Roll (1977) and the Fama-MacBeth procedure, Gibbons et al. (1989) argue that it is not the CAPM's theory which is being tested. Since the theory is equivalent of the market portfolio being mean-variance efficient, the authors develop a test for evaluating whether a given portfolio is ex-ante mean-variance efficient. To test if all N alphas in the time-series regressions $Z_{jt} = \alpha_j + \beta_{jm}Z_{mt} + \epsilon_{jt}$ are jointly equal to zero is actually the test of a joint hypothesis: H_0 : $\alpha = 0$ and m is on the efficient frontier against H_a : $\alpha \neq 0$ and m is on the efficient frontier. If Z_{mt} is mean-variance efficient, then $E[Z_{jt}] = \beta_{jm}E[Z_{mt}]$, so that the hypothesis H_0 : $\alpha_j = 0$, $\forall j = 1, \dots, N$, can be tested using Hotelling's T^2 test, which is the multivariate generalization of the t -test, from which the J_0 statistics can be derived (Gibbons et al. (1989)):

$$J_0 = T \left(1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim \chi_N^2, \quad (6)$$

where $\hat{\alpha}$ is the vector $N \times 1$ of alphas and $\hat{\Sigma}$ is the $N \times N$ covariance matrix of the time-series residuals multiplied by $1/T$ (unbiased estimator of Σ).

Under joint normality, the finite sample equivalent of the J_0 statistics (Gibbons et al. (1989), Campbell et al. (1997)) is:

$$J_1 = \frac{(T - N - 1)}{N} \left(1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{N, T-N-1}. \quad (7)$$

From this testing framework, several authors in the 1980 and 1990 decades found consistent evidences that the empirical version of the CAPM generates positive alphas both in the time-series and

in the cross-section dimension. These evidences, then called anomalies, were gathered throughout the years to form empirically based risk-factors to complement the market-risk captured by the market portfolio to explain stock returns, generating the multi-factor asset pricing models.

The size effect was firstly documented by Banz (1981), who found a strong negative relation between firms' size, defined by the market value, and firms' stock returns, showing that small firms consistently generates higher returns than large firms. Bhandari (1988) found, then, that when including both market betas and the size effect, firms' leverage also help to explain the cross section of expected average returns.

The evidences brought by Stattman (1980), Rosenberg et al. (1985) and Chan et al. (1991) formed the documented book-to-market (BTM) effect, which is the systematic evidence that growth firms, that is, firms with lower BTM have lower average returns while value firms, that is, firms with higher BTM have higher average returns. Additionally, Basu (1983) documents that earnings-price (E/P) ratios also explain the levels of stock returns.

The reasons for these documented effects lie mainly in the argument that they are capturing omitted risk factors. Ball (1978), for instance, argues that E/P is likely higher for stocks with higher risks and, consequently, higher expected returns. What the sources of such risks would be, then, were not yet defined. For Fama and French (1992), the argument of Ball (1978) also apply to the size, leverage and BTM effects. According to the authors, all these variables can be seen as a way to scale stock prices, as a way to extract information from prices about risk and returns. Further, the authors argue that some of them might capture the same risk, so they can be redundant when jointly considered.

Fama and French (1992) show that the positive relationship between market betas from the CAPM and expected returns in the period pre-1969, as documented by Black (1972) and Fama and MacBeth (1973), has disappeared in the period from 1963 to 1990, even when β is considered alone in the equation. So, the basic prediction from the CAPM does not hold. On the other hand, the relationship between returns and size, leverage, E/P and BTM are strong. Analyzing the joint role of these variables, Fama and French (1992)

found that the beta from the CAPM does not help to explain the cross-section of average returns, and the combination of Size and BTM seems to absorb the roles of leverage and E/P.

Fama and French (1992) suggest stocks' risks are multidimensional. While some offer behavioral explanations (DeBondt and Thaler (1987), Lakonishok et al. (1994), Haugen (1995)) for these empirical results, they can be indicating the need of a more complex model, since the CAPM may have oversimplifying assumptions. This line of thought gave rise to the multi-factor theories of the Intertemporal CAPM (ICAPM) (Merton (1973)) and the Arbitrage Pricing Theory (APT) (Ross (1976)).

Based on the idea that idiosyncratic shocks can be arbitrated away and, therefore, should not be priced, Ross (1976) shows that there is at least an approximate beta pricing model with k factors. However, the APT theory remain silent on what variables should be such factors. The ICAPM theory parts from an intertemporal portfolio choice problem, so that if investments opportunities are time-varying according to k state variables, expected returns are described, in equilibrium, by a $(k + 1)$ -factor beta-pricing model. So, although multi-factor models are mainly empirically-based, the APT and ICAPM theories offer theoretical fundamentals for the existence of multidimensional sources of risk, allowing the existence of multi-factor asset pricing models. Formally, a multi-factor beta-pricing model exists if there are factors $\tilde{F} = (\tilde{f}_1, \dots, \tilde{f}_k)'$, a constant R_z and a $k \times 1$ vector γ such that for each $\tilde{R}$:

$$E[\tilde{R}] = R_z + \gamma' \Sigma_F^{-1} \text{Cov}[\tilde{F}, \tilde{R}], \quad (8)$$

where $\beta = \Sigma_F^{-1} \text{Cov}[\tilde{F}, \tilde{R}]$ is the vector of betas, which are the quantities of risk, γ is the vector of risk premia, which is the price of a unit of risk, and R_z is the return of a zero-beta portfolio, so that if exists a risk-free asset, then $R_z = R_f$.

Fama and French (1993) argue that although Size and BTM are not state variables per se, as would determine the ICAPM theory, the patterns of returns according to these variables reflect non-identified state variables that covary with stock returns which are not captured by the market risk. Based on the evidence that returns covary inside each category of firms, Fama and French (1993, 1996) propose the 3-factor model, including the market portfolio as one source of risk

along with two portfolios that mimics variation in Size and BTM. Specifically, the authors built Size and BTM portfolios by double-sorting the stocks of their sample, dividing it into Big ($> 50\%$) and Small ($< 50\%$) and into High BTM ($> 70\%$), Medium BTM ($< 70\%$ and $> 30\%$) and Low BTM ($< 30\%$), creating the risk factor *SMB* (small minus big), which is the difference between the returns of small and big stocks, and *HML* (high minus low), which is the difference between the returns of value stocks (high BTM) and growth stocks (low BTM). So, the 3-factor model for expected returns is:

$$R_j - R_f = a_j + b_j(R_m - R_f) + s_jSMB + h_jHML + e_j. \quad (9)$$

Fama and French (1996) show that the 3-factor model explain several CAPM anomalies previously documented by other authors. The authors show that *HML* captures variations according to E/P, cash flow/price and sales growth. The model also captures the reversal effect, which is the tendency of firms with long-term high (low) returns to present future low (higher) returns, documented by DeBondt and Thaler (1985): loser stocks (low long-term past returns) tend to have positive *SMB* and *HML* slopes and higher future average returns, while winner stocks (high long-term past returns) tend to have negative slopes on *HML* and low future returns. However, as also documented by Fama and French (1996), the 3-factor model cannot capture the continuation, or momentum, anomaly, which is the tendency that short-term winners (losers) have to continuing yielding high (low) returns, as documented by Jegadeesh and Titman (1993).

Carhart (1997) proposes, then, a fourth factor to capture the momentum effect, known as *WML* (winners minus losers), which is the difference between the returns of a portfolio composed by short-term winner stocks and a portfolio composed by short-term losers stocks. Therefore, the 4-factor model can be expressed by:

$$R_j - R_f = a_j + b_j(R_m - R_f) + s_jSMB + h_jHML + w_jWML + e_j. \quad (10)$$

Several other risk factors have been proposed in the literature, such as the liquidity risk factor (Pastor and Stambaugh (2003), Acharya and Pedersen (2005)), idiosyncratic volatility (Ang et al. (2006, 2007)), betting against beta (Frazzini and Pedersen (2014)) and accruals and profitability (Novy-Marx (2013), Fama and French (2015, 2016), Ball

et al. (2016)). However, the 4-factor model by Fama-French-Carhart remains the most popular one.

Besides the empirically chosen risk factor seen above, some authors followed different approaches for developing multi-factor models. Chen et al. (1986), for instance, explore how macroeconomic variables fit as state variables that influences stock returns. Analyzing industrial production, inflation, bonds' risk premia, term-structure of interest rates and stock market indexes, the authors found all economic variables are able to explain the cross-sectional differences in stock returns, while the stock market indexes are only significant in the time-series dimension.

Further, other authors followed a statistical approach. Roll and Ross (1980) use statistical factor analysis as a way to identify latent factors to test the APT. The authors extracted three and four factors from individual returns for the period of 1962 to 1972, finding that such factors are indeed priced and adding other variables do not increase explanatory power. Connor and Korajczyk (1988) also used the principal component analysis as a way to test the APT. Extracting five common factors, the authors compared their multi-factor with the CAPM, finding the statistical factor model better describes assets' expected returns, although some mispricing still remains.

Therefore, although searching factors by an empirical strategy is the most common approach for building multi-factor models, these other approaches can also yield models with good explanatory power for the expected average returns.

3. Data, Models and Tests

Following previous works, we test the statistical and the 4-factor model in monthly excess log-returns of industry portfolios. The data initially selected consisted on all available firms listed on the São Paulo Stock Exchange (Bovespa) from 2001 to 2015 obtained at Economatica, and we built the industry portfolios according to Bovespa's sub-sector classification with 41 categories on the 382 initially available firms. However, since not all 41 industry portfolio have available price data to calculate returns for all months, we firstly selected the industries with more than 80% of months with available data and then we selected the longest contiguous time-series data, ending up with 109 months, from 2006-December to 2015-December,

of 30 industries' returns. As the risk-free rate, we consider the Selic Over rate's monthly log-return, obtained at IPEA's website.

To build the statistical factor model, we went through the following steps. First, we calculated the monthly excess log-returns for the 382 individual stocks, and then selected the same 109 months of the data as the industry portfolios. Second, from the 382 individual stocks, we selected the mostly liquid stocks. To do so, we kept only the stocks from which the latest available liquidity measure from Economática is over 0.01. This filter is necessary in an emerging stock market like the Brazilian one, where the majority of listed stocks are not continuously traded and, therefore, do not carry much information in their prices. With this filter, we would keep 165 stocks, however, even these most liquid stocks have several continuous months with missing data from 2006-12 to 2015-12 (specially in the first years of the sample). The PCA analysis does not allow for missing data, so we excluded these remaining stocks with missing returns, so that our final sample for the PCA consisted of 109 monthly excess log-returns for 78 individual stocks.

Third, we evaluated the adequateness of the Principal Component Analysis according to the Kaiser-Meyer-Olkin (KMO) statistics, which evaluates the variables' correlations and partial correlations with an index from 0 to 1, and the Bartlett's sphericity test, which test the null hypothesis that the correlation matrix of the data is an identity matrix. Higher values of the KMO statistics and the rejection of the null hypothesis of the Bartlett's sphericity test indicate that there is enough correlation in the data so the extraction of a few common components is possible. The analysis on the 78 individual stocks yielded a KMO statistics of 0.92, which is very high, and a Bartlett's sphericity test statistics of 11646.43, which is very far of the non-rejection area.

Fourth, we proceed to the PCA itself. The principal component analysis consists in extracting a few factors from a set of variables according to their common variation, in order to summarize the information contained in that larger set of data. This technique allows identifying latent (non-observable) factors that carry common information from the original data. Formally, the first principal component can be defined as $f_{1t} = x_1^{*'} R_t$, where x_1^* is the argument that

maximizes the $x_1' \hat{\Sigma} x_1$:

$$x_1^* = \underset{x_1}{\operatorname{argmax}} x_1' \hat{\Sigma} x_1 \quad \text{s.t.} \quad x_1' x_1 = \mathbf{1}, \quad (11)$$

where $\hat{\Sigma}$ is a consistent estimator of the sample covariance matrix of the individual returns. The second principal component is $f_{2t} = x_2^{*'} R_t$, where:

$$x_2^* = \underset{x_2}{\operatorname{argmax}} x_2' \hat{\Sigma} x_2 \quad \text{s.t.} \quad x_2' x_2 = \mathbf{0}, \quad (12)$$

and so on until the k -th factor, where $k \leq N$, being N is the number of individual returns used in the analysis. In other words, each extracted factor is a linear combination of the returns, whose weights maximizes their total variance, subject to the restriction that all factors are orthogonal to each other.

Table 1 shows that the first 10 components account for 81% of total variance of the original 78 individual stock returns. Extracting only the components with eigenvalues higher than one, as plotted in figure 1, we have six common factors that summarize 76.6% of all information from the 78 stocks.

Table 1
PCA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Eigenvalues	52.603	2.268	1.410	1.351	1.076	1.049	0.957	0.899	0.830	0.752
Proportion of Variance	67.4%	2.9%	1.8%	1.7%	1.4%	1.3%	1.2%	1.2%	1.1%	1.0%
Cumulative Proportion	67.4%	70.3%	72.2%	73.9%	75.3%	76.6%	77.8%	79.0%	80.1%	81.0%

Figure 2 shows the factor loadings of each individual stocks in the first two extracted components. The first factor can be roughly interpreted as an average of all stocks, mimicking an equally-weighted market portfolio. The other five factors, as usually is the case in PCA, are difficult to interpret.

Having extracted six latent factors from the PCA $F = (f_1, \dots, f_6)'$, that are expected to capture common risk factors for the Brazilian stocks, we define the statistical factor model as:

$$Z_{jt} = \alpha_j + F' \beta_j + \epsilon_{jt} = \alpha_j + \beta_{j1} f_1 + \beta_{j2} f_2 + \beta_{j3} f_3 + \beta_{j4} f_4 + \beta_{j5} f_5 + \beta_{j6} f_6 + \epsilon_{jt}, \quad (13)$$

Figure 1
Scree Plot

where the Z_{jt} is the excess log-return of the industry portfolio $j = 1, \dots, 30$ at month $t = 1, \dots, 109$, the slopes β_{jk} , $k = 1, \dots, 6$ are the estimated quantities of risk of each factor k for each industry portfolio j . If, for each portfolio j , the model captures all common risk variation, then $\alpha_j = 0$, $\forall j$. So we use the multivariate version of the J_1 GRS statistics Cochrane (2005) to test this hypothesis:

$$J_1 = \frac{T - N - K}{N} \left(1 + E_T[F]' \hat{\Sigma}^{-1} E_T[F] \right)^{-1} \hat{\alpha}' \hat{\Sigma}_f^{-1} \hat{\alpha} \sim F_{N, T-N-K}. \quad (14)$$

To test whether the six latent factors captures all variation in expected returns, we conduct a Fama-MacBeth Fama and MacBeth (1973) analysis. First, we estimate the quantities of risk of each factor for each portfolio β_{jk} running $j = 1, \dots, 30$ time-series regressions as in equation (13). In the second step, we run $t = 1, \dots, 109$ cross-section regressions:

$$Z_{jt} = \gamma_{0t} + \gamma_{1t}\beta_{j1} + \gamma_{2t}\beta_{j2} + \gamma_{3t}\beta_{j3} + \gamma_{4t}\beta_{j4} + \gamma_{5t}\beta_{j5} + \gamma_{6t}\beta_{j6} + \eta_{jt}, \quad (15)$$

to estimate the risk premia $\gamma + k$ for each factor at each month, and analyze the time-series of γ_0 and each risk premia, as described in section 2, to test the hypotheses that $\gamma_0 = 0$ and $\gamma_k > 0$, calculating standard errors according to the multivariate version of Shanken

Figure 2

Loading Plot of the first Two Factors

(1992) correction Cochrane (2005):

$$\hat{\sigma}_{\gamma_k}^2 = (1 + \gamma' \Sigma_f^{-1} \gamma) \hat{\sigma}_k^2, \quad (16)$$

where Σ_f is the variance-covariance matrix of the factors.The 4-factor model can be expressed in terms of excess returns Z as:

$$Z_{jt} = \alpha_j + b_j Z_{mt} + s_j SMB_t + h_j HML_t + w_j WML_t + \epsilon_{jt}, \quad (17)$$

where the Z_{jt} is the excess log-return of the industry portfolio $j = 1, \dots, 30$ at month $t = 1, \dots, 109$, so that the GRS statistics in equation (14) can be used to test if $\alpha_j = 0$, $\forall j$. Once estimated the quantities of risk of each factor of the model b_j, s_j, h_j and w_j , the second step of the Fama-MacBeth analysis is:

$$Z_{jt} = \gamma_0 t + \gamma_{Z_{mt}} b_j + \gamma_{SMB_t} s_j + \gamma_{HML_t} h_j + \gamma_{WML_t} w_j + \eta_{jt}, \quad (18)$$

so we can test the hypotheses that $\gamma_0 = 0$ and that the risk premia $\gamma_k > 0$. The data from the four risk factors $R_m - R_f$, SMB , HML

and WML are calculated and made available by the Brazilian Center for Research in Financial Economics of the University of São Paulo (NEFIN).

4. Empirical Results

4.1 Time-Series Analysis

Tables 2 and 3 shows the 30 time-series regressions for the statistical factor model. The R^2 are usually very high (around 0.8 and 0.9), with a few exceptions, and the first factor f_1 mimicking an market portfolio is highly significant for all industry portfolios. However, nine portfolios present time-intercepts statistically different from zero when individually taken.

Tables 4 and 5 shows the time-series regressions for the 4-factor model. The R^2 are generally much smaller than for the statistical factor (around 0.3 and 0.4), while the first two factors are statistically significant for all portfolios, while only for some of them HML and WML are significant. Individually taken, only three portfolios have statistically significant time-intercepts.

The joint test through the GRS J_1 statistics shows that the time-intercepts of the statistical factor model are not equal to zero, as shown in table 6. The test statistics is too high and, therefore, we reject the null hypothesis of all intercepts being equal to zero. However, the test for the 4-factor model indicates that the intercepts are jointly statistically zero. Therefore, the statistical factor model fails the time-series test while the 4-factor model passes. This indicates that, at least in the time series, the 4-factor model is better able to price assets in the Brazilian stock market, instead of a latent statistically-chosen multifactor model.

Table 2
Time-Series Regressions for the Statistical Factor Model (Industries 2 to 20)

	<i>Dependent variable:</i>														
	Ind2	Ind3	Ind4	Ind5	Ind6	Ind7	Ind8	Ind9	Ind10	Ind13	Ind15	Ind16	Ind17	Ind19	Ind20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
f_1	-13.6*** (0.5)	-13.2*** (0.5)	-14.8*** (0.8)	-11.3*** (0.5)	-13.5*** (0.4)	-15.1*** (0.7)	-12.4*** (0.8)	-15.1*** (0.4)	-11.9*** (0.8)	-11.8*** (0.2)	-14.0*** (0.4)	-10.7*** (0.6)	-11.6*** (0.9)	-12.7*** (0.3)	-12.8*** (0.7)
f_2	0.4 (0.5)	-0.5 (0.5)	-1.4* (0.8)	2.3*** (0.5)	-1.9*** (0.4)	0.1 (0.7)	0.5 (0.8)	-3.6*** (0.4)	0.2 (0.8)	1.6*** (0.2)	-0.9** (0.4)	1.6*** (0.6)	-1.1 (0.9)	-0.4 (0.3)	-2.1*** (0.7)
f_3	0.2 (0.5)	0.7 (0.5)	1.2 (0.8)	1.2** (0.5)	1.1*** (0.4)	0.2 (0.7)	0.5 (0.8)	1.3*** (0.4)	-1.3* (0.8)	-0.8*** (0.2)	1.0** (0.4)	-0.7 (0.6)	-0.6 (0.9)	-0.4 (0.3)	0.5 (0.7)
f_4	-0.3 (0.5)	1.0* (0.5)	2.4*** (0.8)	1.4*** (0.5)	-0.1 (0.4)	1.3* (0.7)	0.4 (0.8)	-0.9** (0.4)	2.4*** (0.8)	0.6*** (0.2)	0.7* (0.4)	0.9 (0.6)	2.5*** (0.9)	0.1 (0.3)	1.6** (0.7)
f_5	-1.2** (0.5)	-0.4 (0.5)	1.0 (0.8)	0.9* (0.5)	0.2 (0.4)	1.4* (0.7)	-0.2 (0.8)	-1.1** (0.4)	-0.3 (0.8)	-0.2 (0.2)	-0.1 (0.4)	0.1 (0.6)	1.0 (0.9)	0.1 (0.3)	0.2 (0.7)
f_6	2.6*** (0.5)	0.1 (0.5)	-0.3 (0.8)	-0.6 (0.5)	0.7** (0.4)	0.6 (0.7)	1.9** (0.8)	2.0*** (0.4)	-0.6 (0.8)	-0.2 (0.2)	1.5*** (0.4)	-0.8 (0.6)	-1.0 (0.9)	-0.1 (0.3)	-0.5 (0.7)
<i>Constant</i>	0.7 (0.5)	-0.2 (0.5)	-0.9 (0.8)	1.7*** (0.5)	-0.2 (0.4)	-0.7 (0.7)	-1.6** (0.8)	-1.2*** (0.4)	0.6 (0.8)	0.3 (0.2)	-0.3 (0.4)	1.3** (0.6)	-0.6 (0.9)	0.2 (0.3)	0.9 (0.7)
Obs.	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109
R ²	0.9	0.9	0.8	0.9	0.9	0.8	0.7	0.9	0.7	1.0	0.9	0.8	0.7	0.9	0.8
Adj. R ²	0.9	0.9	0.8	0.8	0.9	0.8	0.7	0.9	0.7	1.0	0.9	0.8	0.6	0.9	0.8
F St.	109.1***	111.6***	58.4***	96.4***	250.8***	78.2***	40.8***	251.1***	39.3***	461.8***	194.8***	63.4***	32.1***	254.0***	63.4***

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 3

Time-Series Regressions for the Statistical Factor Model (Industries 21 to 41)

	<i>Dependent variable:</i>														
	Ind21	Ind22	Ind23	Ind26	Ind27	Ind28	Ind30	Ind31	Ind32	Ind36	Ind37	Ind38	Ind39	Ind40	Ind41
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
f_1	-12.2*** (0.4)	-14.3*** (0.9)	-12.4*** (0.6)	-14.7*** (0.7)	-9.9*** (3.3)	-14.9*** (0.6)	-12.3*** (1.1)	-13.7*** (0.6)	-13.2*** (0.5)	-13.6*** (0.3)	-12.2*** (0.5)	-12.2*** (0.6)	-14.1*** (0.4)	-11.5*** (0.9)	-13.5*** (1.0)
f_2	-0.5 (0.4)	-1.3 (0.9)	-0.9* (0.6)	-4.1*** (0.7)	-5.8* (3.3)	-1.5** (0.6)	2.0* (1.1)	-0.4 (0.6)	-0.4 (0.5)	-3.0*** (0.3)	-0.01 (0.5)	2.4*** (0.6)	-0.2 (0.4)	1.0 (0.9)	0.2 (1.0)
f_3	-0.8** (0.4)	-2.0** (0.9)	-1.0* (0.6)	-2.8*** (0.7)	-1.1 (3.3)	-2.4*** (0.6)	0.5 (1.1)	0.8 (0.6)	0.1 (0.5)	-2.2*** (0.3)	-0.5 (0.5)	-3.8*** (0.6)	-0.1 (0.4)	-1.1 (0.9)	-2.0** (1.0)
f_4	2.6*** (0.4)	1.8** (0.9)	1.4** (0.6)	1.9*** (0.7)	9.1*** (3.3)	0.5 (0.6)	-0.5 (1.1)	2.5*** (0.6)	1.7*** (0.5)	0.3 (0.3)	2.0*** (0.5)	2.5*** (0.6)	0.4 (0.4)	2.5*** (0.9)	0.3 (1.0)
f_5	0.8** (0.4)	-0.2 (0.9)	0.3 (0.6)	-2.4*** (0.7)	4.4 (3.3)	-0.7 (0.6)	0.4 (1.1)	-0.2 (0.6)	-0.2 (0.5)	-0.4 (0.3)	0.1 (0.5)	1.3** (0.6)	-0.02 (0.4)	-0.7 (0.9)	-0.4 (1.0)
f_6	-0.4 (0.4)	1.6* (0.9)	0.8 (0.6)	-2.4*** (0.7)	-3.2 (3.3)	1.5** (0.6)	-1.3 (1.1)	0.6 (0.6)	-0.6 (0.5)	-0.2 (0.3)	-0.1 (0.5)	2.2*** (0.6)	0.05 (0.4)	1.3 (0.9)	0.9 (1.0)
<i>Constant</i>	-1.1*** (0.4)	-0.8 (0.9)	-0.3 (0.6)	-2.1*** (0.7)	1.1 (3.3)	-2.0*** (0.6)	-1.5 (1.1)	0.4 (0.6)	-1.2** (0.5)	-0.9*** (0.3)	-0.2 (0.5)	-0.8 (0.6)	-0.7 (0.4)	0.2 (0.9)	-1.3 (1.0)
Obs.	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109
R ²	0.9	0.7	0.8	0.9	0.2	0.9	0.6	0.8	0.9	1.0	0.8	0.8	0.9	0.6	0.7
Adj. R ²	0.9	0.7	0.8	0.8	0.1	0.9	0.5	0.8	0.8	0.9	0.8	0.8	0.9	0.6	0.6
F St.	204.3***	47.2***	86.4***	99.5***	3.7***	105.5***	22.9***	90.4***	100.5***	330.5***	93.1***	92.6***	173.9***	27.4***	33.5***

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4

Time-Series Regressions for the 4-Factor Model (Industries 1 to 20)

	<i>Dependent variable:</i>														
	Ind2	Ind3	Ind4	Ind5	Ind6	Ind7	Ind8	Ind9	Ind10	Ind13	Ind15	Ind16	Ind17	Ind19	Ind20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
<i>Z_m</i>	1.0*** (0.2)	0.9*** (0.2)	0.8*** (0.3)	0.6*** (0.2)	1.0*** (0.2)	1.1*** (0.3)	0.8*** (0.2)	1.3*** (0.2)	0.6*** (0.2)	0.7*** (0.2)	1.1*** (0.2)	0.7*** (0.2)	0.6** (0.2)	1.0*** (0.2)	1.1*** (0.2)
<i>SMB</i>	1.2*** (0.3)	1.0*** (0.3)	1.4*** (0.4)	0.7** (0.3)	1.2*** (0.3)	1.3*** (0.4)	1.3*** (0.3)	1.4*** (0.3)	1.5*** (0.3)	0.9*** (0.3)	1.3*** (0.3)	0.7** (0.3)	1.2*** (0.3)	1.0*** (0.3)	1.1*** (0.3)
<i>HML</i>	0.1 (0.3)	-0.01 (0.3)	0.1 (0.4)	-0.2 (0.3)	-0.2 (0.3)	-0.005 (0.4)	-0.2 (0.3)	-0.2 (0.3)	0.4 (0.3)	0.2 (0.3)	-0.1 (0.3)	0.2 (0.3)	0.03 (0.4)	-0.003 (0.3)	-0.1 (0.3)
<i>WML</i>	0.3 (0.3)	0.02 (0.3)	0.1 (0.4)	0.3 (0.3)	-0.03 (0.3)	0.2 (0.3)	0.3 (0.3)	-0.1 (0.3)	0.3 (0.3)	0.3 (0.3)	0.3 (0.3)	0.3 (0.3)	0.1 (0.3)	0.2 (0.3)	0.3 (0.3)
<i>Constant</i>	1.9 (1.3)	1.1 (1.2)	0.8 (1.5)	2.1* (1.2)	1.5 (1.1)	0.9 (1.4)	-0.5 (1.3)	1.0 (1.1)	2.0 (1.2)	1.2 (1.1)	1.0 (1.1)	1.9* (1.1)	0.8 (1.3)	1.3 (1.0)	2.0* (1.2)
Obs.	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109
R ²	0.3	0.3	0.3	0.2	0.5	0.3	0.3	0.6	0.3	0.3	0.4	0.2	0.3	0.4	0.3
Adj. R ²	0.3	0.3	0.3	0.1	0.4	0.3	0.3	0.5	0.3	0.3	0.4	0.2	0.2	0.4	0.3
F St.	12.0***	13.4***	11.2***	5.1***	22.5***	14.0***	11.5***	33.3***	12.1***	10.1***	19.8***	6.6***	9.0***	17.6***	13.6***

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 5

Time-Series Regressions for the 4-Factor Model (Industries 21 to 41)

	<i>Dependent variable:</i>														
	Ind21	Ind22	Ind23	Ind26	Ind27	Ind28	Ind30	Ind31	Ind32	Ind36	Ind37	Ind38	Ind39	Ind40	Ind41
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
<i>Z_m</i>	0.7*** (0.2)	1.1*** (0.3)	0.9*** (0.2)	1.3*** (0.2)	0.4 (0.7)	1.1*** (0.2)	0.6** (0.3)	0.9*** (0.2)	0.7*** (0.2)	1.2*** (0.2)	0.7*** (0.2)	0.5* (0.2)	1.1*** (0.2)	0.4* (0.2)	1.2*** (0.3)
<i>SMB</i>	1.1*** (0.3)	1.3*** (0.4)	1.1*** (0.3)	1.5*** (0.3)	1.8* (0.9)	1.2*** (0.3)	0.8** (0.4)	1.2*** (0.3)	1.3*** (0.3)	0.9*** (0.3)	1.2*** (0.3)	1.1*** (0.3)	1.3*** (0.3)	1.3*** (0.3)	1.3*** (0.4)
<i>HML</i>	0.2 (0.3)	0.2 (0.4)	0.1 (0.3)	0.7* (0.3)	-0.1 (1.0)	0.3 (0.3)	-0.1 (0.4)	-0.6* (0.3)	0.2 (0.3)	0.4 (0.3)	0.1 (0.3)	0.3 (0.4)	-0.1 (0.3)	-0.01 (0.4)	0.1 (0.4)
<i>WML</i>	0.2 (0.3)	0.1 (0.3)	0.2 (0.3)	0.2 (0.3)	0.1 (0.9)	-0.2 (0.3)	0.1 (0.4)	0.1 (0.3)	0.1 (0.3)	-0.2 (0.3)	0.3 (0.3)	0.2 (0.3)	0.2 (0.3)	0.1 (0.3)	0.5 (0.3)
<i>Constant</i>	0.1 (1.1)	1.0 (1.4)	1.0 (1.1)	0.002 (1.3)	3.0 (3.7)	0.1 (1.3)	-0.5 (1.6)	1.7 (1.2)	0.3 (1.2)	1.0 (1.1)	0.9 (1.1)	0.3 (1.4)	0.7 (1.2)	1.6 (1.4)	-0.2 (1.4)
Obs.	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109
R ²	0.3	0.4	0.3	0.5	0.1	0.4	0.1	0.4	0.3	0.5	0.3	0.2	0.4	0.2	0.3
Adj. R ²	0.3	0.3	0.3	0.5	0.03	0.4	0.1	0.3	0.3	0.5	0.3	0.1	0.4	0.2	0.3
F St.	12.3***	14.8***	13.6***	23.1***	1.8	20.3***	4.2***	15.4***	13.1***	24.2***	11.9***	5.7***	18.8***	8.5***	13.0***

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 6
GRS Statistics

	Statistical Factor Model	4-Factor Model
GRS Statistics (J_1)	2.794	1.062
Degrees of Freedom	30, 73	30, 75
Critical Value (5%)	1.615	1.611
P-value	0.0002	0.4047

4.2 Fama-Macbeth Analysis

Table 7 shows the statistics of the second step of the Fama-MacBeth analysis for the two models. The estimates of abnormal returns given by γ_0 and the respective variances and t -statistics show that they are not indistinguishable from zero, indicating that no returns are left unexplained considering the set of risk factors of each model. Both models' corrected t -statistics are outside the rejection area at fair significance levels, although the statistical factor model have values nearer the rejection area. However, the statistical factor model have a lower mean absolute error (MAE), which is expected due to its larger number of factors.

Therefore, both models pass the first implication in the cross-section analysis, which is that $\gamma_0 = 0$. However, the second implication, which is positive risk premia, is not attended by either of the models. None of the risk factors of both models have higher enough corrected t -statistics, so, none of them can be considered different from zero. Therefore, although this result can be due to lack of power of the tests, neither of the models can be considered fully adequate to explain why different assets have different expected returns in the Brazilian stock market.

4.3 Robustness Checks

Using Brazilian data for this type of analysis is a challenge. The extent of available data is very restricted when compared with the U.S. market, for example, where these models were originally developed. In this paper, we started our data in 2001, when the Nefin risk factor series start. Initially we have 41 industries, 384 firms and 180 months, but the data present several missing values cases (only few firms have been consistently traded for a long time) even after

Table 7
Fama-MacBeth Statistics

Statistical Factor Model							
	γ_0	γ_{f1}	γ_{f2}	γ_{f3}	γ_{f4}	γ_{f5}	γ_{f6}
Estimate	5.196	0.435	0.024	0.295	0.160	-0.107	0.108
σ^2	6.929	0.052	0.041	0.018	0.028	0.077	0.026
t Stat.	1.974	1.902	0.117	2.194	0.950	-0.386	0.664
Shanken's (1992) σ^2	9.187	0.069	0.055	0.024	0.038	0.102	0.035
Shanken's (1992) t Stat.	1.714	1.652	0.101	1.905	0.825	-0.335	0.576
MAE	4.208						
4-Factor Model							
	γ_0	γ_{Z_m}	γ_{SMB}	γ_{HML}	γ_{WML}		
Estimate	1.406	-1.252	-0.792	-0.882	1.964		
σ^2	1.797	1.672	2.083	0.454	0.826		
t Stat.	1.049	-0.968	-0.549	-1.310	2.161		
Shanken's (1992) σ^2	2.179	2.028	2.526	0.550	1.002		
Shanken's (1992) t Stat.	0.952	-0.879	-0.498	-1.190	1.963		
MAE	4.865						

accounting for liquidity, so we need to balance the number of firms and the lengths of the series. Therefore, we ended up conducting our analysis in the previous sections using only 30 industries, 78 firms and 109 months. The results found so far could be, therefore, specific for this set of time and firms. Trying to overcome this concern, in this section we present and discuss some results using different periods and number of firms.

In the first robustness analysis we prioritize series length, and work with 180 months, but only 26 industries and 49 firms. The PCA analysis indicates 3 common factors for these 49 assets, and the statistical factor gauges a GRS statistics of 3.590 (p-value: 0.000), rejecting the null hypothesis that the 26 industry time-series intercepts are jointly equal to zero. The Fama-Macbeth analysis indicates a Shanken's corrected t-statistic of 2.15 for γ_0 and none of the risk premia are positive. Therefore, the statistical factor models passes neither the time-series or the cross-section tests in this analysis. The 4-factor model have a GRS statistics of 1.428 (p-value: 0.096) and a Shanken's corrected t-statistic of 2.06 for γ_0 and no positive risk premia. Again, the results are similar to the main analysis, where the 4-factor model does not reject the null hypothesis of jointly zero time-

series intercepts but fails the cross-section pricing tests. Therefore, when we increase the time length but decrease the number of assets, the two models are less likely to pass the tests, but the 4-factor models still performs better if compared to the statistical factor model.

Second, we perform the analysis in two subperiods of 90 months each. We first select the period of jan/2001–jun/2008, in which we have 26 industries and 63 firms. For the period of jul/2008–dec/2015 we have 39 available industries and 183 firms. As mentioned before, the most recent periods have greater availability of data. When compared with the main analysis, the statistical factor model performs worse in each of the subperiods. In the time-series dimension it generates a GRS statistic of 12.978 (p-value: 0.000) in the first subperiod and 15.184 (p-value: 0.000) in the second one. And although the Fama-Macbeth analysis finds no statistically significant abnormal returns among the industries, its risk premia also cannot be considered statistically significant, regardless which subperiod we evaluate.

The 4-factor model also performs worse in each of the subperiods compared to the main analysis. In the time-series dimension it generates a GRS statistic of 1.657 (p-value: 0.055) in the first subperiod and 2.059 (p-value of 0.009) in the second one. So, while the 4-factor model passes the test in the time-series test in the main analysis, it almost fails in the first subperiod and fails in the second one. Finally, the Fama-Macbeth analysis also does not show cross-section intercepts statistically different from zero but does not show any positive risk premia as well. Therefore, as in the main analysis, the cross-section subperiods analyses also show that the two models passes the first implication of zero alphas but not the second implications of positive risk premia.

Therefore, these robustness analyses show worse performances for both the models but, in general, generates the same conclusion that the 4-factor model is better for pricing assets in the time dimension and that the two of them generates no abnormal returns in the cross-section analyses, but do not generate positive risk premia either. Therefore, although the 4-factor model performs slightly better than the statistical factor model neither of them is fully adequate to price Brazilian asset prices.

5. Concluding Remarks

This paper aimed to compare the time-series and cross-section performance of a statistical factor asset pricing model, with latent risk factors obtained through principal component analysis, with the 4-factor model of Fama and French (1993, 1996) and Carhart (1997) to price industry portfolios in the Brazilian stock market.

Extracting six common factors from a sample of 78 liquid stocks listed on the São Paulo Stock Exchange, we found the 4-factor model is better able to price the industry portfolios in the time-series dimension. The statistical factor models generates jointly significant time-series abnormal returns while the 4-factor does not. However, in the cross-section dimension both models generates zero abnormal returns but either generates positive risk premia. Similar results, although even weaker, are found if we consider different time periods and different numbers of industries and firms. Therefore, neither of them can adequately explain the differences between assets' expected returns in the Brazilian capital market.

References

- Acharya, V. V. and Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77:375–410.
- Alexander, C. (2000). A primer on the orthogonal GARCH model. *manuscript ISMA Centre, University of Reading, UK*, 2.
- Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, 61(1):259–299.
- Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2007). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23.
- Ball, R. (1978). Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics*, 6:103–126.
- Ball, R., Gerakos, J., Linnainmaa, J. T., and Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, 121:28–45.

- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9:3–18.
- Basu, S. (1983). The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence. *Journal of Financial Economics*, 12:129–156.
- Bhandari, L. C. (1988). Debt/Equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, 43:507–528.
- Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, 45(3):444–455.
- Bornholt, G. (2007). Extending the Capital Asset Pricing model: The reward beta approach. *Accounting & Finance*, 47(1):69–83.
- Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton, New Jersey.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52:57–82.
- Chague, F. D. (2007). The CAPM and Fama-French models in Brazil: A comparative study. Master's thesis, Fundação Getúlio Vargas, São Paulo.
- Chan, L. K., Hamao, Y., and Lakonishok, J. (1991). Fundamentals and stock returns in Japan. *Journal of Finance*, 46:1739–1789.
- Chen, N., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 59(3):383–403.
- Cochrane, J. (2005). *Asset Pricing*. Princeton, New Jersey.
- Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21(2):255–289.
- DeBondt, W. F. M. and Thaler, R. H. (1985). Does the stock market overreact? *Journal of Finance*, 40:793–805.

- DeBondt, W. F. M. and Thaler, R. H. (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance*, 42(2):557–581.
- Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2):213–237.
- Fama, E. F. and French, K. (1992). The cross section of expected returns. *Journal of Finance*, 47:427–465.
- Fama, E. F. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.
- Fama, E. F. and French, K. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1):55–84.
- Fama, E. F. and French, K. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.
- Fama, E. F. and French, K. (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies*, 29:69–103.
- Fama, E. F. and MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607–636.
- Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111:1–25.
- Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, 57:1121–1152.
- Haugen, R. (1995). *The new finance: The case against efficient markets*. Prentice Hall, Englewood Cliffs, New Jersey.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48:65–91.
- Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49(5):1541–1578.

- Lintner, J. (1969). The aggregation of investor's diverse judgments and preferences in purely competitive security markets. *Journal of Financial and Quantitative Analysis*, 4:347–400.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.
- Matos, P. R. F. and Rocha, J. A. T. (2009). Stocks and mutual funds: common risk factors? *Brazilian Business Review*, 6(1).
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41(5):867–887.
- Mussa, A., Famá, R., and Santos, J. O. (2012). A adição do fator de risco momento ao modelo de precificação de ativos dos três fatores de Fama & French aplicado ao mercado acionário brasileiro. *Revista de Gestão*, 19(3).
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28.
- Pastor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *The Journal of Political Economy*, 111(3):642–685.
- Rogers, P. and Securato, J. R. (2009). Estudo comparativo no mercado brasileiro do Capital Asset Pricing Model (CAPM), modelo 3-fatores de Fama e French e reward beta approach. *Revista de Administração Contemporânea*, 3(1):159–180.
- Roll, R. (1977). A critique of the asset pricing theory's tests. *Journal of Financial Economics*, 4:129–176.
- Roll, R. and Ross, S. A. (1980). An empirical investigation of the Arbitrage Pricing Theory. *The Journal of Finance*, 35(5):1073–1103.
- Rosenberg, B., Reid, K., and Lanstein, R. (1985). Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, 11:9–17.
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360.

- Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, 5(1):1–33.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442.
- Stattman, D. (1980). Book values and stock returns. *The Chicago MBA: A Journal of Selected Papers*, 4:25–45.