ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Bachelor Thesis

[International Bachelor Econometrics and Operations Research]

Model Comparison Based On Alpha

Author: Yujie Li
Student Number: 425429
Supervisor: Karolina Scholtus
Second assessor: dr. Xuan Leng
July 8, 2018

Abstract

Alpha is used as a measure of model comparison as model mispricing, usually by modeling test-asset returns. However, Barillas and Shanken (2017) give an idea that test assets are irrelevant in model comparison and excluded factors are more reliable evidence. In this paper, we replicate the work of Barillas and Shanken (2017), with illustrations of non-nested model comparison. Most of the replication is quite successful, except that the result of model comparison based on the likelihood metric is different, which is acceptable because likelihood and model alpha have different norms in model comparison. As extensions, we further confirm the power of different methods of comparison for nested models, as well as with a nonsensical factor. It turns out that both excluded factors and test assets perform well for nested models, while the likelihood always gives inaccurate results. With a nonsensical factor, excluded factors fail to make a correct conclusion, but the test-asset evidence is still reliable, which gives an opposite opinion with Barillas and Shanken (2017). Another extension is about model comparison with nontraded factors, where we find that the test-asset evidence turns out to be inaccurate due to the effect of mimicking portfolios.

1 Introduction

Alpha is commonly used to measure the performance of mutual funds and all types of investments in modern financial market. It is often considered as the active return on an investment, which indicates the performance of that investment against a benchmark. In asset pricing, alpha can also serve as a measure of model mispricing as the deviation from the expected value determined by benchmark factors. Jensen (1968) in the early paper first raises "Jensen's alpha", where he considers excess market returns as the only benchmark factor. Over the decades, more and more factors have been introduced in asset pricing model. The Fama-French (FF3, 1993) three-factor model is one of the most significant factor models, which includes the small minus big (SMB) size factor and high minus low (HML) value factor. As extensions of FF3, researchers further come up with other factor models such as Fama-French (FF5, 2015) five-factor model and the eight-factor model recently introduced by Skocir and Loncarski (2018).

Alpha is useful in obtaining an efficient portfolio as well. Sharpe ratio is a kind of tool to measure portfolio performance, which equals to the expected excess return divided by the standard deviation of a portfolio. A nonzero alpha means the Sharpe ratio can be improved by complementing investment in the benchmark portfolios with a position in the given asset. Thus the aim of asset pricing is equivalent to spanning an efficient portfolio.

Normally when comparing asset pricing models, researchers emphasize the competition in pricing test-asset returns. While in the paper of Barillas and Shanken (2017), they argue that models should be compared in terms of their ability to price all returns, both test assets and factors. In particular, the ability to price excluded factors plays the key role in model comparison, and test assets should be irrelevant. Besides, Barillas and Shanken (2017) also gives a likelihood metric, where models can be ranked based on their likelihoods.

In this paper, we replicate the work of Barillas and Shanken (2017). During the replication, we further confirm that excluded factor is a more reliable evidence for model comparison compared to test asset for non-nested models, while deviation arises in the replication of the illustration of likelihood, which is explainable since likelihood and alpha have different objectives in model comparison. However, it is still unclear whether the conclusion still holds for nested models, and whether these methods are still applicable with certain disturbance. Therefore, as an extension, we give further exploration for these questions. The extension consists of two parts. The first part focuses on the research question whether excluded factors are still a better evidence than test assets for nested model comparison, where we use the FF5 model which nests FF3 as an illustration. It turns out that excluded factors and test assets have almost equivalent performance. In the second part, a random walk process is generated based on the standard normal distribution, which serves as a nonsensical factor in order to answer the research question whether excluded factors and test assets still work for model comparison with a nonsensical factor. The result exposes the defect of excluded-factor evidence because in this case it favors the inefficient "FF3 with the nonsensical factor" against FF3. In contrast, test assets can still give the correct conclusion without being affected by the nonsensical factor, which is opposite with Barillas and Shanken (2017).

In terms of model comparison with nontraded factors, Barillas and Shanken (2017) do not dive into it. The difference with model comparison with traded factors is that test assets can be used to construct mimicking portfolios of nontraded factors. Since the weight of each test asset in mimicking portfolios are different, the conclusions of different test-asset regressions might also be diverse. Thus in the second extension, we further explore this problem based on the research question that how test assets impact model comparison with nontraded factors. To answer this question, an empirical illustration is conducted, where we still use FF3 as the basic model, and add nontraded factors to form a larger model to compare. Harvey, Liu and Zhu (2015) collects hundreds of risk factors and estimate their significance. Here we use the civilian unemployment rate and consumer price index (CPI), which are both nontraded and proved significant factors. Besides, we also include the impressive nontraded liquidity factor, introduced by Pastor and Stambaugh (2003). Even though we choose three nontraded factors, it turns out that only the liquidity factor can be captured by the test assets, thus it is the only nontraded factor which we finally adopt. With the mimicking portfolio as a substitute for the nontraded factor, we draw the conclusion that test-asset evidence is not reliable in this case since it is highly affected by which test asset we use.

Overall, this paper contributes for alpha-based model comparison, and helps financial institutions or managers to choose the proper method to compare models with different conditions, therefore optimize portfolios.

The paper is organized as follows. Section 2 introduces the structure and sources for the data applied in this paper. It follows with replication. Section 3 summarizes the basic idea of test-asset irrelevance in Barillas and Shanken (2017). Section 4 discusses how to perform alpha-based model comparison for nested and non-nested models. Section 5 introduces the likelihood metric and a corresponding illustration as well. Section 6 gives an example of non-nested model comparison provided in Barillas and Shanken (2017). The subsequent sections expand Barillas and Shanken (2017) with two extensions. Section 7 discusses nested model comparison and gives some examples as well. Section 8 incorporates nontraded factors and discusses how test assets make a difference in model comparison. Finally Section 9 gives a summary for this paper.

2 Data

In asset pricing models, most of the risk factors we use in this paper are introduced by Fama and Fench (1993, 2015). There are three traditional Fama-Fench factors: the excess return of the market portfolio (Mkt), small minus big (SMB) size factor and high minus low (HML) value factor. Besides, there are two newly introduced factors: the profitability factor (RMW) and the investment factor (CMA). The data of these factors, as well as the sorted portfolios which we use as test assets, are available in Kenneth French's website. The data of HML^{m}, which is another version of HML, is downloaded from the website of AQR Capital Management: "The

Devil in HML's Details" data set. For the nontraded factors, data of civilian unemployment rate and consumer price index (CPI) is obtained from the website of Federal Reserve Economic Data (FRED), and the data for the nontraded liquidity factor is provided by Lubos Pastor's Research. During the replication, we use the same sample period as Barillas and Shanken (2017), which is from July 1963 to December 2013. The data for the two extensions are relatively more recent, while still considering the limitation of data availability, which is from January 1968 to December 2017. All of the data stated above is monthly data.

3 Test-Asset Irrelevance under the Sharpe improvement metric

The main idea of Barillas and Shanken (2017) is that model comparison should be based on the extent to which each model is able to price the factors in the other models, indicated by the excluded-factor alphas, which should be irrelevant with test assets. To clarify this idea, they start with a theoretical evidence under Sharpe improvement metric to show that test assets are irrelevant in model comparison. The proof is as follows:

Assume there are two imperfect pricing models, M_1 with factors f_1 and M_2 with factors f_2. Now a set of test-asset returns R is provided to evaluate the models. Since they are imperfect models, adding additional information can attain improvement in the squared Sharpe ratio. In this way, if M_1 performs better, then the Sharpe improvement gained from adding the excluded factors and test assets to M_1 is smaller than that to M_2. The following relation holds,

$$Sh^2(f_1, f_2, R) - Sh^2(f_1) < Sh^2(f_2, f_1, R) - Sh^2(f_2), \tag{1}$$

where $Sh^2(\cdot)$ denotes the maximum available squared Sharpe ratio from the portfolios of the given returns. Since both sides of Equation (1) are the sum of Sharpe improvement gained from adding the excluded factors and test assets, the improvement from test assets can be canceled. Then we have the following relation,

$$Sh^2(f_1, f_2) - Sh^2(f_1) < Sh^2(f_2, f_1) - Sh^2(f_2). \tag{2}$$

This process implies the test-asset irrelevance in model comparison. Intuitively, test-asset pricing is a kind of absolute test, thus it cannot measure the relative performance of two models. However, it is still important to evaluate the "absolute" ability of the better model by pricing test assets.

4 Alpha-based model comparison

In this section, we replicate the work of Barillas and Shanken (2017) about how to perform model comparison with traded factors based on alpha. On the basic of Gibbons, Ross, and Shanken

(GRS, 1989) test, Barillas and Shanken (2017) give a relation between a quadratic form of alphas and improvement in squared Sharpe ratio, which is the center of the methodology. During the replication, we clarify the relation first, furthermore derive specific implications for nested and non-nested model comparison.

4.1 Relation between alpha and Sharpe ratio

Before giving the relation between model alpha and Sharpe ratio, it is important to make it clear that what asset pricing factor model is. It is a kind of factor models for the asset price. For example, capital asset pricing model (CAPM) can be viewed as a sing-factor model with excess market returns, and FF3 is a model with three factors to price asset returns. The essential idea of factor models is to describe the N asset returns R_{it} in terms of k common factors f_{it}, as

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + ... + \beta_{ik} f_{kt} + \varepsilon_{it}, \tag{3}$$

where R_{it} is the return on asset i in period t, f_{jt} is the j-th common factor and β_{ij} is the factor loading for asset i on factor j, with error terms ε_{it}. (i= 1, ..., N and j= 1, ..., k).

We can also write the returns $R_t = (R_{1t}, R_{2t}, ..., R_{Nt})'$ on N assets compactly as

$$R_t = \alpha + \beta f_t + \varepsilon_t \tag{4}$$

where $f_t = (f_{1t}, f_{2t}, ..., f_{kt})'$ and β is a $N \times k$ matrix with factor loadings. There are three assumptions for factor models: (1) the factor realizations f_t are stationary with constant mean and variance; (2) error terms ε_{it} have zero mean and invertible covariance matrix, and they are uncorrelated serially as well as across assets; (3) error terms ε_{it} are uncorrelated with each of the factors f_{it}.

Here we refer the assets R_t as test assets, and denote the alpha of test assets on factors in Equation (4) as α_R. Gibbons, Ross, and Shanken (1989) show that the improvement in the squared Sharpe ratio from adding test-asset returns to the universe of factors is equivalent to a quadratic form of the test-asset alphas:

$$\alpha_R' \Sigma^{-1} \alpha_R = Sh(f, R)^2 - Sh(f)^2, \tag{5}$$

where Σ is the covariance matrix of the error terms ε_{it}. With this relation, we can perform alpha-based model comparison in terms of Sharpe improvement.

4.2 Implications for nested model comparison

A model is nested in a larger model when its factors are all included in the larger model. In this way, M_1 with factors f_1 is nested in a larger model M with factors $f = \{f_1, f_2\}$. Since test assets are irrelevant in model comparison as proved in Section 3, we focus on excluded factors solely. Regressing f_2 on f_1, then we can obtain the excluded-factor regression as follows:

$$f_{2t} = \alpha_{21} + df_{1t} + u_t. \tag{6}$$

where α_{21} is the excluded-factor alpha, d denotes (matrix of) coefficients and u_t is the error term. As shown in Fama and French (2016), the idea behind the excluded-factor regression is that, if a factor can be captured by its exposures to the other factors in a model, then the factor plays no role in the explanation of the model. According to Equation (6), when the alphas are zeros for all the excluded factors, which means $\alpha_{21} = 0$, it holds that $Sh(f)^2 = Sh(f_1)^2$. With the equivalent Sharpe ratios of two models, we favor M_1 which has less factors according to the parsimony principle. Alternatively, if $\alpha_{21} \neq 0$, which means $Sh(f)^2 \neq Sh(f_1)^2$. It must hold that $Sh(f)^2 > Sh(f_1)^2$ because the left-hand side of Equation (5) is a quadratic form of alphas with positive definiteness. So that model M is a better model compared to M_1. Thus we can conclude that the nested model is the better model if and only if $\alpha_{21} = 0$.

4.3 Implications for non-nested model comparison

When we compare non-nested models, we can transform them into nested models and follow the same method above. More explicitly, suppose we need to compare models M_1 and M_2 with factors f_1 and f_2 respectively. M_1 and M_2 are non-nested in each other, which means there are excluded factors for both of the models. However, M_1 and M_2 can be seen as nested models of a larger model M with factors $f = f_1 \cup f_2$. Let f_1^* denotes the factors of f after removing factors of f_1 and f_2^* denotes the factors of f after removing factors of f_2. When we perform the excluded-factor regression of M_1, we regress f_1^* on f_1 and obtain excluded-factor alpha α_1^*. Similarly, the excluded-factor regression for M_2 is regressing f_2^* on f_2 and get α_2^*. However, instead of testing whether the alphas are equal to zeros as nested model comparison, we need to compare the magnitudes of α_1^* and α_2^*. A smaller alpha implies a better ability to price excluded factors, which indicates that the model is better than the other.

5 Likelihood-based model comparison

In the last section, we summarize how to perform alpha-based model comparison. In fact, in the paper of Barillas and Shanken (2017), they also provide a likelihood-based model comparison metric, which can prove test-asset irrelevance as well. In this section, we introduce the likelihood metric, and further replicate the corresponding illustration performed by Barillas and Shanken (2017).

To elaborate this metric, we require additional notations. Let f_{it}^* denotes the factors in f_t excluded from f_{it} at period t, and let α_i^* denotes the intercept in the regression of f_{it}^* on f_{it}, $i = 1, 2$. D_t represents all the data at period t, D_t=(f_t, R_t).

Now we try to draw the likelihood function for model M_i. Since the data universe D_t for each model i can be divided into three parts, f_t, f_t^* and R_t, we can use their marginal and conditional densities to rewrite the joint density. Thus the likelihood of model M_i consists of three parts:

(1) the marginal density of f_{it}. Assume normal distribution, it can be written as a function of two parameters μ_{f_i} and Σ_{f_i}, then the likelihood is represented by $L_i(\mu_{f_i}, \Sigma_{f_i}|f_{it})$;

(2) the conditional density of f_{it}^* given f_{it}. If we write the regression of f_{it}^* on f_{it} as follows,

$$f_{it}^* = \alpha_i^* + \beta_i^* f_{it} + \varepsilon_{it}^*, \tag{7}$$

with $Cov(\varepsilon_{it}^*|f_{it}) = \Sigma_i^*$, then the corresponding likelihood is $L_{i^*}(\alpha_i^*, \beta_i^*, \Sigma_i^*|f_t)$;

(3) the conditional density of R_t given f_t. Assume the regression of R_t on f_t is

$$R_t = \alpha_R + \beta_R f_t + \varepsilon_{Rt} \tag{8}$$

with $Cov(\varepsilon_{Rt}|f_t) = \Sigma_R$, the likelihood would be $L_R(\alpha_R, \beta_R, \Sigma_R|D_t)$.

The product of the three likelihoods gives an expression for the likelihood of M_i. But it is not enough since we still need to impose the restrictions that model M_i can price both the test-asset returns and excluded-factor returns ($\alpha_R = 0$ and $\alpha_i^* = 0$). With the restrictions, now it gives the likelihood of M_i:

$$L(M_i) = L_i(\mu_{f_i}, \Sigma_{f_i}|f_{it}) \times L_{i^*}(0, \beta_i^*, \Sigma_i^*|f_t) \times L_R(0, \beta_R, \Sigma_R|D_t). \tag{9}$$

5.1 An illustration of model comparison under the Likelihood metric

Now we replicate the model comparison based on likelihood as Barillas and Shanken (2017). The objective is to see whether we can succeed in replicating their result. If not, we need to find out the reasons which could be causing the differences.

The maximal likelihood $L^{max}(M_i)$ of each model M_i is obtained by maximizing the three components in Equation (9) respectively, and a higher likelihood value indicates a better model. However, a model with more parameters is likely to overfit even though it could has a higher likelihood value. To avoid such situation, we apply the well-known Akaike information criterion (AIC) as another instrument where the number of parameters m_i is incorporated:

$$AIC_i = -2ln[L^{max}(M_i)] + 2m_i. \tag{10}$$

AIC can serve as a measure of adjusted likelihood. Unlike the likelihood, a better model should have a lower value of AIC.

With these criterion, now we introduce the two models which Barillas and Shanken (2017) use to compare: FF3 model with factors Mkt, SMB and HML, and 4FM model with factors Mkt, SMB, HML^{m} and UMD. Most of the definitions of the factors are already given in Section

2. There are two common factors in the models. For 4FM, there is an additional factor UMD as well as HML^{m} which is a substitution of HML. The standard method to construct HML factor is calculating book-to-market using lagged data only. This method ignores the most recent price movements and lacks update. HML^{m} factor solves this problem. It is constructed with more timely data and using all the necessary lags to measure book instead of using lagged book data only, raised by Asness and Frazzini (2013). The up-minus-down factor (UMD) is a momentum factor, obtained by the average return on two high prior return portfolios minus the average return on two low prior return portfolios, as Jegadeesh and Titman (1993).

Both of FF3 and 4FM can be nested in the larger model M={Mkt SMB HML HML^{m} UMD}, then for FF3, f_1={Mkt SMB HML} and f_1^*={HML^{m} UMD}, and for 4FM, f_2={Mkt SMB HML^{m} UMD} and f_2^*={HML}. In this example, we assume all the densities multivariate normal, then for a multivariate linear regression $Y = \alpha + X\beta + U$, with covariance matrix Σ_U, the log likelihood of the data under joint normality and serial independence of U given X is given by:

$$ln(L(\alpha, \beta, \Sigma_U|D_t)) = -\frac{1}{2}[NTln(2\pi) + Tln|\Sigma_U| + trace(U'U\Sigma_U^{-1})], \quad (11)$$

where T is the number of time-series observations and N is the number of independent variables. The likelihood is maximized at the usual OLS estimates. In particular, when we calculate the likelihood of multivariate normal distribution, we use the difference between observations and means of factors for U, and Σ_U equals the maximum likelihood estimate of covariance matrix for factors. Besides, we take the twenty-five portfolios formed on size and momentum as the test asset. Then we have:

$$\begin{aligned} ln[L^{max}(FF3)] &= ln(L^{max}_{FF3}) + ln(L^{max}_{FF3^*}) + ln(L^{max}_{R}) \\ &= 3080.7 + 2733.1 + 26446.8 = 32260.6 \end{aligned}$$

$$\begin{aligned} ln[L^{max}(4FM)] &= ln(L^{max}_{4FM}) + ln(L^{max}_{4FM^*}) + ln(L^{max}_{R}) \\ &= 4212.0 + 1015.9 + 26446.8 = 31674.7 \end{aligned} \quad (12)$$

The likelihood of the test-asset portion is the same for both models, then it has no impact for the model comparison, which establishes the test-asset irrelevance again. To compute the AIC values, we need to find out the number of parameters of each model first. After imposing the zero-alpha constraints, for FF3 model, there are three factor means, six factor variances and covariances, six betas and three residual variances and covariances, thus there are eighteen parameters in total. For 4FM model, there is one more parameter compared to FF3 due to one fewer zero-alpha constraint. Thus the AIC of FF3 model is -2*32260.6+2*18=-64485.2, and for 4FM model is -2*31674.7+2*19=-63311.4.

Based on the result, FF3 model has a higher likelihood value and lower AIC, both of which imply that FF3 is a better model against 4FM. The conclusion is different with the original paper. To find out the reasons, we look at the three components of each likelihood in detail. Since the third component is always the same for the two models, it cannot affect the relative magnitude of likelihoods. There is a slight decrease in the relative value of the first component for FF3, but it is not significant, so the key lies in the second component. In the original paper, the ratio of the second component between FF3 and 4FM is 1.6, and now it turns to be 2.7, which means that the relative ability of FF3 to explain the excluded factors increases dramatically. It is largely due to the worse performance of 4FM to price its excluded factor HML, since the absolute value of FF3 does not change a lot. It could be caused by the update of data for HML or certain factors of 4FM, which affects the fitting of data. However, since the data Barillas and Shanken (2017) use is not available, this speculation is short of confirming.

6 Example of Non-nested model comparison

In this section, we replicate the example of non-nested model comparison provided in Barillas and Shanken (2017). As mentioned before, we aim to find out the difference during the replication and come up with some other insights. Based on FF3 and 4FM, excluded-factor regressions and test-asset regressions are performed respectively.

Table 1: Excluded factor regression for non-nested models

Panel A: Excluded-factor regressions for the FF3 model:

	Coefficients				
LHS	Alpha	Mkt	SMB	HML	R^2
HML^m	-0.66	0.06	0.03	0.97	0.61
	(0.09)	(0.02)	(0.03)	(0.03)	
UMD	10.87	-0.18	0.02	-0.33	0.06
	(0.17)	(0.04)	(0.06)	(0.06)	

Panel B: Excluded-factor regressions for the 4FM model:

	Coefficients					
LHS	Alpha	Mkt	SMB	HML^m	UMD	R^2
HML	-2.00	-0.02	-0.06	0.92	0.36	0.80
	(0.06)	(0.01)	(0.02)	(0.02)	(0.02)	

The table shows the OLS estimates of coefficients and R^2 for excluded regressions of two models. Panel A is for FF3 model, which regress the excluded factors HML^m and UMD on FF3={Mkt, SMB, HML} respectively. Panel B shows of result of regression of HML on 4FM={Mkt, SMB, HML^m, UMD}. The alphas are annualized and the numbers in parentheses are the corresponding standard errors of the coefficients.

The result of excluded-factor regressions is summarized in Table 1. Since what we use is monthly data, we simply multiply the alphas by twelve to get annualized alphas as reported in the table. For the excluded-factor regressions of FF3, the annualized alpha estimates are -0.66% and 10.87% for HML^m and UMD respectively. As for the excluded-factor regression for 4FM of HML, the alpha estimate is -2%. The alpha of UMD on FF3 factors is significantly larger than the others, which suggests terrible ability of FF3 to price the momentum factor. Since the alphas for included factors are always zero, we can calculate the average absolute alpha over the five factors of each model as follows: for FF3 it is (0.66+10.87+0+0+0)/5=2.31% and for 4FM the value is (2.00+0+0+0+0)/5=0.40%. Thus the excluded-factor evidence implies that 4FM is a better model against FF3. This result is consistent with the original paper. However, it is worth noting that in our replication, the result of excluded-factor evidence is inconsistent with that based on likelihood in the previous section. Even though it seems illogical, the truth is that model comparison based on alpha and likelihood are substantially different. Alpha means to measure the failure of pricing. More explicitly, it is the difference between the expected excess return formed by model factors and the actual return on the security. On the other hand, likelihood estimation is a kind of statistical tool used to find the parameters maximizing the likelihood function given observations. Thus when we use two methods with different objectives to compare models, it is no surprise that the results could be different.

The result of test-asset regressions is shown in Table 2. For each test-asset regression, we perform Gibbons, Ross, and Shanken (1989, GRS) test to see whether all the cross-section regression intercepts are equal to zero. The GRS test is widely used in asset pricing, but as shown in Harvey and Liu (2014), there are some problems with this test. The first problem is about its assumption of normal distribution of test-asset returns. It is a common problem for many statistical tests and the result will be inaccurate if the test-asset returns are not normally distributed. The second problem is that the GRS test almost always rejects, which leads to that when applying this test in asset pricing models, researchers focus on the relative magnitude of GRS statistics instead of whether the models are rejected or not. However, it is not what GRS test is designed to. To overcome these problems, Harvey and Liu (2014) create another method based on bootstrap approach. Their method allows for general types of distribution, cross-sectional dependency and time-series dependency, which is more robust and reliable than GRS test, making it a better alternative. However, since the application of this method is quite complex compared to GRS test, and it is not what this paper focuses on, we decide to still follow Barillas and Shanken (2017) and adopt GRS test in our replication.

Among the test-asset regressions, the regression of 16 size-B/M-OP-INV portfolios is missed because the data of this test asset is not accessible on Ken French's website. For most of the other test assets, the result is quite similar to the original paper. While for some test assets, the result is a bit different. For example, in the original paper, the annualized alphas of 25 size-B/M portfolios for FF3 and 4FM are 1.21 and 1.30 respectively, which means it favors FF3 over 4FM. But in our result, the alphas turn to be 1.27 and 1.26, which increases for FF3 and decreases for 4FM. So that the 25 size-B/M portfolios favors 4FM now. Since the result of Table 1 is quite close to that of Barillas and Shanken (2017), which indicates there is no big change in the data of the factors. Thus the difference in our replication might largely due to the changes in the data of test-asset returns. It further proves the defect of model comparison based on test assets.

Table 2: Test-asset regression for non-nested models

Model factors	A$\|a_i\|$	$A\|a_i\|$ /A$\|r_i\|$	GRS	p(GRS)
17 industries				
FF3	**2.21**	2.19	3.43	.0000
4FM	2.40	2.38	4.90	.0000
25 size-B/M				
FF3	1.27	0.56	3.64	.0000
4FM	**1.26**	0.56	4.12	.0000
25 size-UMD				
FF3	3.90	1.14	4.76	.0000
4FM	**1.43**	0.42	3.70	.0000
25 size-INV				
FF3	1.33	0.64	4.65	.0000
4FM	**0.93**	0.45	3.45	.0000
25 B/M-INV				
FF3	**1.32**	0.88	1.99	.0030
4FM	1.82	1.21	2.42	.0002
25 size-OP				
FF3	**1.31**	0.69	2.63	.0000
4FM	1.32	0.70	2.95	.0000
32 size-B/M-OP				
FF3	1.89	0.61	2.71	.0000
4FM	**1.88**	0.61	3.31	.0000
32 size-B/M-INV				
FF3	**1.57**	0.67	2.99	.0000
4FM	1.76	0.75	3.32	.0000
32 size-OP-INV				
FF3	2.21	0.80	4.58	.0000
4FM	**1.90**	0.68	3.68	.0000

This table shows the result of test-asset regressions with FF3 and 4FM. The first column of this table gives nine different kinds of portfolios as the test assets. They are formed on the following characteristics: size, book-to-market (B/M), momentum (UMD), investment (INV), operating profitability (OP), and industry; the second column shows the annualized average absolute value of alphas A$|a_i|$ and the smaller value between two models of each test asset is marked in bold; the third column is the annualized average absolute value of alphas over the average absolute value of return deviation of portfolio i, where A$|r_i|$ is the average absolute difference between the returns of portfolio i and the cross-sectional average return over all portfolios; and the fourth and fifth columns shows the Gibbons, Ross, and Shanken (1989) (GRS) F-statistic as well as the corresponding p-values for testing whether the alphas all equal to zero.

Because the result of test-asset regression is related to the return of test assets, which could lead the conclusion of model comparison inaccurate. Conversely, the excluded-factor regression is not affected by the change of test-asset returns, thus the conclusion based on that is more robust and dependable.

7 Extension 1: Example of Nested model comparison

In the previous example, we prove that the excluded-factor evidence is more reliable than test-asset evidence for non-nested models, but is it still the case for nested models? In this extension, we focus on model comparison of nested models and try to answer the research question whether excluded factor is still a better evidence than test asset.

First we need to determine two nested models. Fama and French (2015) develop the FF3 model by introducing two additional factors: profitability factor (RMW) which is the difference between the returns of diversified portfolios with robust and weak profitability, and investment factor (CMA) which is calculated by the difference between the returns on diversified portfolios of the stocks of low and high investment firms. In this way, a five-factor model (FF5) is formed. As in Fama and French (2015), this model has a problem of failure to fully capture the low average returns of small stocks whose returns perform like those firms that invest aggressively due to low profitability. However, as proved in Fama and French (2015) as well as Martins and Eid (2015), FF5 still has a better performance than FF3.

Table 3: Excluded factor regression for nested models

	Coefficients				
LHS	Alpha	Mkt	SMB	HML	R^2
RMW	4.07	-0.07	-0.25	0.02	0.16
	(0.09)	(0.02)	(0.03)	(0.03)	
CMA	2.67	-0.11	0.03	0.44	0.53
	(0.06)	(0.01)	(0.02)	(0.02)	

The table shows the OLS estimates of coefficients and R^2 for excluded regressions for FF3 model. There are two excluded factors for FF3: profitability factor (RMW) and investment factor (CMA). The alphas are annualized and the numbers in parentheses are the corresponding standard errors of the coefficients.

Similar with the non-nested example, we perform excluded-factor regressions as well as test-asset regressions here with the sample period from January 1968 to December 2017. Since they are nested models, there are only two excluded factors for FF3. The result of the excluded-factor regressions for FF3 is shown in Table 3. The annualized alphas for RMW and CMA are smaller than that of UMD in the previous example, which means the ability of FF3 to price these two factors is better than that to price UMD. The average absolute alpha over the five factors for FF3 is (4.07+2.67+0+0+0)/5=1.35%, and for FF5 is 0. Based on this result, FF5 model is favored over FF3.

For test-asset regressions, we choose portfolios formed based on size, book-to-market (B/M), investment (INV), operating profitability (OP) and industry as test assets. As shown in Table 4, nine of ten test assets prefer FF5 over FF3, and only 17-industry portfolios favors FF3. The result is almost consistent with the conclusion drawn by the excluded-factor evidence. In this case, the relative performance of test assets against excluded factors is different with comparison of non-nested models. Thus we can conclude that the test-asset evidence has a better performance for

Table 4: Test-asset regression for nested models

Model factors	A$\|a_i\|$	$A\|a_i\|$ /A$\|r_i\|$	GRS	p(GRS)
17 industries				
FF3	**2.09**	2.08	2.95	.0001
FF5	2.42	2.40	3.13	.0000
25 size-INV				
FF3	1.37	0.75	4.60	.0000
FF5	**0.93**	0.51	3.44	.0000
25 size-OP				
FF3	1.31	0.82	2.48	.0001
FF5	**0.76**	0.48	1.93	.0047
25 size-B/M				
FF3	1.20	0.66	3.87	.0000
FF5	**1.13**	0.62	3.23	.0000
25 OP-INV				
FF3	1.98	1.19	2.49	.0001
FF5	**1.00**	0.60	1.24	.1926
25 B/M-INV				
FF3	1.33	0.98	2.05	.0022
FF5	**1.32**	0.98	1.70	.0181
25 B/M-OP				
FF3	1.52	0.73	1.90	.0055
FF5	**1.36**	0.66	1.42	.0845
32 size-B/M-OP				
FF3	1.61	0.66	2.50	.0000
FF5	**1.31**	0.54	2.17	.0003
32 size-B/M-INV				
FF3	1.69	0.87	3.32	.0000
FF5	**1.28**	0.66	2.48	.0000
32 size-OP-INV				
FF3	2.21	0.92	4.48	.0000
FF5	**1.33**	0.55	3.40	.0000

This table shows the result of test-asset regressions for FF3 and FF5. The structure is the same with Table 2, thus the explanation is omitted.

nested models compared to non-nested models, and the difference between these two evidences is quite small.

We also apply the likelihood to further compare the two nested models. The difference with non-nested models is that, for the larger model FF5, the data universe D_t can only be divided into two parts, f_t and R_t, since there is no excluded factor for FF5. Thus the likelihood of FF5 is computed by the product of two components, L_{FF5}^{max} and L_R^{max}. The log-likelihoods are as follows:

$$\begin{aligned} ln[L^{max}(FF3)] &= ln(L_{FF3}^{max}) + ln(L_{FF3^*}^{max}) + ln(L_R^{max}) \\ &= 3063.4 + 2346.6 + 24299.2 = 29709.2 \end{aligned}$$

$$
\begin{aligned}
ln[L^{max}(FF5)] &= ln(L^{max}_{FF5}) + ln(L^{max}_{R}) \\
&= 4289.4 + 24299.2 = 28588.6
\end{aligned}
\tag{13}
$$

As shown above, the likelihood ratio of FF3 to FF5 is exp(5603/5), which is overwhelmingly in favor of FF3. This result is inconsistent with excluded factors neither test assets. However, since in this case the components of the two models are different. Thus we cannot say that it is completely fair" to compare the models based on their likelihoods.

7.1 Nested models with nonsensical factors

A nonsensical factor, also called "useless" factor, is defined as being independent of all the test-asset returns. In the paper of Kan and Zhang (1999), they argue that with large sample size, there is a large chance to reject the null hypothesis that the risk premium of a useless factor equals zero. Gospodinov, Kan and Robotti (2013) also prove that the presence of a useless factor makes it difficult to infer the remaining model factors and conclude the correct specification. They all point to a fact that the inclusion of a useless factor can interfere the judgment for model factors, thus it might also cause problems in model selection. In this section, we focus on the research question that what the effect of nonsensical factors on different metrics in model comparison. A random walk variable, denoted by RW, is generated as the nonsensical factor, and added in FF3. We define the new model as $FF3_{rw}$={Mrt SMB HML RW} which nests FF3. Then model comparison is performed based on excluded factors, test assets as well as the likelihood.

The random walk is generated by the following equation:

$$y_t = y_{t-1} + \varepsilon_t, \tag{14}$$

where y_t is computed by the lagged value plus a random error term ε_t, $\varepsilon_t \sim N(0, 1)$. The random walk with a drift is not applicable here because a drift term would imply a non-zero alpha in the excluded-factor regression by default. We treat the random walk as a time-series variable with the same sample size of other factors. The result of excluded-factor regression is as follows:

Table 5: Excluded-factor regression for nested models

	Coefficients				
LHS	Alpha	Mkt	SMB	HML	R^2
RW	-110.53	-0.05	0.04	-0.05	0.002
	(0.23)	(0.05)	(0.08)	(0.08)	

The table shows the OLS estimates of coefficients and R^2 for excluded-factor regression of FF3 model. The alpha is annualized and the numbers in parentheses are the corresponding standard errors of the coefficients.

The annualized alpha in the regression of random walk on FF3 factors is large enough with -110.53, and R^2 is pretty small. It is inevitable because the excluded factor is totally nonsensical and random, thus for sure it cannot be explained by any of these financial factors. The average absolute annualized alpha over the four factors is 27.63 for FF3 and 0 for $FF3_{rw}$, which is a strong indication that $FF3_{rw}$ is a much better model than FF3. But is it really the case? The only difference between the two models is the random walk factor, which is totally useless for explaining excess returns. Now the larger model turns to be a much better one just because the factors in the nested model cannot explain the nonsensical factor, which does not make sense.

On the other hand, test-asset regression does not have such problem. As shown in Table 6, all the test assets favor FF3 against $FF3_{rw}$, which draws an opposite conclusion with the excluded-factor evidence.

Table 6: Test-asset regression for nested models with nonsensical factor

Model factors	A$\|a_i\|$	$A\|a_i\|$ /A$\|r_i\|$	GRS	p(GRS)
17 industries				
FF3	**2.09**	2.08	2.95	.0001
$FF3_{rw}$	2.53	2.51	1.16	.2936
25 size-INV				
FF3	**1.37**	0.75	4.60	.0000
$FF3_{rw}$	2.36	1.29	2.37	.0002
25 size-OP				
FF3	**1.31**	0.82	2.48	.0001
$FF3_{rw}$	1.66	1.04	1.74	.0152
25 size-B/M				
FF3	**1.20**	0.66	3.87	.0000
$FF3_{rw}$	1.76	0.96	1.77	.0124
25 OP-INV				
FF3	**1.98**	1.19	2.49	.0001
$FF3_{rw}$	2.07	1.25	1.03	.4241
25 B/M-INV				
FF3	**1.33**	0.98	2.05	.0022
$FF3_{rw}$	2.07	1.53	1.35	.1189
25 B/M-OP				
FF3	**1.52**	0.73	1.90	.0055
$FF3_{rw}$	2.01	0.96	0.81	.7313
32 size-B/M-OP				
FF3	**1.61**	0.66	2.50	.0000
FF5	1.82	0.75	1.19	.2174
32 size-B/M-INV				
FF3	**1.69**	0.87	3.32	.0000
$FF3_{rw}$	2.64	1.36	2.28	.0001
32 size-OP-INV				
FF3	**2.21**	0.92	4.48	.0000
$FF3_{rw}$	2.26	0.94	2.01	.0010

This table shows the result of test-asset regressions for FF3 and FF3 with random walk. The structure is the same with Table 2, thus the explanation is omitted.

In fact, Equation (5) implies that for nested models, the squared Sharpe ratio of the larger model is always at least as high as that for the nested model, due to the positiveness of the quadratic form of alphas. However, the larger model is not always superior to the nested one. In this example, FF3 is supposed to be better than $FF3_{rw}$. To further confirm it, we look at the coefficients of random walk in test-asset regressions. Without surprise, for most of the test assets, the coefficients of random walk are insignificant, which implies the poor ability of random walk to price excess returns. Overall, the excluded-factor evidence does not give us a correct conclusion.

This finding exposes the drawback of model comparison based on excluded factors, especially for nested models. In our example, a nonsensical excluded factor produces a large excluded-alpha, which suggests that the larger model is better than the nested one. In practice, without knowing whether the excluded factors are useful or not, if we only focus on the excluded-factor test, we could choose a model including a useless factor just like the random walk, while abandon the more efficient and accurate model. To avoid such situation, I suggest to be careful when all the test-asset regressions imply an opposite conclusion with the excluded-factor evidence. In this case, it needs to be figured out whether the excluded factors are useful to price excess returns on earth.

Lastly we apply the likelihood metric. As mentioned before, due to the different components of likelihood functions, the likelihood metric might not be a good method to compare nested models. The result further proves our expectation. By Equation (15), the difference between the log likelihoods of FF3 and $FF3_{rw}$ is 951.5, which obviously favors FF3. Thus the likelihood metric again gives a wrong conclusion for nested model comparison.

$$\begin{aligned} ln[L^{max}(FF3)] &= ln(L^{max}_{FF3}) + ln(L^{max}_{FF3^*}) + ln(L^{max}_{R}) \\ &= 3063.4 + 2263.5 + 251933.8 = 257260.7 \end{aligned}$$

$$\begin{aligned} ln[L^{max}(FF3_{rw})] &= ln(L^{max}_{FF3_{rw}}) + ln(L^{max}_{R}) \\ &= 4375.4 + 251933.8 = 256309.2 \end{aligned} \tag{15}$$

8 Extension 2: Model comparison with nontraded factors

Until now we have discussed about model comparison with traded factors and how it is related with test assets. However, for models with nontraded factors, it might be a different story. For nontraded factors, we can use test assets to form mimicking portfolios, so that test assets can still impact asset pricing. In this section, we focus on model comparison with nontraded factors and try to answer the research question that how test assets affect model comparison with nontraded factors through mimicking portfolios.

8.1 Mimicking portfolio

As mentioned above, when we deal with nontraded factors, we use traded mimicking portfolios to replace them. There are several ways to form mimicking portfolios, such as projecting these factors on a set of base assets, where test assets are still irrelevant. Another method is to use test assets to construct mimicking portfolios. In this way, test assets are involved in model comparison. To get such mimicking portfolios, we need to regress the nontraded factors respectively on a constant and returns of test assets:

$$c_t = a + \sum_k b_k r_{kt} + \epsilon_t, \tag{16}$$

where c_t denotes a nontraded factor and r_{kt} denotes the return of the k-th test asset at time t. The mimicking portfolio is constructed by the weighted average over the individual test asset, and the weight of each test asset is equal to the absolute estimated coefficient of the test asset divided by the sum of absolute coefficients over all the test assets: $w_k = |b_k| / \sum_k |b_k|$. Then the return of the mimicking portfolio is $\sum_k w_k r_k$.

8.2 Example

To further analyze how test assets affect mode comparison with nontraded factors, an example is performed here. First we need to select some nontraded factors to add in FF3 and form a larger model. When choosing risk factors, the most commonly used statistical significance is a t-statistics that exceeds 2.0. However, Harvey, Liu and Zhu (2015) raise a objection to this method and suggest that a newly introduced risk factor should have a t-statistic exceeding 3.0. Civilian unemployment rate (UMR) and consumer price index (CPI) are two of the satisfactory nontraded factors. Intuitively, the conditions of labor market can reflect the state of investment market to some extent, and CPI, the indicator of inflation, can also imply the depreciation or appreciation of assets price. Besides, there is another factor that draws attention which is the nontraded liquidity factor created by Pastor and Stambaugh (2003), who argue that the sensitivities of stock returns to fluctuations in aggregate liquidity is important in asset pricing. There is also a traded counterpart for the liquidity factor which might be useful in mimicking the nontraded liquidity factor, but since we aim to find out how test assets impact model comparison by forming mimicking portfolios, including other factors other than test assets might be a disturbance, thus it is left out of consideration.

Table 7: Statistics for regressions of nontraded factors on test assets

	UMR	CPI	LIQ
R^2_{adj}	0.12	0.03	0.31
F-statistics	1.31	1.08	2.02
p-value	0.01	0.24	0.00

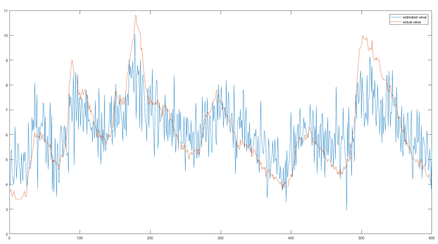
The table shows adjusted R^2 of the OLS regressions of three non-traded factors on test assets, as well as F-statistics of joint significance and the corresponding p-value.

During the construction of mimicking portfolios, it turns out that single test asset gives terrible fitting for the nontraded factors. Thus to fit them at the most extent, we include all the ten test assets in the regression of nontraded factors. However, in this way there are: 17+25*6+3*32 = 263 predictors in total, which might cause overfitting and produce misleadingly high value of R^2. To avoid such situation, we use adjusted R^2 instead of ordinary R^2. The adjusted R^2 is calculated by:

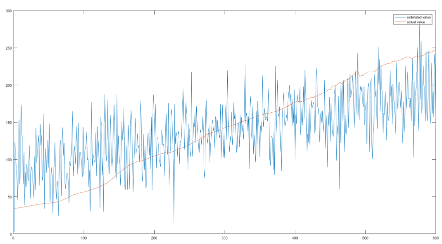
$$R^2_{adj} = 1 - \frac{(n-1)(1-R^2)}{(n-k-1)}, \tag{17}$$

where n is the number of observations and k is the number of explanatory variables excluding the constant. Table 7 implies how the test assets fit the nontraded factors. According to adjusted R^2 , the test assets are explanatory for LIQ and UMR, but not for CPI. Figure 1 further plots the actual values (red line) and estimated values (blue line) of the nontraded factors, which shows obvious overfitting for UMR and CPI. It means that even though the regression of UMR has a proper value of adjusted R^2 and does not fail the F-test, there still exists overfitting. Overall, the test assets are able to price LIQ, and fail to capture the other two factors. Thus we only include LIQ in FF3 and define the new model as FN={Mkt, SMB, HML, LIQ}. Table 8 gives the resulting weight of each test asset in the mimicking portfolio of LIQ. It can be seen that the three 25 size-sorted test assets together with the three 32 size-sorted test assets account for higher weights in the mimicking portfolio compared to the others.

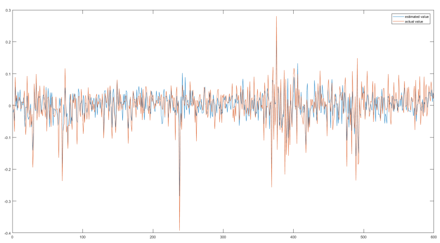
Figure 1: Fitting of test assets for nontraded factors



(a) UMR



(b) CPI



(c) LIQ

Substitute the nontraded factor with mimicking portfolio $\sum_k w_k r_k$, we now focus on how test assets influence model comparison. Since we aim to find out the difference of results drawn

Table 8: Weight of test asset in mimicking portfolio of LIQ

Test assets	Weights
17 industries	0.02
25 size-INV	0.15
25 size-OP	0.12
25 size-B/M	0.11
25 OP-INV	0.06
25 B/M-IN	0.09
25 B/M-OP	0.07
32 size-B/M-OP	0.13
32 size-B/M-INV	0.12
32 size-OP-INV	0.13

The table shows the weight of each test asset in the mimicking portfolio of LIQ. It is calculated by the sum of absolute coefficients of portfolios in one test asset divided by the sum of absolute coefficients of portfolios over all the test assets.

by different test assets, we do not perform excluded-factor regression neither the likelihood metric here. Table 9 summarizes the test-asset regressions of FF3 and FN. The alphas imply that FN model has a better performance again FF3 for seven of ten test assets. If we look at it in detail, we can find that most of these test assets take high weights in the mimicking portfolio. Intuitively, if the mimicking portfolio is determined largely by one test asset, then as a factor the mimicking portfolio can price this test asset to more extent. It is a disadvantage for model comparison based on test-asset regression since the result is highly determined by which test asset we use instead of which model is better. Thus it comes to a conclusion that for model comparison with nontraded factors and test assets are used to form mimicking portfolios, test-asset evidence is not reliable.

9 Conclusion

Barillas and Shanken (2017) argue that, test assets are irrelevant in model comparison and excluded-factor evidence is more reliable, thus model comparison can be performed based on the excluded-alpha. In this paper, the work of Barillas and Shanken (2017) is replicated, including the non-nested example of model comparison and the illustration of the likelihood metric. The replication of the former one is quite successful, while some deviation appears in the latter one. More explicitly, during the replication of model comparison between FF3 and 4FM based on likelihoods, we obtain the result that FF3 is better than 4FM, which is different with the original paper. Through analyzing we conclude that it is caused by the worse ability of 4FM to price the excluded factor HML, which mainly comes from the update of data. Nevertheless, there is a limitation that the data of the original paper is unavailable, leading to the factors with changed data unknown.

There are two problems Barillas and Shanken (2017) do not give further discussion: nested

Table 9: Test-asset regression for FF3 and FN

Model factors	A$\|a_i\|$	$A\|a_i\|$ /A$\|r_i\|$	GRS	p(GRS)
17 industries				
FF3	**2.09**	2.08	2.95	.0001
FN	2.26	2.51	3.01	.0000
25 size-INV				
FF3	1.37	0.75	4.60	.0000
FN	**1.26**	1.34	4.55	.0000
25 size-OP				
FF3	1.31	0.82	2.48	.0001
FN	**1.13**	0.98	2.44	.0001
25 size-B/M				
FF3	1.20	0.66	3.87	.0000
FN	**1.21**	0.98	3.83	.0000
25 OP-INV				
FF3	1.98	1.19	2.49	.0001
FN	**1.87**	0.97	2.47	.0001
25 B/M-INV				
FF3	**1.33**	0.98	2.05	.0022
FN	1.37	1.50	2.00	.0029
25 B/M-OP				
FF3	**1.52**	0.73	1.90	.0055
FN	1.56	0.96	2.18	.0009
32 size-B/M-OP				
FF3	1.61	0.66	2.50	.0000
FN	**1.58**	0.75	2.86	.0000
32 size-B/M-INV				
FF3	1.69	0.87	3.32	.0000
FN	**1.68**	1.39	3.27	.0000
32 size-OP-INV				
FF3	2.21	0.92	4.48	.0000
FN	**2.11**	0.94	4.43	.000

This table shows the result of test-asset regressions with FF3 and FN. The nontraded factor LIQ in FN is substituted by a traded mimicking portfolio constructed by the test assets. The structure is the same with Table 2, thus the explanation is omitted.

model comparison and model comparison with nontraded factors, based on which we further develop two extensions. In the first extension, we raise the research question that whether excluded factor is still a better evidence than test asset for nested models, and the answer is negative because they turn out to have almost equivalent performance. We also include a nonsensical factor to perform another nested model comparison based on the research question that what the effect of nonsensical factors on different metrics. The result exposes a significant defect of excluded-factor evidence since excluded factors always favor the larger model no matter what the excluded factors are. It can cause serious consequence during the application of excluded-factor evidence because in practice we do not know whether the additional factors are useful or not. In contrast, all the test assets give the correct conclusion. Therefore the conclusion is that test assets are still important in model comparison because in the case that all the test-asset regressions give an opposite result with excluded factors, it sends a signal that the additional factors can be useless.

The second extension is constructed on the research question that how test assets affect model comparison with nontraded factors through mimicking portfolios. It turns out that in test-asset regressions, traded factors with the mimicking portfolio of the nontraded factor show better performance to price those test assets which take higher weights in the mimicking portfolio. Thus we can say that in this case, test-asset evidence is not accurate because its result is highly determined by which test asset we use, and the excluded-factor evidence is recommended.

There are some limitations in this paper. First, as mentioned before, the data of the original paper is not accessible, making it hard to confirm the explanation of the deviation in our replication. Second, in model comparison with nontraded factors, test assets fail to capture two of three nontraded factors, leading to that only one nontraded factor is included in the model, which can make the conclusion non-representative. In the future, researchers can include more nontraded factors and try different methods of regression to avoid overfitting. In this way, one can obtain a more accurate conclusion.

References

[1] Barillas, F. and Shanken, J.(2017). Which alpha? *The Review of Financial Studies, Volume 30, Issue 4.*

[2] Kroenckez, T., Schindlerx, F., Sebastian, S. and Theissen, E.(2013). GDP Mimicking Portfolios and the Cross-Section of Stock Returns. *Discussion Paper No. 13-026.*

[3] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33:356.*

[4] Fama, E. F. and French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics Volume 116, Issue 1, April 2015, Pages 1-22.*

[5] Asness, C., and A. Frazzini (2013). The devil in HMLs details. *Journal of Portfolio Management 39:4968..*

[6] Barillas, F. and Shanken J.(2016). Comparing asset pricing models. *Working Paper, Emory University.*

[7] Barillas, F., Kan, R., Robotti, C. and Shanken, J.(2017). Model Comparison with Sharpe Ratios. *Rotman School of Management Working Paper No. 3013149.*

[8] Pastor, L. and Stambaugh, R. (2003). Liquidity risk and expected stock returns. *The Journal of Political Economy; Jun 2003; 111, 3.*

[9] Ødegaard, B. (2018). Factor Mimicking Portfolios.

[10] Harvey, C., Liu, Y. and Zhu, H. (2015). ... and the Cross-Section of Expected Returns. *The Review of Financial Studies, Volume 29, Issue 1, 1 January 2016, pages 5-68*

[11] Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics, Volume 68, Issue 1, April 2003, Pages 47-73*

[12] Harvey, C. and Liu, Y. (2014). Lucky factors.

[13] Gibbons, M., Ross, S. and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrics, Vol. 57, No. 5 (September, 1989), 1121-1152*

[14] Gospodinov, N., Kan, R. and Robotti, C (2013). Misspecification-robust inference in linear asset pricing models with irrelevant risk factors. *Working Paper, Federal Reserve Bank of Atlanta, No. 2013-9*

[15] Kan, R. and Zhang, C. (1999). Two-Pass Tests of Asset Pricing Models with Useless Factors. *THE JOURNAL OF FINANCE VOL. LIV, NO. 1 FEBRUARY 1999*

[16] Stambaugh, R. and Yuan, Y. (2017). Mispricing Factors. *The Review of Financial Studies, Volume 30, Issue 4, 1 April 2017, Pages 12701315*

[17] Fama, E. F. and French, K. R.(2016). Dissecting Anomalies with a Five-Factor Model. *The Review of Financial Studies, Volume 29, Issue 1, 1 January 2016, Pages 69103*

[18] Martins, C. and Eid, W. (2015). Pricing Assets with Fama and French 5-Factor Model: a Brazilian market novelty. *Conference: 15o Encontro Brasileiro de Finanas*

[19] Loncarski, I. and Skocir, M. (2018). Multi-factor asset pricing models: Factor construction choices and the revisit of pricing factors. *Journal of International Financial Markets, Institutions and Money, March 2018.*