AG428 Asset Pricing Coursework Examination

Lewis Britton

Academic Year 2020/2021

Table of Contents

1 May 2018 Paper

1.1 Question 1: Cross-Section

Critically examine the cross-sectional regression approach to evaluate linear factor models.

Background

Basic Ordinary Least Squares (OLS) regression analysis is used to test the integrity and accuracy of factor models. It generally relies on linear factor models following the standard of: $E(r_i) = \sum_{k=1}^{K} \beta_{ik}\lambda_k$ (or, as reflected in the formation of r_i: $R_i = R_f + \sum_{k=1}^{K} \beta_{ik}\lambda_k$). This was recognised by Shanken (1992) where: r_i represents the excess return on asset i ($R_i - R_f$); β_{ik} represents the *beta* of asset i relative to factor k (how i moves relative to k); λ_k represents the risk premium of factor k for K factors. This carries $\forall i\{1, ..., N\}$. As well as being used to test models such as the CAPM, ICAPM, and the APT; it was further brought to light by Fama and French (2015, 2016, 2017, 2018) across their analysis of various three-factor and five-factor models, such as the FF3 and FF5. This has been further extended more recently by other academics such as Hou, et al. (2020) with their contribution of models like variants of the HMXZ.

From here, methodologies extend to both time-series and cross-sectional analysis (Fama, 2015). The time-series approach is more commonly used to test linear factor models when the factors are portfolio returns such that $\lambda_k = E(r_{kt}) = E(r_m) - R_f$. This tests portfolio efficiency and assumes single or K factors, no folio contraints apart from the budget constraint, and the existence of a risk-free asset. Fama and MacBeth (1973) also utilise cross-sectional regression analysis to account for economic-focused factors, also. Current focus is on the cross-sectional approach.

Traditional OLS Assumptions

The Fama and MacBeth (1973) method generally makes use of methods which satisfy these requirements, such as exploitations of various aspects of estimated moments.

[1] $\Sigma\varepsilon_i \cong 0$

The sum of all residuals should approximately zero, meaning the degree to which observations outlie the line of best fit should be consistent. Errors should be minimised to satisfy $\min \left\{ \Sigma_{i=1}^{N} \varepsilon_i^2 = \Sigma_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2 \right\}$ for $\hat{\alpha} = \hat{\beta}_0;\ \hat{\beta} = \hat{\beta}_1$

[2] $\mathrm{var}(\varepsilon_i) = \mathrm{E}(\varepsilon_i^2) = \sigma^2$

All observations have constant errors, meaning that the variance of the residuals is constant. Therefore, they are homoskedastic; no need for a heteroskedasticity test (White, 1980) or the adjustment to a robust model.

[3] $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ (for $i \neq j$)

Error terms uncorrelated with one another; observations are exogenous (X_1 does not cause changes in X_2), not endogenous. There is no need for Instrumental Variable approach. No endogeneity problem.

Step 1

This cross-sectional method first makes use of time-series regression in its preliminary stage. To begin conducting this method, it must be insured that the user has collected a panel of asset returns for $\{r_{i,t}\}_{i=1,\,t=1}^{i=N,\,t=T}$, and a time-series of the market excess return for $\{E(r_t)\}_{t=1}^{t=T}$. This is when we run a time-series regression to view each asset's exposure to the risk factor, for $i \in \{1, ..., N\}$, in order to obtain the *beta* values ($\{\hat{\beta}_i\}_{i=1}^{N}$). Presented as follows:

$$
\begin{pmatrix} r_{i=1,t} \\ r_{i=2,t} \\ \vdots \\ r_{i=N,t} \end{pmatrix} = \begin{pmatrix} 1 & \hat{\beta}_{i=1,1} \\ 1 & \hat{\beta}_{i=2,1} \\ \vdots & \vdots \\ 1 & \hat{\beta}_{i=N,1} \end{pmatrix} \begin{pmatrix} \alpha_t \\ \lambda_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{i=1,t} \\ \varepsilon_{i=2,t} \\ \vdots \\ \varepsilon_{i=N,t} \end{pmatrix}
$$

Step 2

At this stage, treat all estimated *beta* values from the time-series regression as observed. Thus, you have observed *beta* values $\forall t\{1, ..., T\}$; represented as: $\{\hat{\beta}\}_{i=1}^{K}$. So forth, you now wish to compute multiple T cross-sectional regressions at each time stamp for $t \in \{1, ..., T\}$, $\forall i\{1, ..., N\}$ and their repective *beta* values. Presented as follows:

2

$$
\begin{pmatrix} r_{i=1,t} \\ r_{i=2,t} \\ \vdots \\ r_{i=N,t} \end{pmatrix} = \begin{pmatrix} 1 & \hat{\beta}_{i=1,1} \\ 1 & \hat{\beta}_{i=2,1} \\ \vdots & \vdots \\ 1 & \hat{\beta}_{i=N,1} \end{pmatrix} \begin{pmatrix} \alpha_t \\ \lambda_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{i=1,t} \\ \varepsilon_{i=2,t} \\ \vdots \\ \varepsilon_{i=N,t} \end{pmatrix}
$$

Hence, produced now are estiates existing at each time stamp for α and λ, as such: $[\hat{\alpha}_{t=1}, ..., \hat{\alpha}_T] \wedge [\hat{\lambda}_{t=1}, ..., \hat{\lambda}_T]$.

Step 3

We recall that Fama and MacBeth (1973) described the market exposure risk premium as $\lambda_k = E(r_{kt}) = E(r_m) - R_f$. Thus, using the estimates of the *lambda* values generated above, average risk premia for each time stamp is generated for such: $\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_t$. The same process occurs for erros. At this stage, the Fama and MacBeth (1973) method can also use the different market premium estimates at each time period for $t \in \{1, ..., T\}$ to generate standard errors. This follows:

$$
\sigma_{\hat{\lambda}} = \frac{\sigma_{\hat{\lambda}_t}}{\sqrt{T}}
$$

Step 4

At this stage, you wish to conduct a test to determine whether the $\hat{\lambda}$ are statistically different from zero. This is most commonly conducted through a t-test. Deduction of the t-ratio follows:

$$
t\text{-}ratio_{\hat{\lambda}} = \frac{\hat{\lambda}}{\left(\frac{\sigma_{\hat{\lambda}_t}}{\sqrt{T}} \right)}
$$

As expected in this t-test, the null hypothesis is that the premium is equal to zero (H$_0$: $\lambda = 0$). This is commonly rejected, using traditional t-ratio interpretation.

Notes of This Method

In the Fama and MacBeth (1973) approach, there is no requirement to know the absolute limit of T when estimating the risk premium. A large-enough cross-section will suffice. Such that, $\lambda \equiv E(r_t) = E(r_m) - R_f$ exploits the covariation of all i for $i \in \{1, ..., N\}$. Hence, $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} r_m - R_f$ is not relied upon for accuracy in $E(r_m) - R_f$.

1.2 Question 2: Individual Securities

Discuss the benefits of using individual securities to test linear factor models using both time-series and cross-sectional regression approaches.

Piss off.

1.3   Question 3: Comparison Tests

Critically examine the alternative model comparison tests of Fama and French (2015, 2016). Barillas and Shanken (2017, a, b), and Brarillas, Kan, Robotti and Shanken (2017).

Background

When examining and comparing linear factor models such as Fama and French's three-factor model, five-factor model, Hou, et al.'s five-factor model etc., there are various metrics under which components of the models are tested. Generally, the mean-variance efficiency of models tends to be rejected. History shows that no model really presents zero pricing error and thus, instead of look for the model with absolute zero pricing error, it is important to identify the model which shows componenets which minimise pricing error and alike contributories. Model testing usually starts with the Gibbons, Ross and Shanken (1989) Test in which models can be compared on a basic level relative to the null hypothesis of zero pricing error across observations. More expansive model comparisons then extend to pricing error metrics explored by Fama and French (2015, 2016, 2017, 2018), Sharpe performance comparison tests brought ot light by Barillas, Kan, Robotti and Shanken (2020), and factor redundancy tests explored by Fama (2015).

Gibbons, Ross and Shanken Test

Gibbons, Ross and Shanken (1989) developed the Gibbons, Ross and Shanken Test (GRS Test); a method which tests the null hypothesis of zero pricing error (α_i) in asset i for all $i \in \{1, ..., N\}$); "all of the *alpha* values are jointly equal to zero". Thus, this concept relates to the joint distribution of the intercepts of a model being equal to zero. This is a multivariate test for mean-variance efficiency of factor models. It assumes the regression residuals (ε_{it}), when testing for pricing error, have a multivariate normal distribution with zero-mean and a constant covariance matrix. Knowing that $N_{\text{Assets}} < (T_{\text{Observations}} - K_{\text{Factors}})$, the test proceeds. The GRS Test essentially acts as an applied F-test in which all pricing error values (α_i) in a model are equal to zero. Upon observation of the *p-value* produced in the GRS Test relevant to H_0: $\alpha = 0$, we tend to reject the mean-variance efficiency of models as most do present significant pricing error.

Computation of the GRS Test makes use of basic matrix algebraic functions such as

addition/subtraction $(+/-)$, multiplication, transposition $(')$, and inversion $(^{-1})$. We recall the implied models from the Fama and French (2015, 2016, 2017, 2018) time-series factor model analysis:

$$E(r_i) = \alpha_i + \beta_{ik}\lambda_{ik}$$

$$\therefore E(r_i) = \alpha_i + \beta_{ik}E(r_{kt})$$

After computed the relevant estimates, the GRS Test applies the discussed notes to a vector of all the *alpha* values in a given model, as: $\hat{\alpha} = [\hat{\alpha}_1, ..., \hat{\alpha}_N]$. Hence, "are all the *alpha* values jointly equal to zero?". In search of the test-statistic for this methodology, Gibbons, Ross and Shanken (1989) began with the Wald Test:

$$T\left(\frac{1}{(1 + Sh_k^2)}\right)\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

$$\Rightarrow [\hat{\Sigma}^{-1} = \text{cov}(\varepsilon, \varepsilon')^{-1}]$$

We know from Sharpe (1966) that Sh_k^2 is the maximum squared sharpe performance of factor k in the model. And furthermore, that $\hat{\alpha}'$ is the transpose of the discussed vector of all *alpha* values in the model, and $\hat{\alpha}$ is the [N,1] vector of individual *alpha* values. $\hat{\Sigma}^{-1}$ is the *Maximum Liklihood* (ML) estimate of the residual covariance matrix for N regressions. The Maximum squared sharpe performance is given by:

$$Sh_k^2 = \bar{\alpha}_k' V_k^{-1} \bar{\alpha}_k$$

$$\Rightarrow [V_k^{-1} = \text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}]$$

We know from here that $\bar{\alpha}$ is the [K,1] matrix of mean excess returns on K factors, $\bar{\alpha}'$ is the transpose of this, V_k^{-1} is the inverse of ML estimate of the [K,K] covariance matrix of the mean excess returns on K factors. The Wald Test concludes upon an asyptotic $(T \longrightarrow \infty)$ *Chi-Squared* (χ^2) distribution with N degrees of freedom.

From here, Gibbons, et al. (1989) deduce their applied F-test that there are N cases of no pricing errors, under residual multivariate normality (IID normal). This is based on an F-distribution, and N and T-N-K degrees of freedom and is therefore relevant in smaller samples. Intuitively, the Null hypothesis (H_0) is that there is mean-variance efficiency reflected in all the *alpha* values being jointly equal to zero thus, that $\lambda = 0$. The GRS Test is represented as:

$$\left(\frac{T-N-K}{N}\right)\left(\frac{1}{(1+Sh_k^2)}\right)\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

As a final note in relation to portfolio efficiency, if $K = 1$ then the GRS Test examines whether the market portfolio lies on the efficient frontier of the $(N+1)$ assets. However, if $K > 1$ then the GRS Test examines whether a combination of the K factors which lies on the efficient frontier of the $(N + K)$ assets.

Pricing Error Metrics

Its recognised that when usign the GRS Test for N test assets to compare models, $\text{cov}(\hat{\alpha}, \hat{\alpha}')$ varies across these models. Therefore, the GRS Test can generate inaccurately low results due to sampling error. There are eight pricing error metrics explored and practiced by Fama and French (2012, 2015, 2016, 2018) which focus on the actual pricing error (α) values in the models.

The first metric is the absoulte average *alpha* value ($\bar{\alpha}$) which accounts purely for the magnitude of the *alpha* values. That is, absolute average mispricing. It removes any bias which may occur in models which have *alpha* values which are skewed in one direction. The model with the lower absolute average *alpha* value is the better one as lower (closer to 0) mispricing is identified.

The next metric is the average *t-ratio* of the *alpha* values ($\overline{\text{t-ratio}_\alpha}$). Following *t-stat* protocol, this is as always an indicator of the significance of the mispricing. Therefore, as high significance indicates a significantly great mispricing error, the model with a *t-ratio* reflecting low significance is better. In terms of *t-stat* interpretation, this generally requires a lower *t-ratio*.

The third and fourth metrics are, much like *t-ratio*, based on traditional OLS output. They include the average Adjusted R^2 value of the models ($\overline{\text{Adj. } R^2}$) in which a higher value is sought which can aid explanation of lower standard error values. Furthermore, the fourth metric is the average standard error of the *alpha* values ($\overline{\text{std.err}_\alpha}$). Intuitively, a lower value is saught here as it's implied such model would contain mispricing values (α) closer to 0. That is, lower mispricing.

The fifth metric is a ratio which makes use of squared *alpha* values and the volatility of excess returns on the N testing assets. It's represented as: $\frac{\overline{\alpha^2}}{\sigma_r^2}$. This ratio is a fairly

intuitive ratio which primarily acounts for the real spread in mispricing to volatility. Of course, a lower value is desired. An extention of this metric is the sixth metric; another ratio. It is similar to the former however, the numerator accounts also for the standard error of the *alpha* values. It is represented as: $\frac{\overline{\lambda^2}}{\sigma_r^2}$ (where $\lambda^2 = \alpha^2 - \text{std.err}_\alpha^2$), $\therefore \frac{\overline{\lambda^2}}{\sigma_r^2} = \frac{\overline{\alpha^2 - \text{std.err}_\alpha^2}}{\sigma_r^2}$. It is known as the average real mispricing as it is said to account for and remove sampling error. Although not 100% mathematically sound or correctly in-line with various mathematical theory, this method better accounts for vector distance, or 'spread' in the case of 'finance', in relation to mispricing and excess return volatility. Of course again, a lower value is sought.

The seventh metric another based on basic output; the average squared standard error value of *alpha* values over average squared *alpha*: $\frac{\overline{\text{std.err}_\alpha^2}}{\alpha^2}$. It is said to specifically focus upon and capture the proportion of the mispricing across the N testing assets which is attributed to sampling error. Thus, as you wish the 'average real mispricing' metric [6] to present a lower value than the basic mispricing metric [5], you are anticipating larger sampling error in better results. Therefore, the model with a greater value in this metric would be one generally with a better (lower) 'average real mispricing' (greater attribution of mispricing to sampling error, in the case of mispricing). Thus, higher value is better.

The final, eighth, metric is simply the average squared Sharpe performance of the optimal portfolio. Therefore, denoted by: $\overline{\text{Sharpe}_\alpha^2}$. Of course, model with the higher value is the better one.

Sharpe Performance Measures

Furthermore, Barillas and Shanken (2017) conduct further research regarding the integrity of model testing on N assets across time-series models. They make the argument that the choice of testing assets in comparison is irrelevant. This is based on the principal that comparing models using the GRS metric highlights the fact that increases in Sharpe performance approaching $Sh_{k_A, k_B, r}^2$ sees a better-off model A when it's lower than model B. Considering the fixed-nature of $Sh_{k_A, k_B, r}^2$ (and that under the GRS metric, Model A: $(Sh_{k_A, k_B, r}^2 - Sh_{k_A}^2)$ and Model B: $(Sh_{k_A, k_B, r}^2 - Sh_{k_B}^2)$), there would be an indifference in the former and comparison simply upon $Sh_{k_A}^2$ and $Sh_{k_B}^2$. Thus, irrelevant choice of test assets.

The aim is to compare models on a nested or non-nested basis. That is, models which are endogenous/not endogenous of each-other. The methodology is based on the idea that "if you include all the factors in the investment universe, the test assets become irrelevant" . For example, [you are comparing the FF3 and FF5] the Fama and French FF3 is endogenous of the Fama and French FF5 as the FF5 is an expanded version of the FF3 which holds many of the same characteristics. Contrastingly, [you are comparing the FF5 and the HMXZ5] the Hou, et al. HMXZ5 is not endogenous of the FF5. Therefore, the FF3 vs. FF5 is a 'nested' test; the FF5 vs. HMXZ5 is a 'non-nested' test.

Barillas and Shanken (2017) note that the issue to be addressed in model comparisons is "how well a factor model actually prices factors which are not included in the model". They add that comparing models on this basis often presents different results to traditional tests of mispricing. They proceed with what they call 'Proposition 1'; a test of: $Sh^2_{k_A} = Sh^2_{k_B}$. Take, for example, the nested comparison of the FF3 and FF5; how well does the smaller (nested) model (FF3) price the factors not included from the FF5. Thus, let k_A denote factors in the smaller model for L factors (FF3 factors), and let k_B denote factors in the larger model for K factors ($L + J$ (L factors of model A and J additional factors (intuitievly: $J = K - L$))).

Notation simplification:

- Smaller Model (k_A): L Factors
- Larger Model (k_B): $K = L + J$ Factors
- Non-Endogenous Factors: $J = K - L$

So forth, the following test is conducted:

$$k_{jt} = \alpha_j + \sum_{l=1}^{L} \beta_{jl} k_{lt} + \varepsilon_{jt}$$

Note that k_{jt} denotes excess returns on the J factors which are excluded (additional factors in the larger model), and k_{lt} denotes excess returns on the L factors actually accounted for by the smaller model. The test of this model operates on the basis of the null hypothesis (H$_0$) of $\alpha_j = 0 \ \forall j\{1, .., J\}$. Thus, stating that both models produce equivalent excess return predictions and hold the same squared Sharpe performance. If this hypothesis was failed-to-be-rejected, J factors are redundant and the smaller model is preferred. If the hypothesis is rejected however, the larger model is preferred,

9

with a greater squared Sharpe performance.

Furthermore, Barillas, Kan, Robotti and Shanken (2020) extend reearch in this area to formulate the test of maximum squared Sharpe perfromance. The nested testing uses the GRS methodology however, Proposition 1 also allows analysis of the metric 'difference in squared Sharpe performance'. This of course, requires that the Sharpe performance of at least one of the models is greater than zero. In some cases, non-nested models with minor overlapping factors may still support $Sh^2_{k_A} = Sh^2_{k_B}$ if the overlapping factors hold the same Sharpe performance of combined factors of both of the models. Barillas, et al. (2020) further expand upon this with a test of equality of Sharpe performance, which makes use of a 'benchmark' model. It operates on the null hypothesis that the benchmark has an equivalent Sharpe performance as the other models in question. The alternative hypothesis states that at least one of the other models has a *higher* Sharpe performance than the benchmark.

From here, testing can be observed more-closely on a nested/non-nested basis. Using a benchmark which is nested in the other models allows use of a close equivalent of the GRS methodology to test for aggregate zero-*alpha*-value. This uses excluded factors (of the other model with the widest scope) from the benchmark model to test for Sharpe equivalence. An example given for this uses comparison of the CAPM to the Fama and French FF3 and FF5. This would involve the GRS test of H_0: $\alpha = 0$ cross-overs SMB, HML, RMW and CMA fators. Upon rejection of H_0, it's implied that the benchmark has a significantly lesser Sharpe performance than the other models in question. Additionally observing non-nested models, the null hypothesis that the benchmark presents an equivalent or greater Sharpe performance to those of the other, non-nested, models is drawn. Barillas, et al. (2019, 2020) make use of various methods of Wolak (1987) in this context. This denotes $(Sh^2_{kl} - Sh^2_{ki})$ as δ_i for $i \in \{2, ..., p\}$, in the case pf p models. Thus, H_0: $\delta \geq 0$. This of course accounts for benchmark $p-1 = k$.

The output for relevant tests consists of the adjusted squared Sharpe measures of each model in question (and associated *p-values*), and the difference in adjusted squared Sharpe performance (and associated *p-values*) across the models. Ferson and Siegel (2003) highlight a 'bias-adjusted' Sharpe performance as this 'adjusted squared Sharpe performance', which accounts for bias in Sharpe performance where N is great relative to T:

$$\text{Adj. Sharpe}^2 = \text{Sharpe}^2 \left(\frac{(T - K - 2)}{T} \right) - \frac{K}{T}$$

Included also is the associated *p-value* in relation to the above which is based Barillas, et al.'s (2020) 'Proposition 1'.

Factor Redundancy Tests

Fama and French (2015) exercised a method of 'factor spanning' as they referred to it. This methodology asks the question of "can *beta* values of other factors in the model explain the mean excess return of a factor?"; "are all of the factors required?". That is, The role that a factor may play could already be captured by others; coefficients of other factors may already reflect what a factor contributes. Therefore, it may have approximately no effect on excess returns of the model during the sample period; mean excess return of the model would be approximately equal without the factor. To answer this question, factor redundancy regressions are run for each factor portfolio's excess returns. The model is as follows:

$$r_{pt} = \alpha_p + \sum_{k=1}^{K-1} \beta_{pk} r_{kt} + \varepsilon_{pt}$$

As we'd expect, r_{pt} denotes the excess returns of the factor portfolio at time t. p is introduced to denote a single factor portfolio representing one $k \in \{1, ..., K\}$ hence, when left with the explanatory side of the new model: $k \in \{1, ..., (K-1)\}$. The model is based on the null hypothesis that a factor adds ≈ 0 to excess returns (H$_0$: $\alpha_p = 0$). Thus, we wish to reject this null hypothesis; saying a factor is not redundant. If a factor is redundant, it's worth omitting.

Furthermore, Fama and French (2018) the question of marginal contribution. That is, not just if a factor is redundant or worth including but, what its marginal contribution to the model actually is. Numerically, this contribution is given by the marginal contribution to the maximum squares Sharpe performance. It's represented as: $\frac{\alpha^2}{\sigma_\varepsilon^2}$, which accounts for the *alpha* values and the volatility of residuals. There remains the question of what drives a factor's contribution; a high *alpha* value ($\uparrow \alpha$) or a low volatility of residuals ($\downarrow \sigma_\varepsilon$)? It's good to observe fairly distributed marginal contribution; individual components of the metri must also be observed to determine why a factor may have an exceptionally high/low marginal contribution however. Obviously, this method can be used to highlight heavily dominant factors despite its naming-after the opposite.

1.4 Question 4: Test Improvement

Lewellen, Nagel, Shanken (2010) provide a critique of asset pricing tests of linear factor models and make a number of recommendations to improve these tests. Discuss the critique and the recommendations provided by Lewllen, et al. (2010).

Piss off.

2 December 2018 Paper

2.1 Question 1: Models & Comparison Tests

Discuss the main linear factor models and explain the Gibbons, Ross and Shanken (1989) test of mean-variance efficiency.

Background

Many traditional models include the Capital Asset Pricing Model (CAPM), the International Capital Asset Pricing Model (ICAPM), and the Arbitrage Pricing Theory model (APT). The great John H. Cochrane of Standford University (the only 'business' academic I've come across who actually has the care produce notes in LaTeX (even uses LaTeX syntax on his website, although he uses a little too much JavaScript for my liking)) highlights, as part of Cochrane (2005), that linear factor models "act as a better proxy for aggregate utility growth, than consumption data". Suggesting that the financial metric in question better-reflect states of the economy and how to forecast the 'least bad' outcomes.

Capital Asset Pricing Model

The CAPM, explored by Sharpe (1964), is described as using aggregate wealth to represent utility growth in the economy. This is reflected in the *market portfolio*. It makes use of the efficient frontier [of portfolios] where the entire collection of investors are asummed to select portfolios which lie on this *mean-variance efficient frontier*. A risk-free asset is assumed to exist also. It's said that in the market-equilibrium, investors hold an 'optimal' combination of risk-free and risky assets within the market folio which is on the efficient frontier. Therefore as a result, the Security Market Line (SML) states that there is a linear relationship between expected return and the market *beta* value, as: $E(R_i) = R_f + \beta_i(E(R_m) - R_f)$, where $(E(R_m) - R_f) =$ Market Risk Premium and if $\beta = 1$ (the closer β is to 1), the market folio is being tracked (the closer the market folio is being tracked).

International Capital Asset Pricing Model

Additionally, the ICAPM was explored by Merton (1973), accounting for variables which may influence investment opportunity. Primarily, to avoid poor investment movements in the future which could have an effect on current utility. That is, investors may

hedge undesirable movements in investment opportunities. This makes any relevant variable which can be used to predict these changes today, an additional factor used in contribution to aggregate wealth. Fama (1996) describes this idea of the ICAPM as 'multifactor minimum-variance efficiency', as opposed to mean-variance-oriented. The time-series issue is the concern with consumption and invetment opportunity; for each one of these captured by a 'state' variable, a 'mimicking' portfolio is produced $(r_{s(t+1)})$. This is in aid of selecting an optimal portfolio to $\max\{E(r_p)\}$ and $\min\{\sigma_p^2\}$, with a target β_{is} for all state variables $s \in \{1, ..., S\}$. this method produces a result on the 'multifactor minimum variance frontier', with optimal hedged S portfolios. Realistically, this ICAPM method is used heavily by investors in times of precaution. For example, one s hedged folio may be reflected in a recession risk factor so, the investor may give up some current $E(r_p)$ in order to significantly minimise possible risk in the future, and current risk, specifically in the case of a recession. Similar to the CAPM, investors optimally hold a combination of the risk-free asset, market folio and S folios.

Arbitrage Pricing Theory Model

The APT, explored by Ross (1976), is purposed to aid the pricing of assets across a smaller number of commonly occuring factors. The model helps compare unique identifiers of stocks. That is, the APT aims to avoid pricing *idiosynchratic* components of returns as they can be 'diversified away' in the use of a greater-stock (larger) portfolio. Choice of factors is based upon selecting ones which best explain the covariance matrix of stock returns. Hence, common occuring factors in stocks are what should be priced. The model utilised $r_i = E(r_i) + \sum_{k=1}^{K} \beta_{ik} f_k + \varepsilon_i$ states that if there are no idiosynchratic factors, then $\varepsilon_i = 0$. Thus, any asset return contained within is accurately represented by the folio of K factors. In result, a linear relationship between expected return and the *beta* value: $E(R_i) = R_f + \sum_{k=1}^{K} \beta_i \lambda_k$. However, when $\varepsilon \neq 0$ asset returns are not reflected simply by a factor folio thus, an approximate APT model is produced. Connor (1984) (best year ever) highlights the need for an 'equilibrium APT' which accounts exact relationships between expect return and factor *beta* values. Predictions of the APT are similar to those, of the CAPM and variant. A combination of K-factor folios lie on the mean-varince frontier; there is a linear relationship between expected retrun and factor *beta* values; stock characteristics hold no relationship with expected returns after adjustment for factor *beta* values.

Fama & French Stock Characteristic Models

Furthermore, more recent factor models were brought to light by Fama and French (2015, 2016, 2017, 2018) in anticipation of analysing more specific characteristics of the nature and type of assets/portfolios. The Fama & French Three-Factor Model (FF3) and Fama & French Five-Factor Model (FF5) are good examples. As of Fama and French (2015), we know these multi-factor models account for stock characteristics such as *market excess return* $(R_m - R_f)$; *size* (small-minus-big (SMB)) which is 'average return on small-firm portfolios minus average return on big-firm portfolios'; *book-to-market* (high[value]-minus-low[growth] (HML)) which is 'average return on value-firm portfolios minus average return on growth-firm portfolios'; *operating profitability* (robust-minus-weak (RMW)) which is 'average return on high-profit-firm portfolios minus average return on low-profit-firm portfolios'; and *investment prospect* (conservative-minus-aggressive (CMA)) which is 'average return on conservatively-investing-firm portfolios minus aggressively-investing-firm portfolios'. Fama and French (2015) find that small, profitable and valuable firms present the highest returns. The FF5, on which this is based, follows:

$$(R_{it} - R_{ft}) = \alpha_i + \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t$$

$$+ \beta_4 RMW_t + \beta_5 CMA_t + \varepsilon_{it}$$

More recently, Hou, et al. (2020) presented an alternative five-factor model, the HMXZ5. It is descirbed as containing the 'q-factors', based on 'q-theory' and appearning in previous research/versions of their models, and an additional expected growth factor. It accounts for market excess return, size, profitability (return on equity, as opposed to annual profitability as in the FF models) investment and expected growth. The Fama and French models have been found the generally outperform the Hou, et al. models within relevant comparative metrics. Additionally, Stambaugh and Yuan (2017) propose a four-factor model based on market excess return, size and two mispricing factors. Daniel, Hirshleifer and Sun (2020) propose a three-factor model including market excess return, a long-term financing anomaly factor, and a short-term post-earnings announcement factor. Neither of these are quite as commonly-practiced as the former two.

Foundational Basis of Factors

Ferson (2003) highlights various different methods and sets of rationale in the creation of factors in multifactor models. There are three general methods including first, statistical formation based on basic statistical analysis which aims to explain covariance matrices of stock returns. Although this method may be statistically accurate in the relevance of some of the major factor models, it lacks rationale and meaning. The second method, brought to light by Chen, Roll, Ross (1986), tends to fill some of the blank spaces in the first, it is based on economic factors with real meaning. The third method relates specifically to factors from stock characteristics. This formation of factors is the one primarily carried to factors models from Fama and French and Hou, et al., discussed previously.

Gibbons, Ross and Shanken Test

Gibbons, Ross and Shanken (1989) developed the Gibbons, Ross and Shanken Test (GRS Test); a method which tests the null hypothesis of zero pricing error (α_i) in asset i for all $i \in \{1, ..., N\}$); "all of the *alpha* values are jointly equal to zero". Thus, this concept relates to the joint distribution of the intercepts of a model being equal to zero. This is a multivariate test for mean-variance efficiency of factor models. It assumes the regression residuals (ε_{it}), when testing for pricing error, have a multivariate normal distribution with zero-mean and a constant covariance matrix. Knowing that $N_{\text{Assets}} < (T_{\text{Observations}} - K_{\text{Factors}})$, the test proceeds. The GRS Test essentially acts as an applied F-test in which all pricing error values (α_i) in a model are equal to zero. Upon observation of the *p-value* produced in the GRS Test relevant to H_0: $\alpha = 0$, we tend to reject the mean-variance efficiency of models as most do present significant pricing error.

Computation of the GRS Test makes use of basic matrix algebraic functions such as addition/subtraction ($+/-$), multiplication, transposition ($'$), and inversion ($^{-1}$). We recall the implied models from the Fama and French (2015, 2016, 2017, 2018) time-series factor model analysis:

$$E(r_i) = \alpha_i + \beta_{ik}\lambda_{ik}$$

$$\therefore E(r_i) = \alpha_i + \beta_{ik}E(r_{kt})$$

After computed the relevant estimates, the GRS Test applies the discussed notes to a vector of all the *alpha* values in a given model, as: $\hat{\alpha} = [\hat{\alpha}_1, ..., \hat{\alpha}_N]$. Hence, "are all the

alpha values jointly equal to zero?". In search of the test-statistic for this methodology, Gibbons, Ross and Shanken (1989) began with the Wald Test:

$$T \left(\frac{1}{(1 + Sh_k^2)} \right) \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}$$

$$\Rightarrow [\hat{\Sigma}^{-1} = \mathrm{cov}(\varepsilon, \varepsilon')^{-1}]$$

We know from Sharpe (1966) that Sh_k^2 is the maximum squared sharpe performance of factor k in the model. And furthermore, that $\hat{\alpha}'$ is the transpose of the discussed vector of all *alpha* values in the model, and $\hat{\alpha}$ is the [N,1] vector of individual *alpha* values. $\hat{\Sigma}^{-1}$ is the *Maximum Liklihood* (ML) estimate of the residual covariance matrix for N regressions. The Maximum squared sharpe performance is given by:

$$Sh_k^2 = \bar{\alpha}_k' V_k^{-1} \bar{\alpha}_k$$

$$\Rightarrow [V_k^{-1} = \mathrm{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}]$$

We know from here that $\bar{\alpha}$ is the [K,1] matrix of mean excess returns on K factors, $\bar{\alpha}'$ is the transpose of this, V_k^{-1} is the inverse of ML estimate of the [K,K] covariance matrix of the mean excess returns on K factors. The Wald Test concludes upon an asyptotic ($T \longrightarrow \infty$) *Chi-Squared* (χ^2) distribution with N degrees of freedom.

From here, Gibbons, et al. (1989) deduce their applied F-test that there are N cases of no pricing errors, under residual multivariate normality (IID normal). This is based on an F-distribution, and N and T-N-K degrees of freedom and is therefore relevant in smaller samples. Intuitively, the Null hypothesis (H_0) is that there is mean-variance efficiency reflected in all the *alpha* values being jointly equal to zero thus, that $\lambda = 0$. The GRS Test is represented as:

$$\left(\frac{T - N - K}{N} \right) \left(\frac{1}{(1 + Sh_k^2)} \right) \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}$$

As a final note in relation to portfolio efficiency, if $K = 1$ then the GRS Test examines whether the market portfolio lies on the efficient frontier of the $(N+1)$ assets. However, if $K > 1$ then the GRS Test examines whether a combination of the K factors which lies on the efficient frontier of the $(N + K)$ assets.

2.2 Question 2: Comparison Tests

Critically evaluate the model comparison tests that can be used to evaluate linear factor models.

Background

When examining and comparing linear factor models such as Fama and French's three-factor model, five-factor model, Hou, et al.'s five-factor model etc., there are various metrics under which components of the models are tested. Generally, the mean-variance efficiency of models tends to be rejected. History shows that no model really presents zero pricing error and thus, instead of look for the model with absolute zero pricing error, it is important to identify the model which shows componenets which minimise pricing error and alike contributories. Model testing usually starts with the Gibbons, Ross and Shanken (1989) Test in which models can be compared on a basic level relative to the null hypothesis of zero pricing error across observations. More expansive model comparisons then extend to pricing error metrics explored by Fama and French (2015, 2016, 2017, 2018), Sharpe performance comparison tests brought ot light by Barillas, Kan, Robotti and Shanken (2020), and factor redundancy tests explored by Fama (2015).

Gibbons, Ross and Shanken Test

Gibbons, Ross and Shanken (1989) developed the Gibbons, Ross and Shanken Test (GRS Test); a method which tests the null hypothesis of zero pricing error (α_i) in asset i for all $i \in \{1, ..., N\}$); "all of the *alpha* values are jointly equal to zero". Thus, this concept relates to the joint distribution of the intercepts of a model being equal to zero. This is a multivariate test for mean-variance efficiency of factor models. It assumes the regression residuals (ε_{it}), when testing for pricing error, have a multivariate normal distribution with zero-mean and a constant covariance matrix. Knowing that $N_{\text{Assets}} < (T_{\text{Observations}} - K_{\text{Factors}})$, the test proceeds. The GRS Test essentially acts as an applied F-test in which all pricing error values (α_i) in a model are equal to zero. Upon observation of the *p-value* produced in the GRS Test relevant to H$_0$: $\alpha = 0$, we tend to reject the mean-variance efficiency of models as most do present significant pricing error.

Computation of the GRS Test makes use of basic matrix algebraic functions such as addition/subtraction ($+/-$), multiplication, transposition ($'$), and inversion ($^{-1}$). We

recall the implied models from the Fama and French (2015, 2016, 2017, 2018) time-series factor model analysis:

$$E(r_i) = \alpha_i + \beta_{ik}\lambda_{ik}$$

$$\therefore E(r_i) = \alpha_i + \beta_{ik}E(r_{kt})$$

After computed the relevant estimates, the GRS Test applies the discussed notes to a vector of all the *alpha* values in a given model, as: $\hat{\alpha} = [\hat{\alpha}_1, ..., \hat{\alpha}_N]$. Hence, "are all the *alpha* values jointly equal to zero?". In search of the test-statistic for this methodology, Gibbons, Ross and Shanken (1989) began with the Wald Test:

$$T\left(\frac{1}{(1 + Sh_k^2)}\right)\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

$$\Rightarrow [\hat{\Sigma}^{-1} = \text{cov}(\varepsilon, \varepsilon')^{-1}]$$

We know from Sharpe (1966) that Sh_k^2 is the maximum squared sharpe performance of factor k in the model. And furthermore, that $\hat{\alpha}'$ is the transpose of the discussed vector of all *alpha* values in the model, and $\hat{\alpha}$ is the [N,1] vector of individual *alpha* values. $\hat{\Sigma}^{-1}$ is the *Maximum Liklihood* (ML) estimate of the residual covariance matrix for N regressions. The Maximum squared sharpe performance is given by:

$$Sh_k^2 = \bar{\alpha}_k'V_k^{-1}\bar{\alpha}_k$$

$$\Rightarrow [V_k^{-1} = \text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}]$$

We know from here that $\bar{\alpha}$ is the [K,1] matrix of mean excess returns on K factors, $\bar{\alpha}'$ is the transpose of this, V_k^{-1} is the inverse of ML estimate of the [K,K] covariance matrix of the mean excess returns on K factors. The Wald Test concludes upon an asyptotic $(T \longrightarrow \infty)$ *Chi-Squared* (χ^2) distribution with N degrees of freedom.

From here, Gibbons, et al. (1989) deduce their applied F-test that there are N cases of no pricing errors, under residual multivariate normality (IID normal). This is based on an F-distribution, and N and T-N-K degrees of freedom and is therefore relevant in smaller samples. Intuitively, the Null hypothesis (H_0) is that there is mean-variance efficiency reflected in all the *alpha* values being jointly equal to zero thus, that $\lambda = 0$. The GRS Test is represented as:

$$\left(\frac{T-N-K}{N}\right)\left(\frac{1}{(1+Sh_k^2)}\right)\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

As a final note in relation to portfolio efficiency, if $K = 1$ then the GRS Test examines whether the market portfolio lies on the efficient frontier of the $(N+1)$ assets. However, if $K > 1$ then the GRS Test examines whether a combination of the K factors which lies on the efficient frontier of the $(N + K)$ assets.

Pricing Error Metrics

Its recognised that when usign the GRS Test for N test assets to compare models, $\text{cov}(\hat{\alpha}, \hat{\alpha}')$ varies across these models. Therefore, the GRS Test can generate inaccurately low results due to sampling error. There are eight pricing error metrics explored and practiced by Fama and French (2012, 2015, 2016, 2018) which focus on the actual pricing error (α) values in the models.

The first metric is the absoulte average *alpha* value ($\bar{\alpha}$) which accounts purely for the magnitude of the *alpha* values. That is, absolute average mispricing. It removes any bias which may occur in models which have *alpha* values which are skewed in one direction. The model with the lower absolute average *alpha* value is the better one as lower (closer to 0) mispricing is identified.

The next metric is the average *t-ratio* of the *alpha* values ($\overline{\text{t-ratio}_\alpha}$). Following *t-stat* protocol, this is as always an indicator of the significance of the mispricing. Therefore, as high significance indicates a significantly great mispricing error, the model with a *t-ratio* reflecting low significance is better. In terms of *t-stat* interpretation, this generally requires a lower *t-ratio*.

The third and fourth metrics are, much like *t-ratio*, based on traditional OLS output. They include the average Adjusted R^2 value of the models ($\overline{\text{Adj. } R^2}$) in which a higher value is sought which can aid explanation of lower standard error values. Furthermore, the fourth metric is the average standard error of the *alpha* values ($\overline{\text{std.err}_\alpha}$). Intuitively, a lower value is saught here as it's implied such model would contain mispricing values (α) closer to 0. That is, lower mispricing.

The fifth metric is a ratio which makes use of squared *alpha* values and the volatility of excess returns on the N testing assets. It's represented as: $\frac{\overline{\alpha^2}}{\sigma_r^2}$. This ratio is a fairly

intuitive ratio which primarily acounts for the real spread in mispricing to volatility. Of course, a lower value is desired. An extention of this metric is the sixth metric; another ratio. It is similar to the former however, the numerator accounts also for the standard error of the *alpha* values. It is represented as: $\frac{\overline{\lambda^2}}{\sigma_r^2}$ (where $\lambda^2 = \alpha^2 - \text{std.err}_\alpha^2$), $\therefore \frac{\overline{\lambda^2}}{\sigma_r^2} = \frac{\overline{\alpha^2 - \text{std.err}_\alpha^2}}{\sigma_r^2}$. It is known as the average real mispricing as it is said to account for and remove sampling error. Although not 100% mathematically sound or correctly in-line with various mathematical theory, this method better accounts for vector distance, or 'spread' in the case of 'finance', in relation to mispricing and excess return volatility. Of course again, a lower value is sought.

The seventh metric another based on basic output; the average squared standard error value of *alpha* values over average squared *alpha*: $\frac{\overline{\text{std.err}_\alpha^2}}{\alpha^2}$. It is said to specifically focus upon and capture the proportion of the mispricing across the N testing assets which is attributed to sampling error. Thus, as you wish the 'average real mispricing' metric [6] to present a lower value than the basic mispricing metric [5], you are anticipating larger sampling error in better results. Therefore, the model with a greater value in this metric would be one generally with a better (lower) 'average real mispricing' (greater attribution of mispricing to sampling error, in the case of mispricing). Thus, higher value is better.

The final, eighth, metric is simply the average squared Sharpe performance of the optimal portfolio. Therefore, denoted by: $\overline{\text{Sharpe}_\alpha^2}$. Of course, model with the higher value is the better one.

Sharpe Performance Measures

Furthermore, Barillas and Shanken (2017) conduct further research regarding the integrity of model testing on N assets across time-series models. They make the argument that the choice of testing assets in comparison is irrelevant. This is based on the principal that comparing models using the GRS metric highlights the fact that increases in Sharpe performance approaching $Sh_{k_A,k_B,r}^2$ sees a better-off model A when it's lower than model B. Considering the fixed-nature of $Sh_{k_A,k_B,r}^2$ (and that under the GRS metric, Model A: $(Sh_{k_A,k_B,r}^2 - Sh_{k_A}^2)$ and Model B: $(Sh_{k_A,k_B,r}^2 - Sh_{k_B}^2)$), there would be an indifference in the former and comparison simply upon $Sh_{k_A}^2$ and $Sh_{k_B}^2$. Thus, irrelevant choice of test assets.

The aim is to compare models on a nested or non-nested basis. That is, models which are endogenous/not endogenous of each-other. The methodology is based on the idea that "if you include all the factors in the investment universe, the test assets become irrelevant" . For example, [you are comparing the FF3 and FF5] the Fama and French FF3 is endogenous of the Fama and French FF5 as the FF5 is an expanded version of the FF3 which holds many of the same characteristics. Contrastingly, [you are comparing the FF5 and the HMXZ5] the Hou, et al. HMXZ5 is not endogenous of the FF5. Therefore, the FF3 vs. FF5 is a 'nested' test; the FF5 vs. HMXZ5 is a 'non-nested' test.

Barillas and Shanken (2017) note that the issue to be addressed in model comparisons is "how well a factor model actually prices factors which are not included in the model". They add that comparing models on this basis often presents different results to traditional tests of mispricing. They proceed with what they call 'Proposition 1'; a test of: $Sh^2_{k_A} = Sh^2_{k_B}$. Take, for example, the nested comparison of the FF3 and FF5; how well does the smaller (nested) model (FF3) price the factors not included from the FF5. Thus, let k_A denote factors in the smaller model for L factors (FF3 factors), and let k_B denote factors in the larger model for K factors ($L + J$ (L factors of model A and J additional factors (intuitievly: $J = K - L$))).

Notation simplification:

- Smaller Model (k_A): L Factors
- Larger Model (k_B): $K = L + J$ Factors
- Non-Endogenous Factors: $J = K - L$

So forth, the following test is conducted:

$$k_{jt} = \alpha_j + \sum_{l=1}^{L} \beta_{jl} k_{lt} + \varepsilon_{jt}$$

Note that k_{jt} denotes excess returns on the J factors which are excluded (additional factors in the larger model), and k_{lt} denotes excess returns on the L factors actually accounted for by the smaller model. The test of this model operates on the basis of the null hypothesis (H_0) of $\alpha_j = 0 \ \forall j\{1,..,J\}$. Thus, stating that both models produce equivalent excess return predictions and hold the same squared Sharpe performance. If this hypothesis was failed-to-be-rejected, J factors are redundant and the smaller model is preferred. If the hypothesis is rejected however, the larger model is preferred,

with a greater squared Sharpe performance.

Furthermore, Barillas, Kan, Robotti and Shanken (2020) extend reearch in this area to formulate the test of maximum squared Sharpe perfromance. The nested testing uses the GRS methodology however, Proposition 1 also allows analysis of the metric 'difference in squared Sharpe performance'. This of course, requires that the Sharpe performance of at least one of the models is greater than zero. In some cases, non-nested models with minor overlapping factors may still support $Sh^2_{k_A} = Sh^2_{k_B}$ if the overlapping factors hold the same Sharpe performance of combined factors of both of the models. Barillas, et al. (2020) further expand upon this with a test of equality of Sharpe performance, which makes use of a 'benchmark' model. It operates on the null hypothesis that the benchmark has an equivalent Sharpe performance as the other models in question. The alternative hypothesis states that at least one of the other models has a *higher* Sharpe performance than the benchmark.

From here, testing can be observed more-closely on a nested/non-nested basis. Using a benchmark which is nested in the other models allows use of a close equivalent of the GRS methodology to test for aggregate zero-*alpha*-value. This uses excluded factors (of the other model with the widest scope) from the benchmark model to test for Sharpe equivalence. An example given for this uses comparison of the CAPM to the Fama and French FF3 and FF5. This would involve the GRS test of H_0: $\alpha = 0$ cross-overs SMB, HML, RMW and CMA fators. Upon rejection of H_0, it's implied that the benchmark has a significantly lesser Sharpe performance than the other models in question. Additionally observing non-nested models, the null hypothesis that the benchmark presents an equivalent or greater Sharpe performance to those of the other, non-nested, models is drawn. Barillas, et al. (2019, 2020) make use of various methods of Wolak (1987) in this context. This denotes $(Sh^2_{kl} - Sh^2_{ki})$ as δ_i for $i \in \{2, ..., p\}$, in the case pf p models. Thus, H_0: $\delta \geq 0$. This of course accounts for benchmark $p-1 = k$.

The output for relevant tests consists of the adjusted squared Sharpe measures of each model in question (and associated *p-values*), and the difference in adjusted squared Sharpe performance (and associated *p-values*) across the models. Ferson and Siegel (2003) highlight a 'bias-adjusted' Sharpe performance as this 'adjusted squared Sharpe performance', which accounts for bias in Sharpe performance where N is great relative to T:

$$\text{Adj. Sharpe}^2 = \text{Sharpe}^2 \left(\frac{(T - K - 2)}{T} \right) - \frac{K}{T}$$

Included also is the associated *p-value* in relation to the above which is based Barillas, et al.'s (2020) 'Proposition 1'.

Factor Redundancy Tests

Fama and French (2015) exercised a method of 'factor spanning' as they referred to it. This methodology asks the question of "can *beta* values of other factors in the model explain the mean excess return of a factor?"; "are all of the factors required?". That is, The role that a factor may play could already be captured by others; coefficients of other factors may already reflect what a factor contributes. Therefore, it may have approximately no effect on excess returns of the model during the sample period; mean excess return of the model would be approximately equal without the factor. To answer this question, factor redundancy regressions are run for each factor portfolio's excess returns. The model is as follows:

$$r_{pt} = \alpha_p + \sum_{k=1}^{K-1} \beta_{pk} r_{kt} + \varepsilon_{pt}$$

As we'd expect, r_{pt} denotes the excess returns of the factor portfolio at time t. p is introduced to denote a single factor portfolio representing one $k \in \{1, ..., K\}$ hence, when left with the explanatory side of the new model: $k \in \{1, ..., (K-1)\}$. The model is based on the null hypothesis that a factor adds ≈ 0 to excess returns (H$_0$: $\alpha_p = 0$). Thus, we wish to reject this null hypothesis; saying a factor is not redundant. If a factor is redundant, it's worth omitting.

Furthermore, Fama and French (2018) the question of marginal contribution. That is, not just if a factor is redundant or worth including but, what its marginal contribution to the model actually is. Numerically, this contribution is given by the marginal contribution to the maximum squares Sharpe performance. It's represented as: $\frac{\alpha^2}{\sigma_\varepsilon^2}$, which accounts for the *alpha* values and the volatility of residuals. There remains the question of what drives a factor's contribution; a high *alpha* value ($\uparrow \alpha$) or a low volatility of residuals ($\downarrow \sigma_\varepsilon$)? It's good to observe fairly distributed marginal contribution; individual components of the metri must also be observed to determine why a factor may have an exceptionally high/low marginal contribution however. Obviously, this method can be used to highlight heavily dominant factors despite its naming-after the opposite.

2.3   Question 3: Market Frictions

Discuss the impact of market frictions on tests of portfolio efficiency.

Piss off.

2.4   Question 4: Conditioning

Critically examine the impact of conditioning information in tests of portfolio efficiency.

Background

As the great John H. Chochrane says, "we do not live in a world where each day's returns are based on a coin flip". Essentially, variables vary over time; as the name should suggest. This idea of 'conditioning' is based on the idea of more complex models of expected returns, *beta* values, risk, etc., being 'conditional' upon 'information sets' which may not have been previously considered. These information sets such, as varying covariance [matrices], can be produced by time-varying perceptions and attitudes of individuals, markets, and analysts etc. and their ability to produce results/predictions based on the access to information they have. And so on. In other words, methodologies surrounding this have a lot to do with subjectivity of internals and some externals. An alternative is that there are 'unconditional' expectations.

Prior to consideration of conditions, mean-variance effiecny of a factor model has been tested based adjacent to a fixed-weight efficient frontier. This contradicts the idea of conditioning as it relies upon time-constant expected return, covariance and risk etc. Based on the discussed, we define 'conditioning information' as lagged information variables which hold some form of predictive nature of future results, based upon mostly subjective componenets. These lagged information variables are generally used to further aid prediction of expected returns, covariance and risk etc., also known as 'conditional moments'. This allows investors to expand their trading strategies into a more dynamic and organic nature.

Expected Return Framework

Generally, we derive the expected discounted payoff framework, with regrds to the above, considering primrily the 'agent' information set. That is, the information set based on the perceptions and attiudes of individuals in the market. We refer to their subjective expectations about the future. The theory of *the law of iterated expectation* states that:

$$E(E(x)_c) = E(x)$$

The expectation of a conditional expectation is equivalent to the expectation. That

is: "what you think today about what you are going to think tomorrow about the day after tomorrow is equivalent to what you think today about tomorrow" ($E(E(x_{t+2})) = E(x_{t+1}) \ \forall t\{t \to \infty\}$). However, we cannot accurately quantify simply what is 'in people's heads'. So forth, proxies for conditioning information relative to the agent stem from time-series lagged dividend yield on the market index, short-term Treasury 'T-' Bill rate, term/interest spread(s) and default spread. These give accurate indciators of agent behaviour and thus, predictability. Following the theoretical framework above, these can be modelled to satisfy mean-variance efficiency with conditioning information. We use the term 'conditional mean variance' (CMV) frontier to denote the equivalent to the efficient frontier upon which a portfolio lies if it is efficient to the conditional information. In the research of Hansen and Richard (1987), this must satisfy:

$$\min\{E(R_{pt}^2)|Z_{t-1}\}$$

subject to

$$E(R_{pt}|Z_{t-1}) = \text{Target Level} \wedge x'_{Z_{t-1}}1_N = 1$$

In other words, a portfolio lies on the CMV frontier if it satisfies minimization of $E(R_{pt}^2)$ such that a lagged information variable exists and is present, and is subject to $E(R_{pt}|Z_{t-1})$ being present at the target level and optimal weights (x_{Z_t}). Furthermore, Hansen and Richard (1987) also denote the equivalent of the mean-variance frontier (in the presence of conditioning), without conditioning information, as the 'unconditional mean-variance' (UMV) frontier. It displays the set of optimal weights (x_{Z_t}) to satisfy:

$$\min\{E(R_{pt}^2)\}$$

subject to

$$E(R_{pt}) = \text{Target Level} \wedge x'_{Z_{t-1}}1_N = 1$$

In other words, a portfolio lies on the UMV frontier if it satisfies minimization of $E(R_{pt}^2)$ which is subject to $E(R_{pt})$ being present at the target level and optimal weights x_{Z_t}. Hansen and Richard (1987) also highlight, as extensively discussed over various appearances from John H. Chochrane, that some portfolios which lie on the UMV frontier also lie on the CMV frontier however, other portfolios which lie on the CMV frontier do not generally lie on the UMV frontier. Finally, for reference, we recall the most 'naïve', if you will, version of the mean-variance efficient frontier; what we now denote ' 'passive mean-variace' (PMV) frontier' as. It is simply subject to optimal

$E(R_{pt})$ and weights. As follows:

$$\min\{E(R_{pt}^2)\}$$

$$subject\ to$$

$$E(R_{pt}) = \text{Target Level} \wedge x'1_N = 1$$

In effect, the CMV forntier optimizes the trade-off between expected payoff and conditional risk based on information and other attributes, in a time-series format. On the othe rhand, the UMV frontier optimizes the trade-off between expected payoff and unconditional risk based upon non-informational attributes, in a time-series format. The UMV frontier is essentially providing a similar framework to the PMV frontier however, it allows for the consideration of attributes which contribute to more dynamic trading strategies.

Using Information

As we noted, generally we like to consider what the agent is thinking. Ferson and Siegel (2001) explore the case that where clients have less information than portfolio managers, the cleints desire a portfolio which lies on the UMV frontier as opposed to the CMV frontier due to various unaccountable and untransferrable informational and behavioural attributes and characteristics in the case of the porfolio manager's use of the CMV frontier. The use of a CMV frontier portfolio in this case can result in poorer performance compared to the UMV equivalent. Ferson and Siegel (2001) argue that the UMV portfolios can be more resillient to misspecification of conditional moments as attributes of the UMV portfolios are more valid than CMV ones in the case of misspecification. Neither are optimal however.

A slightly different approach to incorporating conditional information without extensive modelling of conditional moments is called the 'multiplicative approach'. The method is in the title. You assume L lagged information variables and multiply asset returns by each according lagged information variable, individually. This results in a larger-scoped pool for an optimal portfolio to be selected from.

Deriving Frontier Componenets

Ferson and Siegel (2009) expand upon their previous work, beginning with the statement that all pricing models, models with economic factors, and models with informa-

tion conditioning see a portfolio on the UMV frontier. They first derive the optimal portfolio weights, in the presence of a fixed risk-free asset set of return of (R_f). This follows:

$$x_Z = \left(\frac{u_s - R_f}{B}\right) Q(u_Z - R_f)$$

Where :

$$Q = \left((u_Z - R_f)(u_Z - R_f)' + cov_Z\right)^{-1}$$

$$B = E\left((u_Z - R_f)'Q(u_Z - R_f)\right)$$

x_Z obviously denotes the optimal weight, u_s in the target expected return, R_f is the risk-free return, u_Z denotes conditional expected returns, and cov_Z is the conditional covariance matrix. First, to estimate conditional moments u_Z and cov_Z, Ferson and Siegel (2009) present the formation of the time-series regression:

$$R_{it} = \alpha_i + \sum_{l=1}^{L} \beta_{il} Z_{l(t-1)} + \varepsilon_{it}$$

This highlights L lagged information variables $Z_{l(t-1)}$ of lag l for $l \in \{1, ..., L\}$. From here, fitted values are interprted as conditional expected returns (u_Z) and the covariance matrix (cov_Z) is calculated from produced components. It can be noted that Ferson and Siegel (2009) also consider a covariance matrix which varies over time however, they find little change in results when accounting for this.

Testing The Model

Ferson and Siegel (2009) extend their research into testing their methodology using a form of the traditional Wald test in a similar form to the Gibbons, Ross and Shanken (1989) test (GRS Test), to test mean-variance efficiency; whether or not a portfolio lies on the UMV frontier. As always, this accounts for K risky assets in the benchmark model and N testing assets. Recall the form of the Wald test:

$$T\left(\frac{S_r^2 - S_{r_p}^2}{1 + S_{r_p}^2}\right)$$

As we know, S_r^2 is the squared Sharpe performance introduced by Sharpe (1966) for the entirity of the $(N + K)$ assets. With the inclusion of optimal conditioning information, this is the squared Sharpe performance of the UMV portfolio. $S_{r_p}^2$ is the squared Sharpe

performance of the K assets in the benchmark. Hence, mean variance efficiency exists for the benchmark model if $S_r^2 = S_{r_p}^2$. Using a statistical probability simulation method, known as the *Monte Carlo Simulation*, which generates hypothetical values in the case of extreme uncertainty. This is present under the null hypothesis of tested folios being present upon the UMV frontier. Ferson and Siegel (2009) find that in the presence of conditioning information, the UMV frontier holds a higher Sharpe performance over a longer period of time. They also reject mean-variance efficiency of the CAPM and the FF of Fama and French (1993), both in the presence of conditional information and not.

Issues

A primary issue highlighted by *the legends themselves* Basu and Fletcher (2016) is that strategies related to UMV optimization require high turnover so frequently when adjusting for trading costs, the UMV approach may require significantly greater outperformance. It does not always hold this characeristic. They ask the question of whether the benefit of the predictability using discussed methods outweigh higher trading costs associated with them. They suggest that the method of prioritizing lagged information variables whih hold high predictability and imply low turnover.

3 April 2020 Paper

3.1 Question 1: Time-Series

Ciritcally examine the time-series regression approach to evaluate linear factor models.

Background

Basic Ordinary Least Squares (OLS) regression analysis is used to test the integrity and accuracy of factor models. It generally relies on linear factor models following the standard of: $E(r_i) = \sum_{k=1}^{K} \beta_{ik} \lambda_k$ (or, as reflected in the formation of r_i: $R_i = R_f + \sum_{k=1}^{K} \beta_{ik} \lambda_k$). This was recognised by Shanken (1992) where: r_i represents the excess return on asset i ($R_i - R_f$); β_{ik} represents the *beta* of asset i relative to factor k (how i moves relative to k); λ_k represents the risk premium of factor k for K factors. This carries $\forall i \{1, ..., N\}$. As well as being used to test models such as the CAPM, ICAPM, and the APT; it was further brought to light by Fama and French (2015, 2016, 2017, 2018) across their analysis of various three-factor and five-factor models, such as the FF3 and FF5. This has been further extended more recently by other academics such as Hou, et al. (2020) with their contribution of models like variants of the HMXZ.

From here, methodologies extend to both time-series and cross-sectional analysis (Fama, 2015). The time-series approach is more commonly used to test linear factor models when the factors are portfolio returns such that $\lambda_k = E(r_{kt}) = E(r_m) - R_f$. This tests portfolio efficiency and assumes single or K factors, no folio contraints apart from the budget constraint, and the existence of a risk-free asset. Fama and MacBeth (1973) also utilise cross-sectional regression analysis to account for economic-focused factors, also. Current focus is on the time-series approach.

Traditional OLS Assumptions

[1] $\Sigma \varepsilon_i \cong 0$

The sum of all residuals should approximately zero, meaning the degree to which observations outlie the line of best fit should be consistent. Errors should be minimised to satisfy $\min \left\{ \Sigma_{i=1}^{N} \varepsilon_i^2 = \Sigma_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2 \right\}$ for $\hat{\alpha} = \hat{\beta}_0$; $\hat{\beta} = \hat{\beta}_1$

[2] $\text{var}(\varepsilon_i) = \text{E}(\varepsilon_i^2) = \sigma^2$

All observations have constant errors, meaning that the variance of the residuals is constant. Therefore, they are homoskedastic; no need for a heteroskedasticity test (White, 1980) or the adjustment to a robust model.

[3] $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ (for i $\neq$ j)

Error terms uncorrelated with one another; observations are exogenous (X_1 does not cause changes in X_2), not endogenous. There is no need for Instrumental Variable approach. No endogeneity problem.

Relevant Factors & Factor Models

Furthermore, more recent factor models were brought to light by Fama and French (2015, 2016, 2017, 2018) in anticipation of analysing more specific characteristics of the nature and type of assets/portfolios. The Fama & French Three-Factor Model (FF3) and Fama & French Five-Factor Model (FF5) are good examples. As of Fama and French (2015), we know these multi-factor models account for stock characteristics such as *market excess return* $(R_m - R_f)$; *size* (small-minus-big (SMB)) which is 'average return on small-firm portfolios minus average return on big-firm portfolios'; *book-to-market* (high[value]-minus-low[growth] (HML)) which is 'average return on value-firm portfolios minus average return on growth-firm portfolios'; *operating profitability* (robust-minus-weak (RMW)) which is 'average return on high-profit-firm portfolios minus average return on low-profit-firm portfolios'; and *investment prospect* (conservative-minus-aggressive (CMA)) which is 'average return on conservatively-investing-firm portfolios minus aggressively-investing-firm portfolios'. Fama and French (2015) find that small, profitable and valuable firms present the highest returns. The FF5, on which this is based, follows:

$$(R_{it} - R_{ft}) = \alpha_i + \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t$$

$$+\beta_4 RMW_t + \beta_5 CMA_t + \varepsilon_{it}$$

More recently, Hou, et al. (2020) presented an alternative five-factor model, the HMXZ5. It is descirbed as containing the 'q-factors', based on 'q-theory' and appearning in previous research/versions of their models, and an additional expected growth factor. It accounts for market excess return, size, profitability (return on eq-

32

uity, as opposed to annual profitability as in the FF models) investment and expected growth. The Fama and French models have been found the generally outperform the Hou, et al. models within relevant comparative metrics. Additionally, Stambaugh and Yuan (2017) propose a four-factor model based on market excess return, size and two mispricing factors. Daniel, Hirshleifer and Sun (2020) propose a three-factor model including market excess return, a long-term financing anomaly factor, and a short-term post-earnings announcement factor. Neither of these are quite as commonly-practiced as the former two.

The Time-Series Model

In the purely time-series method, Fama (2015) provides an overview of the components used to evaluate factors based dominantly upon stock return portfolios. These include: the use of N testing assets/portfolios etc., K factors included in comparison, and T time-series observations of factors and testing assets/portfolios. The regression uses real excess returns (r_{it}) $(R_{it} - R_{ft})$ for asset/portfolio i $(i \in \{1, ..., N\})$, in period t of the total period. The model makes use of excess factor portfolio return also, denoted by r_{kt} $(R_{kt} - R_{ft})$ for factor k $(k \in \{1, ..., K\})$, in period t of the total period. Therefore, it is implied that three sets of returns are required for this model: [1] testing asset/portfolio returns, [2] factor portfolio returns, and [3] risk-free return proxy. Both of these sets of excess returns $(r_{it}$ and $r_{kt})$ are abserved symmetrically over time, for correspodning time-stamps such that $t \in \{1, ..., T\}$. From here, traditional OLS estimates for coefficients are computed and standard error values follow, using basic IID or GMM methods. The model follows:

$$r_{it} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} r_{kt} + \varepsilon_{it}$$

We know that we would expect to produce a random error term which satisfies $E(\varepsilon_{it}) = 0 \ \forall k \{1, ..., K\}$ if pricing error is going to be captured most accurately. So forth, it is implied that:

$$E(r_i) = \alpha_i + \beta_{ik} \lambda_{ik}$$

$$\therefore E(r_i) = \alpha_i + \beta_{ik} E(r_{kt})$$

Pricing Error

In a world of no pricing error, abnormal return or 'pricing error' (α) should equal zero. Basic reduction [of pricing error] methods include the use of portfolios over individual assets, and ensuring T observations is greater than N test assets/portfolios. This never actually removes pricing error however. Testing portfolios are always likely to show abnormality when testing against stock-characteristic-based linear factor models. We still work with the idea that the intercept in the time-series regression is the pricing error, as follows:

$$\alpha_i = r_i - \sum_{k=1}^{K} \beta_{ik} r_k$$

Hence, the pricing error is the abnormal return on asset i; the average excess return of asset i minus expected excess return of asset i from the model. This α_i is also sometimes referred to as the 'Jensen (1968) preformance measure' in relation to other, non-stock-characteristic-based, linear factor models.

Hypothesis Testing

Given the above, we furthermore work with the null hypothesis that this pricing error is equal to zero, that is "intercepts in the time-series regression are all equal to zero". We of course wish to fail-to-reject the null hypothesis to satisfy no pricing error. Hypotheses follow:

$$H_0: \alpha_i = 0;$$
$$H_A: \alpha_i \neq 0;$$
$$\text{For asset } i \text{ where } i \in \{1, ..., N\}$$

From the above, we see this is a 'two-tail' test as the opposing hypotheses are basi c scalars to satisfy "equivalent to or not equivalent to zero?" ($= / \neq 0$), as opposed to a 'one-tail' vector test to satisfy "equivalent to or not equivalent to zero, in which direction?" ($= / \neq 0$ @ < 0, $= / \neq 0$ @ > 0). Observing the α_i ('coefficient' output) value and corresponding signs, for N pricing errors this two-tail test implies:

For Significantly $(+)$ α_i: 'significantly positive pricing error exists'
For Significantly $(-)$ α_i: 'significantly negative pricing error exists'

In relation to the 'significance' portion of this methodology, using very basic hypothesis testing methods, this essentially reports whether or not abnormal returns/pricing error

are/is significant for any assets in the model. Using standard *t-stat* interpretation, we determine the significance of the magnitude of α_i values. The *t-stat* derivation follows:

$$t\text{-}stat = \frac{\alpha_i - \alpha_{i_{H_0}}}{std.err_{\alpha_i}} = \frac{\alpha_i - \alpha_{i_{H_0}}}{\left(\frac{\sigma_{\alpha_i}}{\sqrt{N}}\right)}$$

Using *t-stat* interpretation where *p-values* are not computed but T is large enough, if:

|t-stat| > 1.96: reject the null hypothesis at the 5% significance level;

|t-stat| > 2.58: reject the null hypothesis at the 1% significance level

Alternatively if *p-values* are computed, 1%, 5% and 10% significance are more apparent at the corresponding values *p-values* of 0.01, 0.05 and 0.10, respectively. The 10% significance level tends to be ignored. If:

p-value < 0.05: reject the null hypothesis at the 5% significance level;

p-value < 0.01: reject the null hypothesis at the 1% significance level

Note that 5% and 1% significance levels are also referred to as the 95% and 99% confidence [in rejecting the null] intervals. Correct syntax: refer to 5% and 1% significance levels when referring to p-values; 95% and 99% confidence levels when referring to hypotheses.

For example, let an example $\alpha_i = -0.3396$ and example $t\text{-}stat = -5.1197$. In the most expansive linguistic manner: "we reject the null hypothesis (H$_0$: $\alpha_i = 0$) in favour of the alternative hypothesis (H$_A$: $\alpha \neq 0$) with 99% confidence, in observing a *t-stat* of -5.1197 (which magnitude is greater than 2.58), to say that the pricing error/abormal return of asset i is statistically negative by 0.3396 units at the 1% significance level".

3.2   Question 2: Individual Securities

Discuss the benefits of using individual securities rather than portfolios to test linear factor models using the time-series regression approach.

Piss off.

3.3 Question 3: Market Frictions

Critically evaluate the impact that market frictions can have on tests of portfolio efficiency.

Piss off.

3.4   Question 4: Comparison Tests

Critically evaluate the model comparison tests that can be used to evaluate linear factor models.

Background

When examining and comparing linear factor models such as Fama and French's three-factor model, five-factor model, Hou, et al.'s five-factor model etc., there are various metrics under which components of the models are tested. Generally, the mean-variance efficiency of models tends to be rejected. History shows that no model really presents zero pricing error and thus, instead of look for the model with absolute zero pricing error, it is important to identify the model which shows componenets which minimise pricing error and alike contributories. Model testing usually starts with the Gibbons, Ross and Shanken (1989) Test in which models can be compared on a basic level relative to the null hypothesis of zero pricing error across observations. More expansive model comparisons then extend to pricing error metrics explored by Fama and French (2015, 2016, 2017, 2018), Sharpe performance comparison tests brought ot light by Barillas, Kan, Robotti and Shanken (2020), and factor redundancy tests explored by Fama (2015).

Gibbons, Ross and Shanken Test

Gibbons, Ross and Shanken (1989) developed the Gibbons, Ross and Shanken Test (GRS Test); a method which tests the null hypothesis of zero pricing error (α_i) in asset i for all $i \in \{1, ..., N\}$); "all of the *alpha* values are jointly equal to zero". Thus, this concept relates to the joint distribution of the intercepts of a model being equal to zero. This is a multivariate test for mean-variance efficiency of factor models. It assumes the regression residuals (ε_{it}), when testing for pricing error, have a multivariate normal distribution with zero-mean and a constant covariance matrix. Knowing that $N_{\text{Assets}} < (T_{\text{Observations}} - K_{\text{Factors}})$, the test proceeds. The GRS Test essentially acts as an applied F-test in which all pricing error values (α_i) in a model are equal to zero. Upon observation of the *p-value* produced in the GRS Test relevant to H$_0$: $\alpha = 0$, we tend to reject the mean-variance efficiency of models as most do present significant pricing error.

Computation of the GRS Test makes use of basic matrix algebraic functions such as addition/subtraction ($+/-$), multiplication, transposition ($'$), and inversion ($^{-1}$). We

recall the implied models from the Fama and French (2015, 2016, 2017, 2018) time-series factor model analysis:

$$E(r_i) = \alpha_i + \beta_{ik}\lambda_{ik}$$

$$\therefore E(r_i) = \alpha_i + \beta_{ik}E(r_{kt})$$

After computed the relevant estimates, the GRS Test applies the discussed notes to a vector of all the *alpha* values in a given model, as: $\hat{\alpha} = [\hat{\alpha}_1, ..., \hat{\alpha}_N]$. Hence, "are all the *alpha* values jointly equal to zero?". In search of the test-statistic for this methodology, Gibbons, Ross and Shanken (1989) began with the Wald Test:

$$T\left(\frac{1}{(1 + Sh_k^2)}\right)\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

$$\Rightarrow [\hat{\Sigma}^{-1} = \text{cov}(\varepsilon, \varepsilon')^{-1}]$$

We know from Sharpe (1966) that Sh_k^2 is the maximum squared sharpe performance of factor k in the model. And furthermore, that $\hat{\alpha}'$ is the transpose of the discussed vector of all *alpha* values in the model, and $\hat{\alpha}$ is the [N,1] vector of individual *alpha* values. $\hat{\Sigma}^{-1}$ is the *Maximum Liklihood* (ML) estimate of the residual covariance matrix for N regressions. The Maximum squared sharpe performance is given by:

$$Sh_k^2 = \bar{\alpha}_k' V_k^{-1} \bar{\alpha}_k$$

$$\Rightarrow [V_k^{-1} = \text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}]$$

We know from here that $\bar{\alpha}$ is the [K,1] matrix of mean excess returns on K factors, $\bar{\alpha}'$ is the transpose of this, V_k^{-1} is the inverse of ML estimate of the [K,K] covariance matrix of the mean excess returns on K factors. The Wald Test concludes upon an asyptotic $(T \longrightarrow \infty)$ *Chi-Squared* (χ^2) distribution with N degrees of freedom.

From here, Gibbons, et al. (1989) deduce their applied F-test that there are N cases of no pricing errors, under residual multivariate normality (IID normal). This is based on an F-distribution, and N and T-N-K degrees of freedom and is therefore relevant in smaller samples. Intuitively, the Null hypothesis (H_0) is that there is mean-variance efficiency reflected in all the *alpha* values being jointly equal to zero thus, that $\lambda = 0$. The GRS Test is represented as:

$$\left(\frac{T-N-K}{N}\right)\left(\frac{1}{(1+Sh_k^2)}\right)\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

As a final note in relation to portfolio efficiency, if $K = 1$ then the GRS Test examines whether the market portfolio lies on the efficient frontier of the $(N+1)$ assets. However, if $K > 1$ then the GRS Test examines whether a combination of the K factors which lies on the efficient frontier of the $(N+K)$ assets.

Pricing Error Metrics

Its recognised that when usign the GRS Test for N test assets to compare models, $\text{cov}(\hat{\alpha}, \hat{\alpha}')$ varies across these models. Therefore, the GRS Test can generate inaccurately low results due to sampling error. There are eight pricing error metrics explored and practiced by Fama and French (2012, 2015, 2016, 2018) which focus on the actual pricing error (α) values in the models.

The first metric is the absoulte average *alpha* value $(\bar{\alpha})$ which accounts purely for the magnitude of the *alpha* values. That is, absolute average mispricing. It removes any bias which may occur in models which have *alpha* values which are skewed in one direction. The model with the lower absolute average *alpha* value is the better one as lower (closer to 0) mispricing is identified.

The next metric is the average *t-ratio* of the *alpha* values $(\overline{\text{t-ratio}_\alpha})$. Following *t-stat* protocol, this is as always an indicator of the significance of the mispricing. Therefore, as high significance indicates a significantly great mispricing error, the model with a *t-ratio* reflecting low significance is better. In terms of *t-stat* interpretation, this generally requires a lower *t-ratio*.

The third and fourth metrics are, much like *t-ratio*, based on traditional OLS output. They include the average Adjusted R^2 value of the models $(\overline{\text{Adj. } R^2})$ in which a higher value is sought which can aid explanation of lower standard error values. Furthermore, the fourth metric is the average standard error of the *alpha* values $(\overline{\text{std.err}_\alpha})$. Intuitively, a lower value is saught here as it's implied such model would contain mispricing values (α) closer to 0. That is, lower mispricing.

The fifth metric is a ratio which makes use of squared *alpha* values and the volatility of excess returns on the N testing assets. It's represented as: $\frac{\overline{\alpha^2}}{\sigma_r^2}$. This ratio is a fairly

intuitive ratio which primarily acounts for the real spread in mispricing to volatility. Of course, a lower value is desired. An extention of this metric is the sixth metric; another ratio. It is similar to the former however, the numerator accounts also for the standard error of the *alpha* values. It is represented as: $\frac{\overline{\lambda^2}}{\sigma_r^2}$ (where $\lambda^2 = \alpha^2 - \text{std.err}_\alpha^2$), $\therefore \frac{\overline{\lambda^2}}{\sigma_r^2} = \frac{\overline{\alpha^2 - \text{std.err}_\alpha^2}}{\sigma_r^2}$. It is known as the average real mispricing as it is said to account for and remove sampling error. Although not 100% mathematically sound or correctly in-line with various mathematical theory, this method better accounts for vector distance, or 'spread' in the case of 'finance', in relation to mispricing and excess return volatility. Of course again, a lower value is sought.

The seventh metric another based on basic output; the average squared standard error value of *alpha* values over average squared *alpha*: $\frac{\overline{\text{std.err}_\alpha^2}}{\alpha^2}$. It is said to specifically focus upon and capture the proportion of the mispricing across the N testing assets which is attributed to sampling error. Thus, as you wish the 'average real mispricing' metric [6] to present a lower value than the basic mispricing metric [5], you are anticipating larger sampling error in better results. Therefore, the model with a greater value in this metric would be one generally with a better (lower) 'average real mispricing' (greater attribution of mispricing to sampling error, in the case of mispricing). Thus, higher value is better.

The final, eighth, metric is simply the average squared Sharpe performance of the optimal portfolio. Therefore, denoted by: $\overline{\text{Sharpe}_\alpha^2}$. Of course, model with the higher value is the better one.

Sharpe Performance Measures

Furthermore, Barillas and Shanken (2017) conduct further research regarding the integrity of model testing on N assets across time-series models. They make the argument that the choice of testing assets in comparison is irrelevant. This is based on the principal that comparing models using the GRS metric highlights the fact that increases in Sharpe performance approaching $Sh_{k_A,k_B,r}^2$ sees a better-off model A when it's lower than model B. Considering the fixed-nature of $Sh_{k_A,k_B,r}^2$ (and that under the GRS metric, Model A: $(Sh_{k_A,k_B,r}^2 - Sh_{k_A}^2)$ and Model B: $(Sh_{k_A,k_B,r}^2 - Sh_{k_B}^2)$), there would be an indifference in the former and comparison simply upon $Sh_{k_A}^2$ and $Sh_{k_B}^2$. Thus, irrelevant choice of test assets.

The aim is to compare models on a nested or non-nested basis. That is, models which are endogenous/not endogenous of each-other. The methodology is based on the idea that "if you include all the factors in the investment universe, the test assets become irrelevant" . For example, [you are comparing the FF3 and FF5] the Fama and French FF3 is endogenous of the Fama and French FF5 as the FF5 is an expanded version of the FF3 which holds many of the same characteristics. Contrastingly, [you are comparing the FF5 and the HMXZ5] the Hou, et al. HMXZ5 is not endogenous of the FF5. Therefore, the FF3 vs. FF5 is a 'nested' test; the FF5 vs. HMXZ5 is a 'non-nested' test.

Barillas and Shanken (2017) note that the issue to be addressed in model comparisons is "how well a factor model actually prices factors which are not included in the model". They add that comparing models on this basis often presents different results to traditional tests of mispricing. They proceed with what they call 'Proposition 1'; a test of: $Sh^2_{k_A} = Sh^2_{k_B}$. Take, for example, the nested comparison of the FF3 and FF5; how well does the smaller (nested) model (FF3) price the factors not included from the FF5. Thus, let k_A denote factors in the smaller model for L factors (FF3 factors), and let k_B denote factors in the larger model for K factors ($L + J$ (L factors of model A and J additional factors (intuitievly: $J = K - L$))).

Notation simplification:

- Smaller Model (k_A): L Factors
- Larger Model (k_B): $K = L + J$ Factors
- Non-Endogenous Factors: $J = K - L$

So forth, the following test is conducted:

$$k_{jt} = \alpha_j + \sum_{l=1}^{L} \beta_{jl} k_{lt} + \varepsilon_{jt}$$

Note that k_{jt} denotes excess returns on the J factors which are excluded (additional factors in the larger model), and k_{lt} denotes excess returns on the L factors actually accounted for by the smaller model. The test of this model operates on the basis of the null hypothesis (H_0) of $\alpha_j = 0 \ \forall j\{1,..,J\}$. Thus, stating that both models produce equivalent excess return predictions and hold the same squared Sharpe performance. If this hypothesis was failed-to-be-rejected, J factors are redundant and the smaller model is preferred. If the hypothesis is rejected however, the larger model is preferred,

with a greater squared Sharpe performance.

Furthermore, Barillas, Kan, Robotti and Shanken (2020) extend reearch in this area to formulate the test of maximum squared Sharpe perfromance. The nested testing uses the GRS methodology however, Proposition 1 also allows analysis of the metric 'difference in squared Sharpe performance'. This of course, requires that the Sharpe performance of at least one of the models is greater than zero. In some cases, non-nested models with minor overlapping factors may still support $Sh^2_{k_A} = Sh^2_{k_B}$ if the overlapping factors hold the same Sharpe performance of combined factors of both of the models. Barillas, et al. (2020) further expand upon this with a test of equality of Sharpe performance, which makes use of a 'benchmark' model. It operates on the null hypothesis that the benchmark has an equivalent Sharpe performance as the other models in question. The alternative hypothesis states that at least one of the other models has a *higher* Sharpe performance than the benchmark.

From here, testing can be observed more-closely on a nested/non-nested basis. Using a benchmark which is nested in the other models allows use of a close equivalent of the GRS methodology to test for aggregate zero-*alpha*-value. This uses excluded factors (of the other model with the widest scope) from the benchmark model to test for Sharpe equivalence. An example given for this uses comparison of the CAPM to the Fama and French FF3 and FF5. This would involve the GRS test of H_0: $\alpha = 0$ cross-overs SMB, HML, RMW and CMA fators. Upon rejection of H_0, it's implied that the benchmark has a significantly lesser Sharpe performance than the other models in question. Additionally observing non-nested models, the null hypothesis that the benchmark presents an equivalent or greater Sharpe performance to those of the other, non-nested, models is drawn. Barillas, et al. (2019, 2020) make use of various methods of Wolak (1987) in this context. This denotes $(Sh^2_{kl} - Sh^2_{ki})$ as δ_i for $i \in \{2, ..., p\}$, in the case pf p models. Thus, H_0: $\delta \geq 0$. This of course accounts for benchmark $p-1 = k$.

The output for relevant tests consists of the adjusted squared Sharpe measures of each model in question (and associated *p-values*), and the difference in adjusted squared Sharpe performance (and associated *p-values*) across the models. Ferson and Siegel (2003) highlight a 'bias-adjusted' Sharpe performance as this 'adjusted squared Sharpe performance', which accounts for bias in Sharpe performance where N is great relative to T:

$$\text{Adj. Sharpe}^2 = \text{Sharpe}^2 \left(\frac{(T - K - 2)}{T} \right) - \frac{K}{T}$$

Included also is the associated *p-value* in relation to the above which is based Barillas, et al.'s (2020) 'Proposition 1'.

Factor Redundancy Tests

Fama and French (2015) exercised a method of 'factor spanning' as they referred to it. This methodology asks the question of "can *beta* values of other factors in the model explain the mean excess return of a factor?"; "are all of the factors required?". That is, The role that a factor may play could already be captured by others; coefficients of other factors may already reflect what a factor contributes. Therefore, it may have approximately no effect on excess returns of the model during the sample period; mean excess return of the model would be approximately equal without the factor. To answer this question, factor redundancy regressions are run for each factor portfolio's excess returns. The model is as follows:

$$r_{pt} = \alpha_p + \sum_{k=1}^{K-1} \beta_{pk} r_{kt} + \varepsilon_{pt}$$

As we'd expect, r_{pt} denotes the excess returns of the factor portfolio at time t. p is introduced to denote a single factor portfolio representing one $k \in \{1, ..., K\}$ hence, when left with the explanatory side of the new model: $k \in \{1, ..., (K-1)\}$. The model is based on the null hypothesis that a factor adds ≈ 0 to excess returns (H$_0$: $\alpha_p = 0$). Thus, we wish to reject this null hypothesis; saying a factor is not redundant. If a factor is redundant, it's worth omitting.

Furthermore, Fama and French (2018) the question of marginal contribution. That is, not just if a factor is redundant or worth including but, what its marginal contribution to the model actually is. Numerically, this contribution is given by the marginal contribution to the maximum squares Sharpe performance. It's represented as: $\frac{\alpha^2}{\sigma_\varepsilon^2}$, which accounts for the *alpha* values and the volatility of residuals. There remains the question of what drives a factor's contribution; a high *alpha* value ($\uparrow \alpha$) or a low volatility of residuals ($\downarrow \sigma_\varepsilon$)? It's good to observe fairly distributed marginal contribution; individual components of the metri must also be observed to determine why a factor may have an exceptionally high/low marginal contribution however. Obviously, this method can be used to highlight heavily dominant factors despite its naming-after the opposite.

3.5 Question 5: Conditioning

Discuss the benefits of using conditioning information in mean-variance analysis.

Background

As the great John H. Chochrane says, "we do not live in a world where each day's returns are based on a coin flip". Essentially, variables vary over time; as the name should suggest. This idea of 'conditioning' is based on the idea of more complex models of expected returns, *beta* values, risk, etc., being 'conditional' upon 'information sets' which may not have been previously considered. These information sets such, as varying covariance [matrices], can be produced by time-varying perceptions and attitudes of individuals, markets, and analysts etc. and their ability to produce results/predictions based on the access to information they have. And so on. In other words, methodologies surrounding this have a lot to do with subjectivity of internals and some externals. An alternative is that there are 'unconditional' expectations.

Prior to consideration of conditions, mean-variance efficiecny of a factor model has been tested based adjacent to a fixed-weight efficient frontier. This contradicts the idea of conditioning as it relies upon time-constant expected return, covariance and risk etc. Based on the discussed, we define 'conditioning information' as lagged information variables which hold some form of predictive nature of future results, based upon mostly subjective componenets. These lagged information variables are generally used to further aid prediction of expected returns, covariance and risk etc., also known as 'conditional moments'. This allows investors to expand their trading strategies into a more dynamic and organic nature.

Expected Return Framework

Generally, we derive the expected discounted payoff framework, with regrds to the above, considering primrily the 'agent' information set. That is, the information set based on the perceptions and attiudes of individuals in the market. We refer to their subjective expectations about the future. The theory of *the law of iterated expectation* states that:

$$E(E(x)_c) = E(x)$$

The expectation of a conditional expectation is equivalent to the expectation. That

is: "what you think today about what you are going to think tomorrow about the day after tomorrow is equivalent to what you think today about tomorrow" ($E(E(x_{t+2})) = E(x_{t+1}) \; \forall t\{t \rightarrow \infty\}$). However, we cannot accurately quantify simply what is 'in people's heads'. So forth, proxies for conditioning information relative to the agent stem from time-series lagged dividend yield on the market index, short-term Treasury 'T-' Bill rate, term/interest spread(s) and default spread. These give accurate indciators of agent behaviour and thus, predictability. Following the theoretical framework above, these can be modelled to satisfy mean-variance efficiency with conditioning information. We use the term 'conditional mean variance' (CMV) frontier to denote the equivalent to the efficient frontier upon which a portfolio lies if it is efficient to the conditional information. In the research of Hansen and Richard (1987), this must satisfy:

$$\min\{E(R_{pt}^2)|Z_{t-1}\}$$

$$subject\ to$$

$$E(R_{pt}|Z_{t-1}) = \text{Target Level} \wedge x'_{Z_{t-1}}1_N = 1$$

In other words, a portfolio lies on the CMV frontier if it satisfies minimization of $E(R_{pt}^2)$ such that a lagged information variable exists and is present, and is subject to $E(R_{pt}|Z_{t-1})$ being present at the target level and optimal weights (x_{Z_t}). Furthermore, Hansen and Richard (1987) also denote the equivalent of the mean-variance frontier (in the presence of conditioning), without conditioning information, as the 'unconditional mean-variance' (UMV) frontier. It displays the set of optimal weights (x_{Z_t}) to satisfy:

$$\min\{E(R_{pt}^2)\}$$

$$subject\ to$$

$$E(R_{pt}) = \text{Target Level} \wedge x'_{Z_{t-1}}1_N = 1$$

In other words, a portfolio lies on the UMV frontier if it satisfies minimization of $E(R_{pt}^2)$ which is subject to $E(R_{pt})$ being present at the target level and optimal weights x_{Z_t}. Hansen and Richard (1987) also highlight, as extensively discussed over various appearances from John H. Chochrane, that some portfolios which lie on the UMV frontier also lie on the CMV frontier however, other portfolios which lie on the CMV frontier do not generally lie on the UMV frontier. Finally, for reference, we recall the most 'naïve', if you will, version of the mean-variance efficient frontier; what we now denote ' 'passive mean-variace' (PMV) frontier' as. It is simply subject to optimal

$E(R_{pt})$ and weights. As follows:

$$\min\{E(R_{pt}^2)\}$$

$$subject\ to$$

$$E(R_{pt}) = \text{Target Level} \wedge x'1_N = 1$$

In effect, the CMV forntier optimizes the trade-off between expected payoff and conditional risk based on information and other attributes, in a time-series format. On the othe rhand, the UMV frontier optimizes the trade-off between expected payoff and unconditional risk based upon non-informational attributes, in a time-series format. The UMV frontier is essentially providing a similar framework to the PMV frontier however, it allows for the consideration of attributes which contribute to more dynamic trading strategies.

Using Information

As we noted, generally we like to consider what the agent is thinking. Ferson and Siegel (2001) explore the case that where clients have less information than portfolio managers, the cleints desire a portfolio which lies on the UMV frontier as opposed to the CMV frontier due to various unaccountable and untransferrable informational and behavioural attributes and characteristics in the case of the porfolio manager's use of the CMV frontier. The use of a CMV frontier portfolio in this case can result in poorer performance compared to the UMV equivalent. Ferson and Siegel (2001) argue that the UMV portfolios can be more resillient to misspecification of conditional moments as attributes of the UMV portfolios are more valid than CMV ones in the case of misspecification. Neither are optimal however.

A slightly different approach to incorporating conditional information without extensive modelling of conditional moments is called the 'multiplicative approach'. The method is in the title. You assume L lagged information variables and multiply asset returns by each according lagged information variable, individually. This results in a larger-scoped pool for an optimal portfolio to be selected from.

Deriving Frontier Componenets

Ferson and Siegel (2009) expand upon their previous work, beginning with the statement that all pricing models, models with economic factors, and models with informa-

tion conditioning see a portfolio on the UMV frontier. They first derive the optimal portfolio weights, in the presence of a fixed risk-free asset set of return of (R_f). This follows:

$$x_Z = \left(\frac{u_s - R_f}{B} \right) Q(u_Z - R_f)$$

Where :

$$Q = \left((u_Z - R_f)(u_Z - R_f)' + cov_Z \right)^{-1}$$

$$B = E \left((u_Z - R_f)' Q(u_Z - R_f) \right)$$

x_Z obviously denotes the optimal weight, u_s in the target expected return, R_f is the risk-free return, u_Z denotes conditional expected returns, and cov_Z is the conditional covariance matrix. First, to estimate conditional moments u_Z and cov_Z, Ferson and Siegel (2009) present the formation of the time-series regression:

$$R_{it} = \alpha_i + \sum_{l=1}^{L} \beta_{il} Z_{l(t-1)} + \varepsilon_{it}$$

This highlights L lagged information variables $Z_{l(t-1)}$ of lag l for $l \in \{1, ..., L\}$. From here, fitted values are interprted as conditional expected returns (u_Z) and the covariance matrix (cov_Z) is calculated from produced components. It can be noted that Ferson and Siegel (2009) also consider a covariance matrix which varies over time however, they find little change in results when accounting for this.

Testing The Model

Ferson and Siegel (2009) extend their research into testing their methodology using a form of the traditional Wald test in a similar form to the Gibbons, Ross and Shanken (1989) test (GRS Test), to test mean-variance efficiency; whether or not a portfolio lies on the UMV frontier. As always, this accounts for K risky assets in the benchmark model and N testing assets. Recall the form of the Wald test:

$$T \left(\frac{S_r^2 - S_{r_p}^2}{1 + S_{r_p}^2} \right)$$

As we know, S_r^2 is the squared Sharpe performance introduced by Sharpe (1966) for the entirity of the $(N + K)$ assets. With the inclusion of optimal conditioning information, this is the squared Sharpe performance of the UMV portfolio. $S_{r_p}^2$ is the squared Sharpe

performance of the K assets in the benchmark. Hence, mean variance efficiency exists for the benchmark model if $S_r^2 = S_{r_p}^2$. Using a statistical probability simulation method, known as the *Monte Carlo Simulation*, which generates hypothetical values in the case of extreme uncertainty. This is present under the null hypothesis of tested folios being present upon the UMV frontier. Ferson and Siegel (2009) find that in the presence of conditioning information, the UMV frontier holds a higher Sharpe performance over a longer period of time. They also reject mean-variance efficiency of the CAPM and the FF of Fama and French (1993), both in the presence of conditional information and not.

Issues

A primary issue highlighted by *the legends themselves* Basu and Fletcher (2016) is that strategies related to UMV optimization require high turnover so frequently when adjusting for trading costs, the UMV approach may require significantly greater outper-formance. It does not always hold this characeristic. They ask the question of whether the benefit of the predictability using discussed methods outweigh higher trading costs associated with them. They suggest that the method of prioritizing lagged information variables whih hold high predictability and imply low turnover.