1 Introduction - Underlying theory and rationale

1.1 Theoretical Framework

Theoretical asset pricing relies on factor modeling to forecast returns. In this paper, I will first present the underlying logic behind these models, then I will argue for the need of machine learning methods to get more efficient results. First, much of Asset Pricing theory is based on the assumption of the law of one price, that is, the premise that all portfolios of securities with the same payoff should have the same price [1]. This is a weak assumption from an economical point of view - as it is a much less restrictive than classical utility functions assumptions for instance - yet, it entails the presence of a payoff pricing functional, i.e. a function that takes as input the payoffs [2] and returns a price in $\mathbb{R}$ [3] . LeRoy et al.[13]define the pricing functional $p : \mathbb{R}^s \to \mathbb{R}$ such that $p(X) = p$; for s the number of different states of nature; $\mathbb{X}$ a subspace of $\mathbb{R}^s$ spanned by the payoffs x and p the price of a portfolio. Notice how $\mathbb{X}$ doesn't span the whole space $\mathbb{R}^s$ as markets are assumed incomplete i.e. investors cannot buy (or synthesize) any contingent claims, and hence the payoff space is constrained; Rendering the model more realistic [4]. In addition, $p(X)$ is assumed to be linear: i.e for two payoffs x and x' we have $p(\alpha x + \gamma x') = \alpha p + \gamma p'$ for any α,γ $\in\mathbb{R}$.

These loose assumptions are the building blocks of asset pricing factor modeling, as they guarantee the existence of a stochastic discount vector (SDF). In fact, by the Riesz Representation theorem: for a linear functional $L(.)$ on some Hilbert space $\mathcal{H}$, $\exists v \in \mathcal{H}$ such that $\forall u \in \mathcal{H}$, $L(u) = \langle u, v \rangle$. Riesz theorem applies in our framework: given that $p(X)$ is a linear functional and $\mathbb{E}[XY]$ (for some some random X, Y in the vector space) is an inner product[5], then, for $p(X)$ s.t $p : \mathbb{R}^s \to \mathbb{R}$ then $\exists$ a unique $m \in \mathcal{X}$ which is defined as the stochastic discount factor , such that $p(X) = \langle x, m \rangle = \mathbb{E}[mx]$. Simply stated we can represent the linear pricing functional by an inner product with a unique vector, the SDF. This fact is more evident when illustrated geometrically:

[1] With a sufficient and necessary condition that every portfolio with no payoff has a price of zero

[2] At all different states of nature

[3] A functional is a mapping from some vector space to real numbers.

[4] To understand this ; here is a concrete example: Assuming two states of natures; we can imagine a simple market in which there is a single security that either pays 1.1 or .9. The payoff space is "constrained" in the sense that we cannot synthesise the security to get any payoff we want : we either get 1.1 or .9 or some scaled version of them; the codomain $\mathbb{X}$ is thus a 1 dimensional plane defined by a scaled version of the payoff in this state space in R^2. This logic is generalizable to realistic markets by assuming high dimensional state space representations.

[5] $\mathbb{E}[XY]$ is an inner product as it respects the three defining axioms of inner products: symmetry, bi-linearity and positive definiteness)

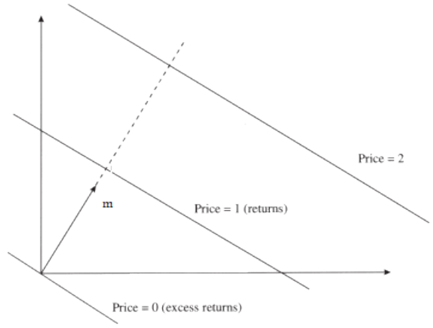

Figure 1: Geometric proof of the existence of a unique SDF m s.t. $E[mX] \equiv p(X)$. To prove it we visualize the SDF in the payoff space $R^{\mathbb{X}}$ - Not to be confused with the space of states of nature R^s

Geometrically (see Figure 1 [3]), This equivalence between $E[mX]$ and $p(X)$ can be represented in a state space representation in $\mathbb{X}$, the span of payoffs in an incomplete market [6]. I will consider for simplicity a two-dimensional payoff plane $\mathbb{X}$, but the same logic is generalizable to any dimension of X. Since each price is represented by a linear function of their corresponding payoffs (Due to the linearity of $p(.)$), we can draw parallel prices hyperplanes [7] linearly in $R^{\mathbb{X}}$ .We pick some vector m orthogonal to the price hyperplanes; then for any payoff belonging to some price plane $= a$, s.t. a represents some fixed price, the inner product between any payoff on p $= a$ and m is the same [8] .Formally: $< m, x_{\text{on price hyperplane a}} > = cst_{\text{for each hyperplane plane a}}$ implies the existence of a vector m orthogonal to all the price hyperplanes in $R^{\mathbb{X}}$ such that $E[mX]$ and $p(X)$ are equivalent.

In addition to the existence of a unique SDF, assuming the absence of arbitrage (along the law of one price) we get that the stochastic discount factor is always positive. In fact, $p(x) = \mathbb{E}(mx) = \int m(s)x(s)\pi(s)ds$ for s the number of States of Nature and $\pi(s)$ the probability of some state of nature. Assuming that prices and payoffs are always positive, and since probabilities are always positive; then $m(s)$ is always positive $\forall$s .

Importantly, one can derive from $p(x) = \mathbb{E}(mx)$ - Also called the "Euler" (or "basic") model - a single Factor model by Simple algebraic manipulations[9] such that

$$E(R^i) - R^f = \beta_{\text{i,m}}\lambda_m \tag{1}$$

That is, the expected premium return for each security is determined by a price risk λ_m ,common to all securities, and the regression coefficient from regressing this security's return on some stochastic

[6] which is a subset of R^s

[7] Or "planes" - like in Figure - depending on the dimensions of the payoff space.

[8] This is implied by construction and is seen clearly in the figure above, since the inner product between x and m is $|\text{proj}\,(x \mid m)| \times |m|$

[9] The payoff R corresponds to a unit price; One can thus write the Euler equation and decompose it as such $1 = \mathbb{E}(mR) = \text{cov}(m,R) + \mathbb{E}(m)\mathbb{E}(R)$. We divide both sides by $E[m]$ and get $\mathbb{E}(R) = \frac{1}{\mathbb{E}(m)} - \frac{\text{cov}(m,R)}{\text{var}(m)} \cdot \frac{\text{var}(m)}{\mathbb{E}(m)}$ with $R^f = \frac{1}{E(m)}$ and with $\frac{\text{Cov(m,}R)}{\text{Var(m)}}$ the linear regression coefficient from regressing R on m with both R and m demeaned variables.

discount factor m.

This approach to asset pricing this Powerful: By assuming the law of One price and the absence of arbitrage, which are mild assumptions from an economical perspective, we were able to derive the Euler equation and represent expected Returns with respect to some unique stochastic discount Factor as a single Factor model. This is also generalizable to multiple Factor models by assuming factors which are linearly related to the SDF. Hence, for any vector of factors f s.t. $m = a + b'f \Longleftrightarrow \mathbb{E}\left(R^i\right) = R^f + \lambda'\beta i$ With β_i Coefficient vectors of the regression of R^ion f . This is the underlying Theory upon which Factor models are built and Asset Pricing Theory (APT) is essentially concerned with constructing an adequate SDF and relating it to the data. In other words, APT begins with the following assumption $p(x) = \mathbb{E}(mx)$, and ultimately aims to comprehend how $m = f(\text{data})$.

"Traditionally", economically motivated stochastic discount factors (and hence Factor models) obey to two different logics: The Arbitrage Pricing Theory And the General Equilibrium logic. They both differ in their economic assumptions and their inspiration for factors. From one side, general equilibrium models are absolute pricing models derived by expressing the stochastic discount factor with respect to marginal utility - for some assumed $u(.)$ - and restricting the model by imposing economic assumptions [10]such that the SDF is a linear function of some factor(s). These models are generally sub models of the broad Intertemporal Capital Asset Pricing model. On the other hand, APT models are relative asset pricing models derived by picking factors that have a small $R2$ when regressed on return. The APT relies on the assumption of an upper bound on the Sharpe ratio, suggesting that a small $R2$ of the regression of expected return on the picked factors implies that the intercept is small and hence that the factors are well specified, implying a factor model. [11] However, ultimately, both logics express the same representation of Returns as linear function of factors. These economically motivated models do not perform well out of sample for evident reasons[12] .This phenomenon is extensively documented. For instance, Fama French disproved ICAPM models by introducing new variables to the model and testing their coefficients; Mehra and Prescott (1985) [17] demonstrated that the widely used consumption-based models (i.e. what I define as "General equilibrium logic" models) do not align with empirical data. Additionally, APT models also exhibit out-of-sample failure. Rapach and Zhou (2013)[21], in their survey on return predictability, clearly highlight the shortcomings of well-known factor models in out-of-sample tests. Furthermore, Bossaerts and Hillion (1999) [1]and Goyal and Welch (2003, 2008)[31] [6]argued that, despite the high in-sample forecastability of some well-known factor models, many popular models fail to outperform naive benchmarks on out-of-sample data. Roll

[10]Those are assumptions on wealth, on time periods, on salary etc...

[11]The ICAPM is considered to be an "absolute" model because it provides a framework that explains the price of any security, while the APT is "relative" because it lacks a fundamental rationale for all assets as it explains returns relative to other returns (without explaining them).

[12]Economic theory is restrictive. And if not tested out of sample, we will always be able to find some model that mimicks the data; but this is not forecasting anymore

(1977)[22] formalized another issue with economically motivated factors in his critique of the validity of empirical tests, arguing that factors are equivalent to mean-variance frontier returns, and that it is always possible to construct overfitting models by picking in-sample mean-variance efficient returns. In other words, many commonly used factor models do not generalize well to new, unseen data; as they tend to overfit on the training data.

1.2 Necessity for Machine Learning Methods

Because economic Theory imposes restrictive and non-realistic assumptions, there has been a growing literature that rather focuses on a pure linear algebra interpretation of stochastic discount factors (and Factor models in general) . The initial problem remains the same from this perspective, the Euler equation remains the foundational structure of the model, but, the stochastic discount factors' construction is tackled differently. Accordingly, many "modern" academic asset pricing papers study the problem as a linear algebra problem and interpret the Stochastic discount factor as factor loadings accordingly: Principal component analysis for example - among other dimension reduction techniques - has been extensively documented in asset pricing literature: Stock and Watson (2002) [24] use principal component to summarize the number of macro economic predictors in macroeconomic forecasts, Ludvigson and Ng (2007)[15]use both Factor Analysis in correspondence with Principal Component Analysis to summarize the feature space, and gets statistically significant out-of-sample results and Nagel (2021)[19], for example, argues for the need of dimension reduction methods for better out of sample performance... References to dimension reduction techniques in empirical asset pricing academia are ubiquitous, reflecting the abundance of dimension reduction methods (and candidate regularization methods associated to each of them). Hence, Machine learning emerges in factor modeling though the prism of linear algebra.

In addition, because of the growing availability of huge financial data, machine learning emerges also as a practical alternative to standard classical models which fail statistically under high dimensional factor modelling . By introducing l_p regularization, dimension reduction and non linearity to the standard factor modeling framework, the model is now able to adapt to high dimensional data and hence incorporate complex dynamics. Accordingly, many academicians tackled the problem as such: Lewellen (2015)[14], for instance, uses 15 features, Freyberger et al. (2020)[5] use 36, Gu et al. (2020b) [9] use approximately a thousand factor and Nagel (2021)[19] argues that the adoption of high-dimensional factor modeling is inevitable and highlights the growing prevalence of factors in literature as evidence supporting this trend.

In addition, machine learning methods, focus on practical out of sample measures to evaluate a model, this is in contrast with standard asset pricing testing, which rely on in-sample goodness of fit tests. Typically, for a generic factor model, $E[R^{\mathrm{ei}}] = \beta_{im}\lambda_m$, one first runs a time series regression to find the beta, regresses the expected excess return on the beta, and tests the significance of the model using GLS, Wald...or some other in sample test. This typical Procedure, is not robust, as it neglects the out of sample performance; while the essence and strength of machine learning lie in its ability to provide reliable out-of-sample performance metrics through the utilization of diverse resampling methods.