Theory of Finance at Work

Carlo A. Favero,[1]

[1]Bocconi University & CEPR

20135 Lecture 1

The Empirical Application of Theory of Finance

- Putting theory of finance at work, implies the specification of optimal portfolio weights and their estimation.
- Theory of Finance specifies optimal portfolio weights as determined by the distribution of future expected returns.
- However, the distribution of future expected returns is not observable.
- The implementation of theory of Finance requires the estimation of the distribution of future expected returns.

Choosing Portfolio Weights

To illustrate the problem we shall consider:

- Standard Modern Portfolio Theory,
- Risk Parity Portfolios

A Static Asset Allocation Problem

Let's denote with $\mathbf{r}$ the random vector of linear total returns from time t to time T from a given menu of N risky assets for interval $[t, T]$, $\mathbf{r} \sim \mathcal{D}(\mu, \Sigma)$

Given a degree of risk aversion λ, a standard *mean-variance* description of this allocation problem is the following:

$$\max_{\mathbf{w}} \left(1 - \mathbf{w}'\mathbf{e}\right) r^f + \mathbf{w}'\mu - \frac{1}{2}\lambda(\mathbf{w}'\Sigma\mathbf{w})$$

where $E[\mathbf{r}] = (1 - \mathbf{w}'\mathbf{e})\, r^f + \mathbf{w}'\mu = r^f + \mathbf{w}'(\mu - r^f\mathbf{e})$ and $Var[\mathbf{w}'\mathbf{r}] = \mathbf{w}'\Sigma\mathbf{w}$

Deriving Optimal Weights

first-order conditions (FOCs) are necessary and sufficient and define the following system of N linear equations in N unknowns, the portfolio weights $\mathbf{w} \in \mathcal{R}^N$:

$$(\mu - r^f \mathbf{e}) - \lambda \Sigma \mathbf{w} = \mathbf{0}.$$

Solving the FOCs yields:

$$\hat{\mathbf{w}} = \frac{1}{\lambda} \Sigma^{-1} \left(\mu - r^f \mathbf{e} \right),$$

No risk free asset and the Tangency Portfolio

Consider now the special case in which $\hat{\mathbf{w}}'\mathbf{e} = 1$, that is no investment in the riskfree bond is allowed. The optimal portfolio in this case is the famous *tangency portfolio*:

$$\mathbf{e}'\hat{\mathbf{w}} = \frac{1}{\lambda}\mathbf{e}'\Sigma^{-1}\left(\mu - r^f\mathbf{e}\right) = 1 \Longrightarrow \lambda = \mathbf{e}'\Sigma^{-1}\left(\mu - r^f\mathbf{e}\right)$$

$$\hat{\mathbf{w}}^T = \frac{\Sigma^{-1}\left(\mu - r^f\mathbf{e}\right)}{\mathbf{e}'\Sigma^{-1}\left(\mu - r^f\mathbf{e}\right)},$$

The Minimum Variance Portfolio

In case the target is to find the minimum variance portfolio:

$$\min_{\mathbf{w}}(\mathbf{w}'\Sigma\mathbf{w})$$

subject to

$$\mathbf{w}'\mathbf{e} = 1$$

the solution will be:

$$\mathbf{w} = \frac{\Sigma^{-1}\mathbf{e}}{\mathbf{e}'\Sigma^{-1}\mathbf{e}}$$

Risk Parity

Decompose the total variance of a portfolio in the sum of the contributions of each asset to the total portfolio variance

$$
\begin{aligned}
Var[\mathbf{w}'\mathbf{r}] &= \sum_{i=1}^{N} w_i Cov\left(r_i, \mathbf{w}'\mathbf{r}\right) \\
\mathbf{w}'\Sigma\mathbf{w} &= \sum_{i=1}^{N} w_i (\Sigma\mathbf{w})_i
\end{aligned}
$$

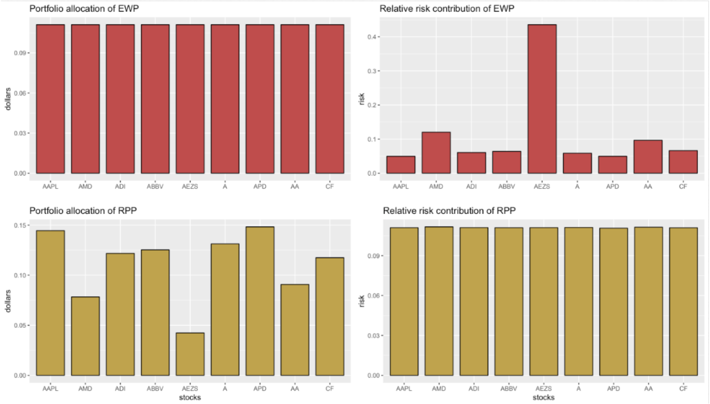
the risk contribution of each asset to total risk can then be written as follows:

$$RRC_i = \frac{w_i(\Sigma\mathbf{w})_i}{\mathbf{w}'\Sigma\mathbf{w}}$$

Risk Parity Portfolios are constructed by choosing weights so that:

$$RRC_i = \frac{1}{N}$$

Risk Parity and Equally Weighted Portfolios



From Theory to Practice

- To apply theory we use past observations on the data to predict the distribution of future expected returns (in the mean-variance approach only the first two moments of the distribution are relevant).
- to map available data into predictions a model is needed
- a model is a structure that captures the feature of the past data and can be simulated to predict the future and the associated uncertainty
  - a model allows first to construct a portfolio based on predictions for μ and Σ
  - a model also allows to simulate the distribution of future returns of a given portfolio to measure risk.

How to build a model for observational data ?

- in finance data are observational: normally we cannot construct experiments to generate data for a treatment and a conrol group and learn form their difference.
- The modelling process uses the "past available data" to predict the future distribution of returns. This is a process that provides useful information only if the features of past data captured by the model persist in the future.
- there are different approach to modelling:
  - select the relevant information with a target variable to predict (supervised machine learning)
  - select the relevant information in absence of a specified target (unsupervised machine learning)

Predicting the distribution of future returns

- Data are not generated by experiments, we have only "observational data"
- Data can be organized in input variables (the information we use to predict) and output variables (the targets of our predictions: returns)
- We use the data by building, estimating and simulating models that relate input variables to target variables
- models need to be validated to minimize the risk of using a "wrong" model
- modelling requires the use of statistical software to analyze the data

The Modelling Process

- The modelling process uses the ”past available data” to predict the future distribution of returns. This is a process that provides useful information only if the features of past data captured by the model persist in the future and involves several steps:
    - Data collection and transformation
    - Graphical and descriptive data analysis
    - Model Specification
    - Model Estimation
    - Model Validation
    - Model Simulation

The Empirical Modelling of Asset Prices

There has been a remarkable evolution of the understanding and empirical modelling of asset prices and financial returns from the 1960s onwards.

- The view from the sixties was based on the Constant Expected Returns (CER) model and the CAPM, when a simple econometric model serves the purpose of modelling returns at all horizons and a one-factor model determines the cross-section of assets returns.
- Several empirical failures of this view have led to the development of Time-Varying Expected Returns (TVER) model where predictability becomes a factor and heterogeneity in predictability is introduced according to the horizon of returns.

The view from the 1960s: Efficient Markets and CER

The history of empirical finance starts with the “efficient market hypothesis”. This view, that dominated the field in the 1960s and 1970s, can be summarized as follows :

- expected returns are constant and normally independently distributed;
- the CAPM is a good measure of risk and thus a good explanation of why some stocks earn higher average returns than others;
- excess returns are close to be unpredictable: any predictability is a statistical artifact or cannot be exploited after transaction costs are taken into account;
- the volatility of returns is constant.

The Model from the 60s': the Constant Expected Returns Model (CER)

In practice, the traditional view can be recast in terms of the simplest possible specification for the predictive models for returns, i.e., the constant expected returns model. Model Specification leads to the following:

$$\begin{aligned} \mathbf{r}_{t,t+1} &= \mu + \mathbf{H}\epsilon_{t+1} \\ \Sigma &= \mathbf{H}\mathbf{H}'. \\ \epsilon_{t+k} &\sim \mathcal{D}\left(\mathbf{0}, \mathbf{I}\right) \end{aligned}$$

where $\mathbf{r}_{t,t+k}$ is the vector of returs between time t and time t+k in which we are interested, μ is the vector of mean returns and the matrix $\mathbf{H}$ determines the time invarying variance-covariance matrix of returns.

Model Estimation

Model estimation allows to find values for $\mu, \mathbf{\Sigma}$. In the case of CER this step is easily solved by n OLS regressions of the n returns on a constant.

$$\hat{\mu^i} = \frac{1}{T} \sum_{t=1}^{T} r^i_{t,t+1}$$

$$\hat{\sigma}_{ii} = \frac{1}{T} \sum_{t=1}^{T} \left(r^i_{t,t+1} - \hat{\mu^i} \right)^2$$

$$\hat{\sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} \left(r^i_{t,t+1} - \hat{\mu^i} \right) \left(r^j_{t,t+1} - \hat{\mu^j} \right)$$

From one-period to multi-period returns in the CER

Notice that once one-step ahead returns are known also n-step ahead returns are known:

$$
\begin{aligned}
E_t(\mathbf{r}_{t,t+n}) &= n\hat{\mu} \\
Var(\mathbf{r}_{t,t+n}) &= n\hat{\Sigma}
\end{aligned}
$$

As a consequence of these properties of the data, weights in an optimal multi-horizon portfolio coincide with weights in a single period horizon portfolio:

$$
\begin{aligned}
\hat{\mathbf{w}}^T &= \frac{\Sigma^{-1}\left(\mu - r^f \mathbf{e}\right)}{\mathbf{e}'\Sigma^{-1}\left(\mu - r^f \mathbf{e}\right)}, \\
&= \frac{\Sigma^{-1}(nn^{-1})\left(\mu - r^f \mathbf{e}\right)}{\mathbf{e}'\Sigma^{-1}(nn^{-1})\left(\mu - r^f \mathbf{e}\right)}
\end{aligned}
$$

Model Simulation

- given some estimates of the unknown parameters (μ, Σ in our case), the model can be used to simulate data and predict the future given the model
- model simulation predicts the entire distribution of all the endogenous variables in the model
- we consider two different approaches to model simulation: Monte-Carlo and Bootstrap Methods

Monte-Carlo methods

- given some estimates of the unknown parameters in the model (μ, Σ in our case).

$$\begin{aligned} \mathbf{r}_{t,t+1} &= \hat{\mu} + \hat{\epsilon}_{t+1} \\ \hat{\epsilon}_{t+1} &\sim \mathcal{D}\left(\mathbf{0}, \hat{\boldsymbol{\Sigma}}\right) \end{aligned}$$

- an assumption is made on the distribution of $\hat{\epsilon}_{t+1}$
- Then an artificial sample for $\hat{\epsilon}_{t+1}$ of the desired length can be computer simulated.
- The simulated residuals are then mapped into simulated returns via the model.
- This exercise can be replicated N times to construct the distribution of model predicted returns.

Bootstrap methods

Do exactly like in Monte-Carlo but rather than using a theoretical distribution for ϵ_t use their empirical distribution (the distribution of the OLS residuals from the n regressions in the CER model) and resample from it with reimmission.

Model Validation

- Backtesting can be used to provide statistical evidence of the capability of the model to replicate the data.
- Statistical testing can be used to evaluate restrictions on coefficients predicted by the theory used to specify the empirical model.

The Cross-Section of Returns

The CER view allows for cross-sectional heterogeneity of returns but such cross-sectional heterogeneity is related to a single factor, the market factor, and the CAPM determines all the cross-sectional variation in μ^i. The statistical model that determines all returns r_t^i and the market return r_t^m,can be described as follows:

$$
\begin{aligned}
\left(r_t^i - r_t^{rf}\right) &= \mu_i + \beta_i u_{m,t} + u_{i,t} \\
\left(r_t^m - r_t^{rf}\right) &= \mu_m + u_{m,t} \\
\begin{pmatrix} u_{i,t} \\ u_{m,t} \end{pmatrix} &\sim n.i.d.\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{ii} & \sigma_{im} \\ \sigma_{im} & \sigma_{mm} \end{pmatrix}\right]
\end{aligned}
$$

where r_t^{rf} is the return onthe risk-free asset. We shall see that $\sigma_{im} = 0$ is a crucial assumption for the valid estimation of the CAPM betas, and that assumption that risk adjusted excess returns are zero (usually known as zero alpha assumption) requires that $\mu_i = \beta_i \mu_m$.

Risk Measurement

Risk of a portfolio is measured by its Value-at-Risk (VaR). The VaR is the percentage loss obtained with a probability at most of α percent:

$$\Pr\left(R^p < -VaR_\alpha\right) = \alpha.$$

where R^p are the returns on the portfolio. If the distribution of returns is normal, then α-percent VaR_α is obtained as follows (assume $\alpha \in (0,1)$):

$$\begin{aligned}\Pr\left(R^p < -VaR_\alpha\right) &= \alpha \Longleftrightarrow \Pr\left(\frac{R^p - \mu_p}{\sigma_p} < -\frac{VaR_\alpha + \mu_p}{\sigma_p}\right) = \alpha \\ &\Longleftrightarrow \Phi\left(-\frac{VaR_\alpha + \mu_p}{\sigma_p}\right) = \alpha,\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative density of a standard normal. At this point, defining $\Phi^{-1}(\cdot)$ as the inverse CDF function of a standard normal, we have that

$$-\frac{VaR_\alpha + \mu_p}{\sigma_p} = \Phi^{-1}(\alpha) \Longleftrightarrow VaR_\alpha = -\mu_p - \sigma_p \Phi^{-1}(\alpha).$$

and, given that μ_p and σ_p are constant over time, VaR_α is also constant over-time. Consider the case of the one per cent value at risk: Because $\Phi^{-1}(0.01) = -2.33$ under the normal distribution, the VaR is:

$$\widehat{VaR}_{0.01} = -\hat{\mu}_p - 2.33\hat{\sigma}_p$$

Empirical Challenges to the traditional model

Over the course of time the traditional view has been empirically challenged on many grounds. In particular it has been observed that

- The tenet that expected returns are constant is not compatible with the observed volatility of stock prices. Stock prices in fact are "too volatile" to be determined only by expected dividends ;
- there is evidence of returns predictability that increases with the horizon at which returns are defined.
- There are anomalies that make returns predictable on occasion of special events.
- The CAPM is rejected when looking at the cross-section of returns and multi factor models are needed to explain the cross-sectional variability of returns
- high frequency returns are non-normal and heteroscedastic, therefore risk is not constant over time and there is predictability of risk at high-frequency

A General Representation of Predictive Models in Finance

All predictive models we shall analyze are special cases of the following general representation:

$$\begin{aligned} \mathbf{r}_{t,t+k} &= f\left(X_t^{\mu}, \Theta_t^{\mu}\right) + \mathbf{H}_{t+k}\epsilon_{t+k} & (1) \\ \mathbf{\Sigma}_{t+k} &= \mathbf{H}_{t+k}\mathbf{H}'_{t+k}. \\ \mathbf{\Sigma}_{t+k} &= g\left(X_t^{\sigma}, \Theta_t^{\sigma}\right) + \sum\nolimits_{j=1}^{q} \mathbf{B}_j \mathbf{\Sigma}_{t+k-j} \mathbf{B}'_j, & (2) \\ \epsilon_{t+k} &\sim \mathcal{D}\left(\mathbf{0}, \mathbf{I}\right) \end{aligned}$$

where $\mathbf{r}_{t,t+k}$ is the vector of returs between time t and time t+k in which we are interested, X_t^{μ} is the vector of predictors for the mean of our returns that we observe at time t, f specifiies the functional relation (thant is potentially time-varying) between the mean returns and the predictors that depends also on a set of parameters Θ_t^{μ}, the matrix $\mathbf{H}_{t+k}$ determines the potentially time varying variance-covariance of the vector of returns..The process for the variance is predictable as there is a functional relation determining the relationship between $\mathbf{H}_{t+k}$ and a vector of predictors X_t^{σ} that is driven by a vector of unknown parameters Θ_t^{σ}.

Predictive Models for Returns at High-Frequency

Consider, for example, the problem of univariate modelling of stock market returns. When k is small and high-frequency returns On the one hand, in the simple asset allocation model, the econometric framework considered for returns is as follows:

$$\begin{aligned} r_{t,t+k} &= 0 + \sigma_{t+k} u_{t+k} \qquad u_{t+k} \sim IID\ \mathcal{D}(0,1), \\ \sigma^2_{t+k} &= \omega + \alpha \sigma^2_{t+k-1} + \beta u^2_{t+k-1}, \quad |\alpha + \beta| < 1 \end{aligned}$$

This is a model that feautres no predictability in the mean of r returns (the expected future return at any horizon is constant at zero), but there is predictability in the variance of returns that it is mean reverting towards a long-term value of $\omega / (1 - \alpha - \beta)$. No assumption of normality is made for the innovation innovation in the process generating returns.

Predictive Models for Returns at High-Frequency

Consider now the case of large k, i.e. long-horizon returns (note that in the continuously compounded case, $r_{t,t+k} \equiv \sum_{j=1}^{k} r_{t,t+j}$), in this case the relevant predictive model can be written as follows:

$$r_{t,t+k} = \alpha + \beta' \mathbf{X}_t + \sigma u_{t+k} \qquad u_{t+k} \sim IID\ \mathcal{N}(0,1),$$

where $\mathbf{X}_t$ is a set of predictors observed at time t. In this case we have that returns feature predictability in mean, constant variance and the innovations are normally distributed. As the horizon k increases, predictability increases and therefore the uncertainty related to the unexpected components of returns decreases (i.e., the annualized variance of returns is a downward sloping function of the horizon). Moreover—as we have already discussed—the dependence of $\sigma_{t,k}$ on time (i.e., its time-varying nature) declines and long-horizon returns can be described as a (conditional) normal homoskedastic processes. In the short-run noise dominates and modelling returns on the basis of fundamentals is very difficult. However, as the horizon increases fundamentals become more important to explain returns and the risk associated to portfolio allocation based on econometric models is reduced. The statistical model becomes more and more precise as k gets large.