4 The Predictability of Asset Returns: Part I

4.1 The Efficient Market Hypothesis (EMH)

Asset price predictability is perhaps the earliest and most discussed topic in financial econometrics. This topic is very important since it could reveal information on how the financial market works, and could lead to potential profits. Of course, the predictability depends on what information the person who carries out the prediction has. Someone who knows a company well may give a preciser forecast to the future price of the company's stock than someone who has never heard of the company. If the person can make a preciser prediction than the market average, there is a chance of superior profits. It could also be possible that the price "reveals" or "represents" information so well that no one is able to make a prediction that is more "correct" than others and therefore generate superior profits. At least starting from the very beginning of the 20th century, people began to think about the problem of market efficiency.

- Samuelson (1965): "In competitive markets there is a buyer for every seller. If one could be sure that a price will rise, it would have already risen."

  "...This means that there is no way of making an expected profit by extrapolating past changes in the futures prices, by chart or any other esoteric devices of magic or mathematics. The market quotation $Y(T, t)$ already contains in itself all that can be known about the future and in that sense has discounted future contingencies as much as is humanly possible."

- Fama (1970): "...In general terms, the ideal is a market in which prices provide accurate signals for resource allocation: that is, a market in which firms can make production-investment decisions, and investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at any time 'fully reflect' all available information. A market in which prices always fully reflect' available information is called 'efficient'."

- Malkiel (1992): "...Formally, the market is said to be efficient with respect to some information set... if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set... implies that it is impossible to make economic profits by trading on the basis of [that information set]."

- Malkiel (2003): "It was generally believed that securities markets were extremely efficient in reflecting information about individual stocks and about the stock market as a whole. The accepted view was that when information arises, the news spreads very quickly and is incorporated into the prices of securities without delay. Thus, neither technical analysis, which is the study of past stock prices in an attempt to predict future prices, nor even fundamental analysis, which is the analysis of financial information such as company earnings and asset values to help investors select undervalued stocks, would enable an investor to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks, at least not with comparable risk."

  "...I will use as a definition of efficient financial markets that such markets do not allow investors to earn above-average returns without accepting above-average risks."

When one talks about market efficiency, one has to be clear about what information set he or she has in mind. In the finance literature, there are three commonly used information sets, and three corresponding market efficiency.

- Week form market efficiency: the market is efficient with respect to the history of prices/returns.
- Semi-strong form market efficiency: the market is efficient with respect to all publicly available information.

- Strong form market efficiency: the market is efficient with respect to all publicly available and private information.

Note that the meaning of "efficiency" here is not the same as the meaning of "efficiency" in Welfare Economics. A consequence of market efficiency with respect to some information set is that given that information set, there is no chance that one could make a superior return without knowing more. As a consequence, one way to test market efficiency is to see whether one can find a trading strategy that could generate superior profits.

However, we make some remarks here. First, inefficiency does not necessarily lead to superior profitability; There are frictions in the financial markets that could prevent profitability that comes from market inefficiency. For example, it could be that the superior returns are too small to offset the transaction costs.

Second, when one is talking about superior returns, there must be some "normal" return that plays the role of a benchmark. One has to be careful when defining what the normal returns are. For example, the normal return for different risks are different.

In practice, people usually test the weak form of EMH by looking at whether one can generate superior returns by studying past stock prices. People test the semi-strong form of EMH by looking at whether one can generate superior returns by studying past stock prices as well as company earnings and asset values. People test strong form EMH by looking at investment performances of professional such as common fund managers, assuming that these professionals have more information than publicly available.

4.2 Return Predictability

In this section we try to formulate the market efficiency hypothesis. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space and $\mathcal{G}_t, \mathcal{H}_t$ be sub-σ-algebras (of $\mathcal{F}$) representing different information sets at time t. Suppose the equilibrium price of an asset, generated from a data generating process(DGP), is P_t^* at time t. If one possess information $\mathcal{G}_t$ at time t, (and if (s)he correctly understands the true DGP of the equilibrium price), his(her) best guess for the equilibrium price (or the true value of the asset) is $\mathbb{E}(P_t^*|\mathcal{G}_t)$.

Let the actual market price be P_t. If his(her) guess is smaller than the actual price, then (s)he would believe that the asset is overpriced. If his(her) guess is greater than the actual price, (s)he would believe that the asset is underpriced. In either case, (s)he would want to arbitrage by longing or shorting the asset. The market is efficient with respect to the information set $\mathcal{G}_t$ means that based on the information set and the above calculation, any investor is not able to make a profit based on the difference between the conditional distribution of P_t given $\mathcal{G}_t$ and the distribution of $\mathbb{E}(P_t^*|\mathcal{G}_t)$.

Here we consider a necessary condition: if the market is efficient with respect to the information set $\mathcal{G}_t$, then any investor is not able to make a profit based on the difference between the market price P_t and his(her) best guess of the equilibrium price $\mathbb{E}(P_t^*|\mathcal{G}_t)$ on average. That is,

$$\mathbb{E}\left(P_t - \mathbb{E}(P_t^*|\mathcal{G}_t)\middle|\mathcal{G}_t\right) = 0.$$

EMH formulated in this way satisfies the following properties.

- If the actual price is the equilibrium price, the market is efficient with respect to any information set: $\mathbb{E}\left(P_t^* - \mathbb{E}(P_t^*|\mathcal{G}_t)|\mathcal{G}_t\right) = 0$ for any $\mathcal{G}_t$.

- If the actual price "fully reflects" the information set $\mathcal{G}_t$, i.e., $P_t = \mathbb{E}(P_t^*|\mathcal{G}_t)$, the market is efficient with respect to $\mathcal{G}_t$:

$$\mathbb{E}\left(\mathbb{E}(P_t^*|\mathcal{G}_t) - \mathbb{E}(P_t^*|\mathcal{G}_t)\middle|\mathcal{G}_t\right) = 0.$$

- If the market is efficient with respect to $\mathcal{G}_t$, then it is efficient with respect to a smaller information set $\mathcal{H}_t \subset \mathcal{G}_t$. That is, if $\mathbb{E}\left(P_t - \mathbb{E}(P_t^*|\mathcal{G}_t)|\mathcal{G}_t\right) = 0$, then

$$\mathbb{E}\left(P_t - \mathbb{E}(P_t^*|\mathcal{H}_t)\middle|\mathcal{H}_t\right) = \mathbb{E}\left(\mathbb{E}\left(P_t - \mathbb{E}(P_t^*|\mathcal{G}_t)\big|\mathcal{G}_t\right)\middle|\mathcal{H}_t\right) = 0.$$

This means that if the market is efficient with respect to the information set $\mathcal{G}_t$ and you have less information than $\mathcal{G}_t$, then you are not able to make a profit on average based on your limited information.

- If the market is efficient with respect to $\mathcal{G}_t$, then it may be not efficient with respect to a larger information set $\mathcal{H}_t \supset \mathcal{G}_t$. For example, if the actual price only "reflect" the information $\mathcal{G}_t$, i.e., $P_t = \mathbb{E}(P_t^*|\mathcal{G}_t)$, then

$$\mathbb{E}\left(P_t - \mathbb{E}(P_t^*|\mathcal{H}_t)\middle|\mathcal{H}_t\right) = \mathbb{E}(P_t^*|\mathcal{G}_t) - \mathbb{E}(P_t^*|\mathcal{H}_t)$$

  in general is not zero (for example, if P_t^* is $\mathcal{G}_t$-measurable but not $\mathcal{H}_t$-measurable). This means that if the market is efficient with respect to the information set $\mathcal{G}_t$ and you hold extra information, then it is possible that you make a profit on average out of the extra information you hold.

Here we make some remarks. First, testing the market efficiency relies on a model for $\mathbb{E}(P_t^*|\mathcal{G}_t)$. Therefore, what is tested is a mixture of the EMH, the model, and the equilibrium. Second, the zero (conditional) *expected* profit condition is only one implication of the efficient market hypothesis. The EMH could have other implications on the (conditional) distribution of the superior returns. Third, the above formulation could also be applied essentially in the same way to the logged prices $\ln P_t$, the simple return R_t, or the log return r_t.

At this moment, we only consider the case in which the information set $\mathcal{G}_t$ consists of only the past prices (or returns if we choose to model the returns). That is, we are considering the test of the weak form of EMH. In a discrete time setting, this means that $\mathcal{G}_t = \sigma(P_{t-1}, P_{t-2}, \ldots)$. For notation convenience, since $\sigma(P_{t-1}, P_{t-2}, \ldots)$ involves only prices up to time $t-1$, we write $\sigma(P_t, P_{t-1}, \ldots) = \mathcal{F}_{t-1}$.

4.3 The Martingale Hypothesis

If our model for the conditional mean asserts that

$$\mathbb{E}(P_t^* | \mathcal{F}_{t-1}) = P_{t-1},$$

then the EMH can be formulated as

$$\mathbb{E}(P_t - P_{t-1} | \mathcal{F}_{t-1}) = 0.$$

It follows that

$$\mathbb{E}(P_t | \mathcal{F}_{t-1}) = P_{t-1}.$$

The above equation shows that the sequence of prices $\{P_t\}$ is a martingale with respect to the *filtration* $\{\mathcal{F}_t\}$ (an increasing sequence of σ-algebras).

Definition 4.1. Let $X_1, X_2, \dots$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{F}_1, \mathcal{F}_2, \dots$ be a sequence of sub-σ-algebras such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all t. If X_n is $\mathcal{F}_n$-measurable, $\mathbb{E}\,|X_n| < \infty$, and

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = X_{t-1} \quad \text{a.s.},$$

then the sequence $\{(X_t, \mathcal{F}_t)\}$ is called a martingale. We may also say that $\{X_t\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t\}$. When the underlying filtration is clear from the context, we may simply say that $\{X_t\}$ is a martingale.

If we replace the equality in the above definition by

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) > X_{t-1} \quad \text{a.s.},$$

then $\{(X_t, \mathcal{F}_t)\}$ is called a sub-martingale. If the equality is replaced by

$$\mathbb{E}(X_t|\mathcal{F}_{t-1}) < X_{t-1} \quad \text{a.s.,}$$

then $\{(X_t, \mathcal{F}_t)\}$ is called a super-martingale.

Suppose that $\{Y_t\}$ is a sequence of random variables and $\{\mathcal{F}_t\}$ is an increasing sequence of sub-σ-algebras. Then $\{(Y_t, \mathcal{F}_t)\}$ is called a martingale difference sequence if

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = 0 \quad \text{a.s..}$$

A martingale difference sequence $\{(Y_t, \mathcal{F}_t)\}$ may be viewed as the sequence of the differences of some martingale, as its name suggests. To see this, Let $X_t = Y_0 + Y_1 + \ldots + Y_t$. It is easy to see that $Y_t = X_t - X_{t-1}$ and $\{(X_t, \mathcal{F}_t)\}$ is a martingale.

Now according to our model for the conditional mean of the asset value, the EMH is equivalent to that $\{(P_t, \mathcal{F}_t)\}$ is a martingale. Or we may write $P_t = P_{t-1} + \varepsilon_t$, then $\{(\varepsilon_t, \mathcal{F}_t)\}$ is a martingale difference sequence. We may show that

$$\mathbb{E}P_t = \mathbb{E}P_{t-1} = \ldots \mathbb{E}P_0$$

and

$$\mathbb{E}\varepsilon_t = \mathbb{E}\varepsilon_{t-1} = \ldots \mathbb{E}\varepsilon_0 = 0.$$

4.4 The Variance Ratio Tests

4.4.1 The Basic Test

A popular procedure to test the martingale hypothesis is first proposed by Lo and MacKinlay (1999, Chapter 2). The test is proposed under a stronger hypothesis:

$$H_0: \quad \varepsilon_t \sim \text{iid } (0, \sigma^2).$$

The test is constructed as follows. Suppose that we obtain $2n+1$ observations $P_0, P_1, \ldots, P_{2n}$. We may obtain two variances

$$\hat{\sigma}_a^2 = \frac{1}{2n} \sum_{t=1}^{2n} (P_t - P_{t-1})^2$$

and

$$\hat{\sigma}_b^2 = \frac{1}{2n} \sum_{k=1}^{n} (P_{2k} - P_{2k-2})^2.$$

We construct two test statistics:

$$J_d = \hat{\sigma}_b^2 - \hat{\sigma}_a^2$$

and

$$J_r = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2} - 1.$$

We have the following result.

Theorem 4.2. *Under the null hypothesis H_0, the asymptotic distributions of J_d and J_r are given by*

$$\sqrt{n} J_d \to_d \mathbb{N}(0, \sigma^4)$$

and

$$\sqrt{n} J_r \to_d \mathbb{N}(0, 1)$$

as $n \to \infty$.

Note that the parameter σ^4, which is unknown, appears in the asymptotic distribution of J_d. This is called a nuisance parameter. The asymptotic distribution of J_r is free of nuisance parameters and therefore J_r is usually preferred.

Proof. Note that both test statistics are functions of $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$. So we first obtain the asymptotic distribution of $(\hat{\sigma}_a^2, \hat{\sigma}_b^2)'$ jointly. Write

$$\hat{\sigma}_a^2 = \frac{1}{2n}\sum_{k=1}^{n}(\varepsilon_{2k-1}^2 + \varepsilon_{2k}^2),$$

and

$$\hat{\sigma}_b^2 = \frac{1}{2n}\sum_{k=1}^{n}(\varepsilon_{2k-1} + \varepsilon_{2k})^2.$$

Then

$$\begin{bmatrix} \hat{\sigma}_a^2 \\ \hat{\sigma}_b^2 \end{bmatrix} = \frac{1}{2n}\sum_{k=1}^{n}\begin{bmatrix} \varepsilon_{2k-1}^2 + \varepsilon_{2k}^2 \\ (\varepsilon_{2k-1} + \varepsilon_{2k})^2 \end{bmatrix}.$$

Since $\{\varepsilon_t\}$ is an iid sequence, $\{(\varepsilon_{2k-1}^2 + \varepsilon_{2k}^2, (\varepsilon_{2k-1} + \varepsilon_{2k})^2)'\}$ is an iid sequence of random vectors. We therefore may use the central limit theorem for an iid sequence to obtain the asymptotic distribution of $(\hat{\sigma}_a^2, \hat{\sigma}_b^2)'$.

It is easy to see that

$$\mathbb{E}\begin{bmatrix} \varepsilon_{2k-1}^2 + \varepsilon_{2k}^2 \\ (\varepsilon_{2k-1} + \varepsilon_{2k})^2 \end{bmatrix} = \begin{bmatrix} 2\sigma^2 \\ 2\sigma^2 \end{bmatrix}.$$

After some tedious calculations, one may show that

$$\operatorname{Var}\left(\begin{bmatrix} \varepsilon_{2k-1}^2 + \varepsilon_{2k}^2 \\ (\varepsilon_{2k-1} + \varepsilon_{2k})^2 \end{bmatrix}\right) = \begin{bmatrix} 2A - 2\sigma^4 & 2A - 2\sigma^4 \\ 2A - 2\sigma^4 & 2A + 2\sigma^4 \end{bmatrix},$$

where $A = \mathbb{E}\varepsilon_1^4$. Therefore, by CLT,

$$\sqrt{n}\left(\begin{bmatrix} \hat{\sigma}_a^2 \\ \hat{\sigma}_b^2 \end{bmatrix} - \begin{bmatrix} \sigma^2 \\ \sigma^2 \end{bmatrix}\right) \to_d \mathbb{N}\left(0, \begin{bmatrix} \frac{A-\sigma^4}{2} & \frac{A-\sigma^4}{2} \\ \frac{A-\sigma^4}{2} & \frac{A+\sigma^4}{2} \end{bmatrix}\right).$$

Now we may write $J_d = h_1(\hat{\sigma}_a^2, \hat{\sigma}_b^2)$ and $J_r = h_2(\hat{\sigma}_a^2, \hat{\sigma}_b^2)$ where $h_1(x_1, x_2) = x_2 - x_1$ and $h_2(x_1, x_2) = \frac{x_2}{x_1} - 1$. We introduce the so called Delta-method: Suppose that $\hat{\beta}$ is an estimator for $\beta \in \mathbb{R}^n$ and

$$\sqrt{n}(\hat{\beta} - \beta) \to_d \mathbb{N}(0, D).$$

If $h : \mathbb{R}^n \to \mathbb{R}$ is a function that is differentiable at β, then

$$\sqrt{n}(h(\hat{\beta}) - h(\beta)) \to_d \mathbb{N}(0, \nabla h^T(\beta) D \nabla h(\beta)).$$

Now we apply the Delta-method to $J_d = h_1(\hat{\sigma}_a^2, \hat{\sigma}_b^2)$ and $J_r = h_2(\hat{\sigma}_a^2, \hat{\sigma}_b^2)$. We have

$$\sqrt{n}(J_d - 0) \to_d \mathbb{N}\left(0, \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{A-\sigma^4}{2} & \frac{A-\sigma^4}{2} \\ \frac{A-\sigma^4}{2} & \frac{A+\sigma^4}{2} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \mathbb{N}(0, \sigma^4),$$

$$\sqrt{n}(J_r - 0) \to_d \mathbb{N}\left(0, \begin{bmatrix} -\frac{1}{\sigma^2} & \frac{1}{\sigma^2} \end{bmatrix} \begin{bmatrix} \frac{A-\sigma^4}{2} & \frac{A-\sigma^4}{2} \\ \frac{A-\sigma^4}{2} & \frac{A+\sigma^4}{2} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sigma^2} \\ \frac{1}{\sigma^2} \end{bmatrix}\right) = \mathbb{N}(0, 1).$$

$\square$

Now we know that ex ante J_r is distributed normally with mean zero and variance one when n is large. (The statistic J_r is a random variable!) Once we observe a particular realization of the sequence of prices $\{P_t\}$, we may use the observed value to calculate the realized value of J_r, say c. (The realized value is a real number!) Now we may calculate the (asymptotic) probability $\mathbb{P}(|\sqrt{n} J_r| \geq c)$ of getting a realization of J_r that is equal to or larger than c in absolute value. This probability is the p-value of the test for the observed data.

4.4.2 The VR test in the Presence of a Drift Term

In this section we consider the case when $\mathbb{E}(P_t^* | \mathcal{F}_{t-1}) = \mu + P_{t-1}$, i.e., when the price follows

$$P_t = \mu + P_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ and μ is a constant. This formulation for stock price is more practical since the additional drift term μ may reflect the effect from inflation. In order to conduct a variance ratio test for ε_t, we need to first estimate μ. To estimate μ, we notice that

$$P_t - P_{t-1} = \mu + \varepsilon_t$$

and

$$\mathbb{E}\varepsilon_t = 0.$$

Therefore, a good estimator of μ is given by

$$\hat{\mu} = \frac{1}{2n}\sum_{t=1}^{2n}(P_t - P_{t-1}) = \frac{1}{2n}(P_{2n} - P_0).$$

Now we may redefine $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ respectively as

$$\hat{\sigma}_a^2 = \frac{1}{2n}\sum_{t=1}^{2n}(P_t - P_{t-1} - \hat{\mu})^2,$$

and

$$\hat{\sigma}_b^2 = \frac{1}{2n}\sum_{k=1}^{n}(P_{2k} - P_{2k-2} - 2\hat{\mu})^2.$$

Now we define J_d and J_r as in the previous section. It can be shown that the asymptotic distribution of the two statistics remain unchanged. The intuition behind is that we may write

$$\hat{\sigma}_a^2 = \frac{1}{2n}\sum_{t=1}^{2n}(P_t - P_{t-1} - \mu + (\mu - \hat{\mu}))^2 = \frac{1}{2n}\sum_{t=1}^{2n}(\varepsilon_t + (\mu - \hat{\mu}))^2$$

and

$$\hat{\sigma}_b^2 = \frac{1}{2n}\sum_{k=1}^{n}(P_{2k} - P_{2k-2} - 2\mu + 2(\mu - \hat{\mu}))^2 = \frac{1}{2n}\sum_{k=1}^{n}(\varepsilon_{2k+1} + \varepsilon_{2k} + 2(\mu - \hat{\mu}))^2.$$

Since $\hat{\mu}$ is a very good estimator of μ, i.e., $\mu - \hat{\mu}$ is small enough, the introduction of the term $\mu - \hat{\mu}$ does not affect the asymptotic distributions of $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$.

4.4.3 A General VR test

When we construct $\hat{\sigma}_b^2$, instead of using the differences of every other observations, we may use the differences of every qth observations. Suppose that we have $nq+1$ observations $P_0, P_1, \ldots, P_{nq}$. Suppose that our null hypothesis is that $P_t = \mu + P_{t-1} + \varepsilon_t$ and $\varepsilon_t \sim \mathrm{iid}(0, \sigma^2)$. We define

$$\hat{\mu} = \frac{1}{nq}(P_{nq} - P_0),$$

$$\hat{\sigma}_a^2 = \frac{1}{nq}\sum_{t=1}^{nq}(P_t - P_{t-1} - \hat{\mu})^2,$$

$$\hat{\sigma}_b^2(q) = \frac{1}{nq}\sum_{k=1}^{n}(P_{qk} - P_{q(k-1)} - q\hat{\mu})^2.$$

Just as before, we define

$$J_d(q) = \hat{\sigma}_b^2(q) - \hat{\sigma}_a^2$$

and

$$J_r(q) = \frac{\hat{\sigma}_b^2(q)}{\hat{\sigma}_a^2} - 1.$$

Using the same approach, we may show that under the null hypothesis, the asymptotic distributions of the two statistics are given respectively by

$$\sqrt{n}J_d(q) \to_d \mathbb{N}\left(0, \frac{2(q-1)}{q}\sigma^4\right)$$

and

$$\sqrt{n}J_r(q) \to_d \mathbb{N}\left(0, \frac{2(q-1)}{q}\right).$$