Empirical Finance: A Review

For Personal Reference

Sai Zhang

Inspired by the Course ***Empirical Finance*** at
London Business School by *Dr. Svetlana Bryzgalova*

April 26, 2021

HERE WE GO!

Empirical finance is an absolutely fascinating field, with some of the most cutting-edge methodologies and the most exploratory techniques. Although it is not my speciality, I am always interested in this literature. During my pre-doc research fellowship at London Business School, I have had the previlege to study in the course ***Financial Economics II: Empirical Finance***. The course instructor Dr. Svetlana Bryzgalova is absolutely one of the most brilliant scholars I have encountered. Thanks to her, I have got to understand this liturature more systematically. In this (personal) review, I summarize the most influential and inspirational works in this field and organize them by different topics. The structure of this review resembles the structure of Dr. Bryzgalova's course, while adjusted according to my personal research interest. I intend to review classic works and discuss some potential directions of future study regarding my personal interest in Behavioral Economics, Game Theory and Network.

Since this review is tailored according to my own research interest and experience, I will not only summarize the theoretical perspectives of the studies, present their findings and discuss how they fit into the literature, but document my replication attempts and pseudo codes as well. All the codes related to this review can be found on my Github page.

I thank Dr. Svetlana Bryzgalova for her valuable intuitions and impressive knowledge of the empirical finance literature. Building this review is truly a memorable journey for me. I would love to share this review and all the related materials to anyone that finds them useful. And unavoidably, I would make some typos and other minor mistakes (hopefully not big ones). So I'd really appreciate any correction. If you find any mistakes, please either set up a branch on Github or send the mistakes to this email address saizhang.econ@gmail.com, BIG thanks in advance!

Contents

CHAPTER 1

TIME-SERIES PREDICTABILITY

Contents

Every investor knows that trading in financial markets is to play games with time itself. Daily trades determine asset prices at every date and hence influence the random distribution of future prices as well as the initial level of prices. One would need "much more careful attention to the process by which both expected payoffs and required rates of return determine asset prices"[1].

[1]See **?**, p. 121

In this chapter, I first, following **?**, Chapter 5, summarize models mapping cash flows and discount rates into prices using present value relations in Section **??**. Then I discuss the early evidence for mean reversion in returns in Section **??**. In Section **??**, I examine the excess volatility puzzle in the predictability debate. To accomodate the stylized facts of time-series predictability, Section **??** presents two of the most influential approaches to decompose prices. In Section **??**, I selectively summarize some researches from the so-called "Prediction Zoo", which satirically describes the floods of price predictors. Finally, I discuss the issues and extensions of time-series predictability in Section **??**.

1.1 Concepts and models

In this section, I follow **?**, Chapter 5 and discuss some of the conceptual building blocks for the strand of time-series empirical finance literature.

1.1.1 Market Efficiency

An intuitive way of explaining ***market efficiency*** is that efficient markets are competitive and allow no easy ways to make economic profit. A more useful and testable definition was given by **?**, p. 127:

> The market is said to be efficient with respect to some information set ϕ, if security prices would be unaffected by revealing that information to all participants.

Some event studies that measure market responses to news announcements can be interpreted as tests of market efficiency regarding the announced information, but in general, this definition is not easy to test. On the other hand, **?** gives a more testable alternative:

> Efficiency with respect to an information set ϕ implies that it is impossible to make economic profits by trading on the basis of ϕ.

This is the idea behind an enormous literature in empirical asset pricing: if an economic model defines the equilibrium return as $\Theta_{i,t}$, then the null hypothesis is

$$R_{i,t+1} = \Theta_{i,t} + U_{i,t+1} \tag{1.1}$$

where $U_{i,t+1}$ is a FAIR game regarding the information set at t, or $\mathbb{E}(U_{i,t+1}|\phi_t) = 0$. Notice that market efficiency is equivalent to rational expectations, one must text a model of expected returns as well when testing market efficiency. After defining a model of expected returns, the variables to be included in the information set must be specified. **?** define three forms of efficient market hypothesis and the corresponding information sets:

- the ***weak form***: past returns

- the ***semi-strong form***: publicly available information such as stock splits, dividends, or earnings
- the ***strong form***: information available to some market participants, but NOT necessarily to all participants.

← this co tested by measurea actions (tr or portfol holdings) potentiall better info agents

In the time-series literature, the simplest economic model is constant return: $\Theta_{i,t} = \Theta$. In Section **??**, I summarize the early literature focusing on this model.

Market efficiency has been widely tested and debated, now the most accepted view of market efficiency hypothesis is that it is a useful benchmark but does not hold perfectly. The debates between long-term versus short-term efficiency, micro versus macro efficiency are still and will continue to be heated. Some noticable alternative hypotheses are:

- ***High-frequency noise***: market prices are contaminated by short-term noise, which can be caused by measurement errors or illiquidity (bid-ask bounce).
- ***Inperfect information processing***: the market reacts sluggishly to information after its releasing
- ***Persistent mispricing***: market prices deviate substantially from efficient levels in a LONG time
- ***Disposition effect***: individual investors are more willing to sell winning stocks then losing stocks, see **?** for details.

1.1.2 Model: autocorrelation of returns

The most basic time-series test of market efficiency is to test "whether past deviations of returns from model-implied expected returns predict future return deviations" (See **?**, p. 124). The leading approach to do so is to test the autocorrelations.

Starting points:

1. The null hypothesis H_0: the stock returns are i.i.d.
2. The standard error for any single sample autocorrelation equals asymptotically $1/\sqrt{T}$, see **?** for a detailed discussion.
3. The standard error would be large, (0.1 if $T = 100$), not so easy to detect small autocorrelation

Any autocorrelation test would have to solve these issues.

1.1.2.1 Q-statistics

Because the stock returns are i.i.d. (H_0), different autocorrelations are uncorrelated with one another. **?** calculates a sum of K squared sample autocorrelations:

$$Q_K = T \sum_{j=1}^{K} \hat{\rho}_j^2 \tag{1.2}$$

where $\hat{\rho}_j = C\hat{o}rr(r_t, r_{t-j})$. Q is asymptotically distributed χ^2 with K degrees of freedom.

Pros: It solves the problem of the large standard errors.

Cons: It does NOT use the sign of the autocorrelations (squared). What could happen is that the expected reutrns are not constant, instead, they are each individually small but all have the same sign.

1.1.2.2 Variance ratio

One way to take the sign of autocorrelations into consideration is the variance ratio statistic. This statistic was introduced to the finance literature by **?** and **?**.

The basic setting is: for a holding period K, the log return of this entire period $r_t(K)$ is the sum of all the one-period returns r_{t+i}:

$$r_t(K) \equiv r_t + r_{t+1} + \cdots + r_{t+K-1}$$

and the variance ratio over the period K would be defined as:

$$VR(K) = \frac{Var(r_t(K))}{K \cdot Var(r_t)}$$

If there are not autocorrelations, then the i.i.d. returns would have identical variance in each period from t to $t+K$, and $Var(r_t(K)) = Var(r_t + \cdots + r_{t+K-1}) = Var(r_t) + \cdots + Var(r_{t+K-1}) = K \cdot Var(r_t)$. Thus, $VR(K) = 1$. If we rewrite the definition of the variance ratio as:

$$VR(K) = \frac{Var(r_t(K))}{K \cdot Var(r_t)} = 1 + \underbrace{2\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)\hat{\rho}_j}_{\text{weighted average of the first } K-1 \text{ sample autocorrelations}} \tag{1.3}$$

Then by comparing $VR(K)$ with 1, we can deduct the direction of the autocorrelations:

$VR(K) > 1$	predominantly **positive** autocorrelations
$VR(K) = 1$	no autocorrelations
$VR(K) < 1$	predominantly **negative** autocorrelations: mean reversion

Notice that the weight term $1 - \frac{j}{K}$ increases as j approaches K[2].

The asymptotic variance of the variance-ratio statistic, under H_0 (i.i.d. returns), is:

$$Var(\hat{VR}(K)) = \frac{4}{T}\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)^2 = \frac{2(2K-1)(K-1)}{3KT} \xrightarrow{K\to\infty} \frac{4K}{3T} \tag{1.4}$$

[2]**?** showed that the estimator of VR(K) can be interpreted in terms of the frequency domain. It is asymptotically equivalent to 2π times the normalized spectral density estimator at the zero frequency, which uses the Bartlett kernel.

When $K \rightarrow \infty, T \rightarrow \infty$, and $K/T \rightarrow 0$ (**?**, p. 463), the true return process can be serially correlated and heteroskedastic, but the variance of the variance-ratio is still given as:

$$Var(\hat{VR}(K)) = \frac{4K}{3T} \cdot VR(K)^2 \tag{1.5}$$

Notice that this can be quite large with a large $VR(K)$. This is due to the fact that $K/T \rightarrow 0$ is a dangerous assumption because in practice K is often large relative to the sample size. To tackle this, **?** develop alternative asymptotics assuming $K/T \rightarrow \delta$ where $\delta > 0$. Through Monte Carlo simulations, they demonstrated that this new distribution is a more robust approximation to the small-sample distribution of the VR statistic. Most current applications of the VR statistic cite $K/T \rightarrow \delta > 0$ as the justification for using Monte Carlo distributions (i.e. set at $K = \delta T$) as representative of the VR statistic's sampling distribution. Some recent challenges of this result are discussed in Section **??**.

To accommodate r_t's exhibiting conditional heteroskedasticity, **?** proposed a heteroskedasticity-robust variance estimation of VR(K) as:

$$Var^*(\hat{VR}(K)) = 4\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)^2 \cdot \frac{\sum_{t=j+1}^{T}(r_t-\bar{r})^2(r_{t-j}-\bar{r})^2}{\left[\sum_{t=1}^{T}(r_t-\bar{r})^2\right]^2}$$

where $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$ is the estimated mean of returns.

1.1.2.3 Regression approach

? established a regression approach to test AR(K). The basic idea is to regress the K-period return on the lagged K-period return:

$$r_t(K) = \alpha_K + \beta_K r_{t-K}(K) + \epsilon_t^K$$

The coefficient β_K would then be:

$$\beta_K = \frac{Cov[r_t(K), r_{t-K}(K)]}{Var[r_{t-K}(K)]} = 2\left[\frac{VR(2K)}{VR(K)} - 1\right] = \frac{2\sum_{j=1}^{K-1}\left(\frac{\min(j,2K-j)}{K}\right)\rho_j}{VR(K)} \tag{1.6}$$

It is clear to see that:

$\beta_K > 0$	predominantly **positive** autocorrelations
$\beta_K = 0$	no autocorrelations
$\beta_K < 0$	predominantly **negative** autocorrelations: mean reversion

1.1.3 Extension: Other variance-ratio tests

As summarized by **?**, the intuition behind the VR test is rather simple, but conducting a statistical inference using the VR test is less straightforward. In this bonus subsection, I briefly summarize some recent development of individual VR tests, multiple VR tests and bootstrapping VR tests. For more detailed discussion, see **?** for a review.

1.1.3.1 Individual VR tests

Conventional VR tests, such as the **?** test, are asymptotic tests: their sampling distributions are approximated by their limiting distributions. In practice, the asymptotic theory provides a poor approximation to the small-sample distribution of the VR statistic, which impeded the use of the statistic. In general, the ability of the asymptotic distribution to approximate the finite-sample distribution depends crucially on the value of K. For a large K relative to T, **?** have proved that the VR statistics are severely biased and right skewed. Several alternative tests try to tackle this issue.

? Test: they suggested a simple power transformtaion of the VR statistic when K is NOT too large. This transformation is able to solve the right-skewness problem and robust to conditional heteroskedasticity. They showed that the transformed VR statistic leads to significant gain in power against mean reverting alternative. They define the VR statistic based on the periodogram and this new statistic is precisely the normalized discrete periodogram average estimate of the spectral density of a stationary process at the origin.

? Test: they proposed a non-parametric alternative using signs and ranks. This test outperforms the **?** test in 2 ways:

(1) As the rank and sign tests have an exact sampling distribution, there is no need to resort to asymptotic distribution approximation.
(2) The tests may be more powerful against a wide range of models displaying serial correlation, including fractionally integrated alternatives.

The rank-based tests display low-size distortions under conditional heteroskedasticity. One thing to notice is that the sign test assumes a zero drift value, **?** extended this test with unknown drift.

? Test: To overcome the issue that the arbitrary and *ad hoc* choice of K, **?** proposed a data-dependent procedure to determine the optimal value of K. This test is based on frequency domain following **?**. However, instead of using Bartlett kernel as **?**, **?** employed the quadratic spectral kernel since it's optimal in estimating the spectral density at the zero frequency.

1.1.3.2 Multiple VR tests

All the tests above are individual tests, where the null hypothesis is tested for an individual value of K. However, to determine whether a time series is a random walk, we need to rule out all possibilities, meaning that for all values of K, the null hypothesis can not be rejected. It is necessary to conduct a joint test where a multiple comparison of VRs over a set of different time horizons is made. However, conducting separate individual tests for a number of K values may lead to over rejection of the null hypothesis. Several tests have been developed for this problem, with the joint null hypothesis $H_0 : \forall K_i, V(K_i) = 1$ against the alternative $H_1 : \exists K_i, V(K_i) \neq 1$

? Test: This test statistic is defined as

$$MVR_1 = \sqrt{T} \max_{1 \leq i \leq m} |M_1(K_i)|$$

where $M_1(K_i) = \frac{VR(K_i)-1}{\sqrt{2(2K-1)(K-1)/3KT}}$. This is based on the idea that the decision regarding the null hypothesis can be obtained from the maximum absolute value of the individual VR statistics. Then, they applied the Sidak probability inequality and give an upper bound to the critical values taken in the studentized maximum modulus (SMM) distribution. The statistic follows the SMM distribution with m and T degrees of freedom, where m is the number of K values. To accommodate heteroskedasticity, one can change $M_1(K_i)$ into a heteroskedasticity-robust individual VR test.

? Test: They use a subsampling technique to develop a multiple VR test. When sample size (T) is relatively small, this test outperforms the conventional VR tests, and shows little to no serious size distortions. The statistic is:

$$MVR_T = \sqrt{T} \max_{1\leq i\leq m} |VR(K_i) - 1|$$

and the sampling distribution function for the MVR_T statistic is asymptotically a maximum of a multivariate normal vector with an unknown covariance matrix, which would be complicated to estimate. Therefore, they proposed to approximate the null distribution by means of the subsampling approach. For a subsample of size b: $(x_t, \cdots, x_{t-b+1})$ where $t = 1, \cdots, T - b + 1$. The statistic MVRs calculated from all individual subsamples would generate a $(1 - \alpha)$th percentile for the $100(1 - \alpha)\%$ critical value. To implement this subsampling technique, a choice of block length b must be made. **?** recommended the interval of $(2.5T^{0.3}, 3.5T^{0.6})$, but they also found that the size and power properties of their test are not sensitive to b.

? Test: This is a multiple rank and sign VR tests, an extension to the **?**'s rank- and sign-based tests. The test is based on the definition of **?** procedure. The rank-based procedures are exact under the i.i.d. assumption whereas the sign-based procedures are exact under both the i.i.d. and martingale difference sequence assumption. They showed that rank-based tests are more powerful than their sign-based counterparts.

?, Wald-Type Test: They suggested a joint test based on the following Wald statistic:

$$MVR_{RS}(K) = T(\mathbf{VR} - \mathbf{1})'\mathbf{\Phi}^{-1}(\mathbf{VR} - \mathbf{1})$$

where **VR** is the $K \times 1$ vector of sample K VRs, **1** is the $K \times 1$ unit vector; Φ is the covariance matrix of **VR**. This statistic $MVR_{RS}(K)$ follows a χ^2 distribution with K degrees of freedom. One thing to remember about this test is that the VR tests are computed over Long lags with overlapping observations, the distribution of the VR test is NON-normal.

?, Wald-Type Test: They also developed a Wald-type multiple VR statistic that incorporates the correlations between VR statistics at various horizon and weights them according to their variances:

$$MVR_{CL}(K) = [\mathbf{VR}(K) - \mathbb{E}[\mathbf{VR}(\mathbf{K})]]'\mathbf{\Psi}^{-1}(K)[\mathbf{VR}(K) - \mathbb{E}[\mathbf{VR}(\mathbf{K})]]$$

Again, $\mathbf{VR}(K)$ is a vector of VR statistics and $\mathbf{\Psi}$ is a measure of the covariance matrix of **VR**; and again, this statistic follows a χ^2 distribution with K degrees of freedom. However, after using Monte Carlo techniques to study the empirical distribution of $MVR_{CL}(K)$, they have found that it has large positive skewness, not χ^2.

?, Wald-Type Test: They proposed a joint VR test based on their individual power transformed VR statistic, also following a χ^2 distribution with K degrees of freedom. One feature sets the **?** test apart from the **?** test and the **?** test: this test is with ONE-sided alternative (i.e., $H_1 : \exists K_i$ s.t. $VR(K_i) < 1$).

1.1.3.3 Bootstrapping VR tests

Instead of using the subsampling method, some researchers proposed to employ a bootstrap method, which is distribution-free and can be used to estimate the sampling distribution of the VR statistic when the distribution of the original population is unknown.

? Test: **?** applied the wild bootstrap to the **?** test and the **?** test in 3 stages:

(1) From a bootstrap sample of T observations $X_t^* = \eta_t X_t$, where η_t is a random sequence with $E(\eta) = 0, E(\eta^2) = 1$ and $t = 1, \cdots, T$
(2) For the bootstrap sample generated in (1), calculate $VR^* = VR(X^*, K_i)$
(3) Repeat (1) and (2) for a sufficient amount of times m, to form a bootstrap distribution of the test statistic $\left\{VR(X^*, K_j; j)\right\}_{j=1}^{m}$

Conditional on X_t, X_t^* is a serially uncorrelated sequence with zero mean and variance X_t^2. Thus, wild bootstrapping approximates the sampling distributions under the null hypothesis, which is a desirable property for a bootstrap test. To perform the test, a specific form of η_t should be chosen. **?** recommended using the standard normal distribution for η_t.

? Test: They used a weighted bootstrap method proposed by **?**, which is heteroskedasticity-robust and done by resampling normalized returns instead of actual returns. The bootstrap scheme can be summarized in 4 steps:

(1) For each t, draw a weighting factor z_t^* with replacement from the empirical distribution of normalized returns $z_t = (r_t - \bar{r})/\sigma(r)$, where $t = 1, \cdots, T$, $\bar{(r)} = E(r)$, $\sigma(r) = Var(r)$.
(2) Form the bootstrap sample of T observations $\tilde{r}_t^* = z_t^* r_t$
(3) Calculate the VR statistic $VR^*(K)$ from the sample r_t^*
(4) Repeat steps (1) and (2) M times, obtaining $VR^*(K; m)_{1 \le m \le M}$

Using this procedure, resampling from normalized returns, the weighted bootstrap method accounts for the possible non-constancy of the variance of returns. This weighted bootstrap scheme is designed to overcome the difficulty that resampling methods may generate data that are less dependent than the original data. One thing to notice is that **?**'s method is not asymptotically pivotal and not supported by any asymptotic theory or Monte Carlo evidence to evaluate its properties, unlike **?**'s method.

1.2 Autocorrelations in returns: empirical evidence

With the models introduced in Section **??**, the mean reversion in stock returns has been examined empirically over the years. The common ground is that there are autocorrelations, but the directions of autocorrelations are indefinite across different settings.

A brief summary is listed below, and the details are discussed in the corresponding sub-sections.

1.2.1 Mean Riverse: Negative autocorrelations

1.2.2 Positive autocorrelations

Due to the positive cross-autocorrelations among individual stocks, **Stock indexes** have predominantly positive high-frequency autocorrelations.

1.2.2.1 ??

?? presented evidence that broad market indexes have predominantly positive high-frequency autocorrelations. In **?**, they found a 30% AR(1) with weekly returns of *September 1962* to *December 1985*, while higher-order ARs are also positive although smaller in magnitude. In **?**, they found similar results with the sample period extended to *December 1987*. They proposed that the equal-weighted index autocorrelation could be rewritten into the sum of own-autocovariances and cross-autocovariances of the component securities. Given that the autocorrelations of individual stock returns are generally negative, they deducted that the cross-autocovariances must be positive and large enough to exceed the sum of the negative own-autocovariances. They built the following model and showed the importance of the forecastability *across* securities for contrarian profits:

Consider a stylized contrarian investment strategy: buy stocks at time t that were losers at time $t-k$ and sell stocks at t that were winners at $t-k$. This strategy can be formally written as

$$\omega_{i,t}(k) = -\frac{1}{N}(R_{i,t-k} - R_{m,t-k}),\ i = 1, \cdots, N$$

where $R_{m,t-k} = \frac{1}{N}\sum_{i=1}^{N} R_{i,t-k}$ is the market (equal-weighted) index return.

By construction, $\vec{\omega}_t(k) \equiv [\omega_{1,t}(k), \cdots, \omega_{N,t}(k)]'$ is an arbitrage portfolio since the weights sum to zero. Such a strategy is designed to take advantage of stock market overreactions since the stocks whose returns deviate more from the market index return will be given higher weights (more positive for huge losers and vice versa). Profit generated from this strategy is $\pi_t(k) = \sum_{t=1}^{N} \omega_{i,t}(k)R_{i,t}$, rewrite this profit, get:

$$\pi_t(k) = \sum_{i=1}^{N} \omega_{i,t}(k)R_{i,t} = -\frac{1}{N}\sum_{i=1}^{N}\left(R_{i,t-k} - R_{m,t-k}\right)R_{i,t} = -\frac{1}{N}\sum_{i=1}^{N} R_{i,t-k}R_{i,t} + R_{m,t-k}R_{m,t}$$

take expectation, get:

$$\begin{aligned} E[\pi_t(k)] &= -\frac{1}{N}\sum_{i=1}^{N} E\left[R_{i,t-k}R_{i,t}\right] + E\left[R_{m,t-k}R_{m,t}\right] \\ &= -\frac{1}{N}\sum_{i=1}^{N}\left(Cov\left[R_{i,t-k},R_{i,t}\right]+\mu_i^2\right) + \left(Cov\left[R_{m,t-k},R_{m,t}\right]+\mu_m^2\right) \end{aligned}$$

where $\mu_m \equiv E\left[R_{m,t}\right] = \mu'\iota/N$. Reorganizing this equation into 3 components:

$$\begin{aligned} E[\pi_t(k)] &= -\frac{1}{N}tr(\Gamma_k) - \frac{1}{N}\sum_{i=1}^{N}\mu_i^2 + \frac{\iota'\Gamma_k\iota}{N^2} + \mu_m^2 \\ &= \underbrace{\left\{\frac{\iota'\Gamma_k\iota}{N^2} - \frac{1}{N}tr(\Gamma_k)\right\}}_{\mathbf{L_k}} - \left\{\frac{1}{N}\sum_{i=1}^{N}(\mu_i-\mu_m)^2\right\} \\ &= \underbrace{\frac{1}{N^2}\left[\iota'\Gamma_k\iota - tr(\Gamma_k)\right]}_{\mathbf{C_k}} + \underbrace{\left(-\frac{N-1}{N}\right)tr(\Gamma_k)}_{\mathbf{O_k}} - \underbrace{\frac{1}{N}\sum_{i=1}^{N}(\mu_i-\mu_m)^2}_{\sigma^2(\mu)} \end{aligned}$$

where $tr(\cdot)$ indicates the trace operator (sum of diagonal elements).

The three components are:

- : $\mathbf{C_k}$ depends ONLY on the off-diagonals of the auto-covariance matrix Γ_k
- : $\mathbf{O_k}$ depends ONLY on the diagonals of the auto-covariance matrix Γ_k
- : $\sigma^2(\mu)$ is independent of Γ_k

Or, cross-autocovariances dictate $\mathbf{C_k}$, own-autocovariances dicate $\mathbf{O_k}$, this is the separation needed.

With this decomposition, the authors explained that the profitability of such a contrarian strategy could be perfectly consistent with a **positively autocorrelated market index** and **negatively autocorrelated individual security returns**.

They also presented 5 illustrative scenarios:

A ***I.I.D. returns***
- $\forall k, \Gamma_k = 0, L_k = C_k = O_k = 0$, thus, $E[\pi_t(k)] = -\sigma^2(\mu) \leq 0$.
- ***Intuition***: When returns do follow random walks, any cross-sectional variation in expected returns would generate negative expected profits when trading with a contrarian strategy. Since the strategy is reduced to shorting the higher and buying the lower mean return securities. BUT, $\sigma^2(\mu)$ is generally small.

B ***Stock market overreaction***
- ***Assumption***: negative self-autocorrelations and zero cross-autocorrelations, i.e., the diagonal elements of Γ_k are negative, the non-diagonal elements are

0. Thus, ignoring the small $\sigma^2(\mu)$, the expected profit is

$$E[\pi_t(k)] \simeq L_k = O_k = -\left(\frac{N-1}{N}\right) tr(\Gamma_k) = -\left(\frac{N-1}{N}\right) \sum_{i=1}^{N} \gamma_{ii}(k) > 0$$

- ***Intuition***: In a overreacting market, "what goes up must come down" and vice versa. Thus, a contrarian investment strategy is profitable on average. A special case is the model of "fads": the sum of a random walk and an AR(1)[3].

C ***White noise and lead-lag relations***

- ***Assumption***: $\forall i \in [1, \cdots, N]$, the return is given by: $R_{i,t} = \mu_i + \beta_i \Lambda_{t-i} + \epsilon_{i,t}$, where $\beta_i > 0$, Λ is a serially independent common factor with zero mean and variance σ^2_λ, $\epsilon_{i,t}$ is assumed to be both cross-sectionally and serially independent. When $k < N$, the autocovariance matrix Γ_k has zeros in all entries except along the kth superdiagonal. On the kth superdiagonal, the elements are $\gamma_{i,i+k} = \sigma^2_\lambda \beta_i \beta_{i+k}$, the profit is:

$$E[\pi_t(k)] \simeq L_k = C_k = \frac{\sigma^2_\lambda}{N^2} \sum_{i=1}^{N-k} \beta_i \beta_{i+k} > 0$$

- ***Intuition***: This is an artifact of the dependence of the ith security's return on a lagged common factor, where the lag is determined by i. Notice that the returns are serially independent, but sotck i's returns can be predicted with past returns of stock j, where $j < i$. With this cross-correlation, a contrarian strategy can still profit, as long as the cross autocovariances are sufficiently large.

D ***Nonsynchronous trading and lead-lag effects***[4]

- ***Assumption***:
"Virtual" return: the returns for security $i \in [1, \cdots, N]$ are generated by: $R_{i,t} = \mu_i + \beta_i \Lambda_i + \epsilon_{i,t}$. Λ_t is some zero-mean, i.i.d. common factor, $\epsilon_{i,t}$ is zero-mean, serially and cross-sectionally independent. But this time the returns are *unobservable*, i.e., "virtual".
Non-trade: in each period t, security i has an i.i.d. probability of p_i to be NOT traded. If not traded, a security's *observed* return $R^o_{i,t}$ is 0 while its true return is still given by $R_{i,t} = \mu_i + \beta_i \Lambda_i + \epsilon_{i,t}$.
Observed return: The observed return of security i in period t is the sum of its virtual returns of all the past **consecutive non-trading** periods. Formally, $R^o_{i,t} = \sum_{k=0}^{\infty} X_{i,t}(k) R_{i,t-k}$. The weights $X_{i,t}(k) \equiv (1-\delta_{i,t})\delta_{i,t-1} \cdots \delta_{i,t-k}$, where δ_i is the (i.i.d.) non-trading indicator. $X_{i,t}(k)$ is also an indicator:

$$X_{i,t}(k) = \begin{cases} 1 & i \text{ is traded at } t\text{, but not in any of the } k \text{ previous periods} \\ 0 & \text{otherwise} \end{cases}$$

[3]Notice that only AR(1) statisfies this conclusion.
[4]See **?** for a detailed discussion.

Nontrading duration: the nontrading duration, or the number of past consecutive periods that security i is not traded, is $\tilde{k}_{i,t} \equiv \sum_{k=1}^{\infty}\left(\prod_{j=1}^{k}\delta_{i,t-j}\right)$, its expectation is $E[\tilde{k}_{i,t}] = \frac{p_i}{1-p_i}$.

Portfolio return: for an equal-weighted portfolio of securities with common nontrading probability p_κ, the observed return to portfolio can be approximated as $R^o_{port,t} \longrightarrow \mu_{port} + (1-p_{port})\beta_{port}\sum_{k=0}^{\infty} p^k_{port}\Lambda_{t-k}$, where β_{port} is the average β of the securities. Then the observed return of the portfolio over q periods is $R^o_{port,T}(q) \equiv \sum_{t=(T-1)q+1}^{Tq} R^o_{port,t}$.

- ***Intuition***:
  The "nontrading" problem aims to fix one problem: the prices of distinct securities are mistakenly assumed to be sampled simultaneously. Prices actually happen in different periods, but are treated as if they were observed at the same time. The "power" of a stock on others is related to how frequently it is traded: For a more frequently traded portfolio a, and a less frequently traded portfolio b, $R_{a,t-1}$ predicts $R_{b,t}$ better than $R_{b,t-1}$ predicts $R_{a,t}$. HOWEVER! This cannot fully explain the magnitude of weekly cross-autocorrelations.

E ***Positively dependent common factor and bid-ask spread***

- $R_{i,t}$ as the sum of: (a) a positively autocorrelated common factor, (b) idiosyncratic white noise, (c) a bid-ask spread process. Formally,

$$R_{i,t} = \mu_i + \beta_i\Lambda_i + \eta_{i,t} + \epsilon_{i,t}$$

  where $E[\Lambda_t] = 0$, $E[\Lambda_{t-k}\Lambda_t] \equiv \gamma_\lambda(k) > 0$ (positively autocorrelated common factor), $E[\epsilon_{i,t}] = E[\eta_{i,t}] = 0$ (idiosyncratic noise), $Var[\epsilon_{i,t}] = \sigma_i^2$.
  The bid-ask spread has a AR(1) as $E[\eta_{i,t-1}\eta_{i,t}] = -s_i^2/4$, where s_i is the percentage bid-ask spread. All higher-order ARs and all cross-correlations are zero.
  The autocovariance matrices are given by:

$$\Gamma_1 = \gamma_\lambda(1)\beta\beta'$$

Based on the size-sorted portfolio returns, the authors have found that returns on larger and more liquid stocks and subsequent returns on smaller and less liquid stocks are positively cross-autocorrelated.

1.3 Excess volatility puzzle

[insert text]

1.4 Decomposing prices

[insert text]

1.4.1 Campbell-Schiller decomposition

[insert text]

1.4.2 Lettau-Ludvigson decomposition

[insert text]

1.5 Prediction zoo

[insert text]

1.6 Issues and extensions

[insert text]

1.6.1 Persistency of most regressors

[insert text]

1.6.2 Aggregate predictors without ex-ante choice

[insert text]

1.6.3 Instability in the prediction relation

[insert text]

1.6.4 Measurement

[insert text]

CHAPTER 2

CROSS-SECTION PREDICTABILITY

Contents

Intro:

2.1 Section 1

Section 1:

CHAPTER 3

GMM AND CROSS-SECTION TEST

Contents

Intro:

3.1 Section 1

Section 1:

CHAPTER 4

ADVANCES IN CROSS-SECTION ASSET PRICING

Contents

Intro:

4.1 Section 1

Section 1:

CHAPTER 5

CONSUMPTION-BASED ASSET PRICING

Contents

Intro:

5.1 Section 1

Section 1:

CHAPTER 6

TERM STRUCTURE OF RETURNS

Contents

In this chapter, I summarize the stylized facts and models of intrest rates, and, combining with the time-series and cross-sectional properties of equities, discuss how the term structure of equity can be incorporated into the asset pricing dynamic. Instead of assuming the risk-free rate to be one period, as classic asset pricing models implying in the Euler equations and SDFs, one would expect that an ideal asset pricing model could not only explain the dynamic of equity, but reconcile the property of the term structure of interest rates as well.

The first part of this chapter summarizes studies of risk free bonds and the term structure of this asset class.

6.1 Section 1

Section 1:

CHAPTER 7

LEARNING

Contents

In this chapter, I summarize the learning in empirical finance. This is one of the most cutting-edge research area now.

7.1 Section 1

Section 1:

CHAPTER 8

CURRENCIES: TIME-SERIES AND CROSS-SECTION

Contents

Intro:

8.1 Section 1

Section 1:

CHAPTER 9

INTERMEDIARY-BASED ASSET PRICING

Contents

In this chapter, I summarize the

9.1 Section 1

Section 1:

CHAPTER 10

FINANCE AND BIG DATA

Contents

Intro:

10.1 Section 1

Section 1: