5 The Predictability of Asset Returns: Part II

5.1 Using Overlapping Differences

In this section we further modify our test to use overlapping q-th differences of P_t to estimate the variance σ^2 on the numerator. To be specific, we let $\hat{\mu}$ and $\hat{\sigma}_a^2$ be defined as in the last section. We define

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

Now we define

$$M_d(q) = \hat{\sigma}_c^2(q) - \hat{\sigma}_a^2$$

and

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

We may show that under the null hypothesis, the asymptotic distribution of the two statistics above are given respectively by

$$\sqrt{n} M_d(q) \to_d \mathbb{N}\left(0, \frac{2(q-1)(2q-1)}{3q^2}\sigma^4\right)$$

and

$$\sqrt{n} M_r(q) \to_d \mathbb{N}\left(0, \frac{2(q-1)(2q-1)}{3q^2}\right).$$

The derivation of the limit distribution of the statistics $M_d(q)$ and $M_r(q)$ is more complicated than in the i.i.d. case in Theorem 4.2. It involves a central limit theorem for q-dependent random variables. Interested readers may refer to the appendix of this chapter for a sketch of the derivation.

A question is, why would we bother to make such an adjustment given that this adjustment will make our derivation of the asymptotic behaviors of our test statistics harder? Lo and MacKinlay (1999) claim that the modified statistics improve finite sample performance of the corresponding tests.

A further finite sample refinement is to use unbiased estimators for σ^2. We define

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

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

$$m = q(nq - q + 1)\left(1 - \frac{1}{n}\right),$$

and define

$$\bar{M}_d(q) = \bar{\sigma}_c^2(q) - \bar{\sigma}_a^2,$$

$$\bar{M}_r(q) = \frac{\bar{\sigma}_c^2(q)}{\bar{\sigma}_a^2(q)} - 1.$$

Since this unbiasedness modification is a very minor one, we may show that the limit distributions of $\sqrt{n}\bar{M}_d(q)$ and $\sqrt{n}\bar{M}_r(q)$ are the same as those of $\sqrt{n}M_d(q)$ and $\sqrt{n}M_r(q)$, respectively.

5.2 Testing for Uncorrelated Increments

In the last section we test the martingale hypothesis

$$P_t = \mu + P_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{m.d.s.}$$

by testing a stronger hypothesis

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

In this section, we replace the i.i.d. null hypothesis by a less strong one, namely the uncorrelated increments hypothesis:

$$H_0 : \{\varepsilon_t\} \text{ is strictly stationary and } \mathbb{E}\varepsilon_t\varepsilon_s = 0 \quad \text{for all } t \neq s.$$

Definition 5.1. A sequence of random variables $\{X_t\}_{t\in\mathbb{Z}}$ is called strictly stationary if the distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})'$ is the same as the distribution of $(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})'$ for any choice of $t_1, t_2, \ldots, t_n$ and $\tau \in \mathbb{Z}$.

Any iid sequence of random variables is strictly stationary.

Suppose that $\{X_t\}$ is a strictly stationary sequence of random variables. Then we have that for any t, t_1, t_2 and τ,

$$\mathbb{E}X_t = \mathbb{E}X_{t+\tau},$$

and

$$\mathrm{Cov}(X_{t_1}, X_{t_2}) = \mathrm{Cov}(X_{t_1+\tau}, X_{t_2+\tau})$$

if they exist. We see that the expectation of X_t is the same for all t, and the covariance of two random variables X_{t_1}, X_{t_2} only depends on the time difference $t_2 - t_1$. Therefore, we define the autocovariance function of the sequence to be

$$\gamma(k) = \mathrm{Cov}(X_t, X_{t+k}) = \mathrm{Cov}(X_1, X_{1+k}).$$

We define the autocorrelation function of the sequence $\{X_t\}$ by

$$\rho(k) = \frac{\mathrm{Cov}(X_t, X_{t+k})}{\mathrm{Var}(X_t)} = \frac{\gamma(k)}{\gamma(0)}.$$

It is easy to see that $\gamma(k) = \gamma(-k)$ and $\rho(k) = \rho(-k)$ for all k. If $\{X_t\}$ is a sequence of uncorrelated random variables, then $\gamma(k) = 0$ and $\rho(k) = 0$ for all k.

Let's come back to our test. The test for our null hypothesis is based on the autocorrelation function $\rho(k)$. We need to estimate it first. Suppose we observe $P_0, P_1, P_2, \ldots, P_T$. We estimate $\rho(k)$ through

$$\hat{\mu} = \frac{P_T - P_0}{T},$$

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=k+1}^{T}(P_t - P_{t-1} - \hat{\mu})(P_{t-k} - P_{t-k-1} - \hat{\mu}),$$

and

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}.$$

Under the null hypothesis and some regularity conditions, we may show that

$$\sqrt{T}\hat{\rho}(k) \rightarrow_d \mathbb{N}(0,1)$$

as $T \rightarrow \infty$.

The estimator $\hat{\rho}(k)$ is a biased estimator. For small samples this bias could be large. A bias-corrected version $\tilde{\rho}(k)$ is therefore proposed:

$$\tilde{\rho}(k) = \hat{\rho}(k) + \frac{T-k}{(T-1)^2}(1 - \hat{\rho}^2(k)).$$

We may show that under the null hypothesis and some regularity conditions, we have

$$\frac{T}{\sqrt{T-k}}\tilde{\rho}(k) \rightarrow_d \mathbb{N}(0,1).$$

5.3 The Q-Statistics

Each of the tests in the above section tests for the zero correlation for a particular k. However, we know that under the null hypothesis the autocorrelations are zero for all k. Box and Pierce (1970) proposed a Q-statistic that utilizes this fact:

$$Q_m = T\sum_{k=1}^{m} \hat{\rho}(k)^2.$$

It can be shown that

$$Q_m \rightarrow_d \chi^2_m.$$

as $T \rightarrow \infty$. That is, the Q-statistic converges in distribution to a chi-square distribution with degree of freedom m.

Ljung and Box (1978) provided a finite sample correction for the Q-statistic:

$$Q'_m = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}^2(k)}{T-k}.$$

This statistic also converges in distribution to χ^2_m.

A robustified version of the Q-statistic can be found in Lobato et al. (2002). This modification is constructed by replacing $\hat{\rho}^2(k)$ with $\tilde{\rho}^2(k)$ in the above Q-statistics where

$$\tilde{\rho}^2(k) = \frac{\hat{\gamma}(k)^2}{\hat{\tau}(k)},$$

$$\hat{\tau}_j = \frac{1}{T-k}\sum_{t=k+1}^{T}(P_t - P_{t-1} - \hat{\mu})^2(P_{t-k} - P_{t-k-1} - \hat{\mu})^2.$$

The limit distributions preserve.

5.4 Appendix: Deriving the Asymptotic Distribution of $M_r(q)$

We only provide a sketch here. We focus on $M_r(q)$ since the derivation of the asymptotic distribution of $M_d(q)$ is simpler. We follow the basic idea in the proof of Theorem 4.2. The main difference is now we need to use a central limit theorem for strictly stationary m-dependent sequences of random variables/vectors.

Theorem 5.2. *Let $\{X_t\}$ be a strictly stationary m-dependent sequence of random vectors with mean zero and autocovariance function $\Gamma(j)$, $j = 0, \pm1, \pm2, \ldots$. If $\sum_{|j|\leq m}\Gamma(j) \neq 0$, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \to_d \mathbb{N}\left(0, \sum_{|j|\leq m}\Gamma(j)\right).$$

We may first assume that $\mu = 0$ and later make an argument that the estimation of μ does not affect the asymptotic distribution of our test statistics, just as in the previous chapter. The core part of the derivation is to obtain the limit distribution of

$$\frac{1}{\sqrt{(n-1)q+1}}\sum_{k=q}^{nq} Z_k$$

where

$$Z_k = \begin{bmatrix} \frac{1}{q}(\varepsilon_{k-q+1}+\varepsilon_{k-q+2}+\ldots+\varepsilon_k)^2-\sigma^2 \\ \varepsilon_k^2-\sigma^2 \end{bmatrix}$$

is mean zero.

Under the null hypothesis, it is easy to see that $\{Z_k\}$ is strictly stationary and $q-1$-dependent. Then we need to obtain the autocovariance function $\Gamma(j)$ of Z_k for $|j|<q$. We have that

$$\Gamma(j) = \begin{bmatrix} \sigma_{11}(j) & \sigma_{12}(j) \\ \sigma_{21}(j) & \sigma_{22}(j) \end{bmatrix}$$

where

$$\begin{aligned}
\sigma_{11}(j) &= \mathrm{Cov}\left(\frac{(\varepsilon_{k-q+1}+\varepsilon_{k-q+2}+\ldots+\varepsilon_k)^2}{q}, \frac{(\varepsilon_{k-q+1-j}+\varepsilon_{k-q+2-j}+\ldots+\varepsilon_{k-j})^2}{q}\right) \\
&= \frac{1}{q^2}\mathrm{Cov}\left((\varepsilon_1+\ldots+\varepsilon_q)^2, (\varepsilon_{1+j}+\ldots\varepsilon_{q+j})^2\right) \\
&= \frac{1}{q^2}\left(\mathbb{E}(\varepsilon_1+\ldots+\varepsilon_q)^2(\varepsilon_{1+j}+\ldots\varepsilon_{q+j})^2 - \mathbb{E}(\varepsilon_1+\ldots+\varepsilon_q)^2\mathbb{E}(\varepsilon_{1+j}+\ldots\varepsilon_{q+j})^2\right)
\end{aligned}$$

for $0\le j<q$, and $\sigma_{11}(-j)=\sigma_{11}(j)$. Write

$$\mathbb{E}(\varepsilon_1+\ldots+\varepsilon_q)^2(\varepsilon_{1+j}+\ldots\varepsilon_{q+j})^2 = \sum_{1\le i_1\le q}\sum_{1\le i_2\le q}\sum_{1+j\le i_3\le q+j}\sum_{1+j\le i_4\le q+j}\mathbb{E}\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{i_3}\varepsilon_{i_4}$$

There are q^2 terms in the sums on the right hand side of the above equation. Each of these individual terms are non-zero only if they belong to the one of the following four categories:

1. $i_1=i_2=i_3=i_4$. There are only $q-j$ such terms. Each term is equal to $\mathbb{E}\varepsilon_1^4$. We denote the fourth moment by A.

2. $i_1 = i_2 \neq i_3 = i_4$. There are $q^2 - (q-j)$ such terms. Each term is equal to σ^4.

3. $i_1 = i_3 \neq i_2 = i_4$. There are $(q-j)(q-j-1)$ such terms. Each term is equal to σ^4.

4. $i_1 = i_4 \neq i_2 = i_3$. There are $(q-j)(q-j-1)$ such terms. Each term is equal to σ^4.

So in the end, we have

$$\sigma_{11}(j) = \frac{(q-j)(A - 3\sigma^4) + 2(q-j)^2\sigma^4}{q^2}$$

for all $0 \leq j < q$, and $\sigma_{11}(-j) = \sigma_{11}(j)$.

Also, similar calculation shows that

$$\begin{aligned}\sigma_{12}(j) &= \mathrm{Cov}\left(\frac{(\varepsilon_{k-q+1} + \varepsilon_{k-q+2} + \ldots + \varepsilon_k)^2}{q}, \varepsilon_{k-j}^2\right)\\ &= \begin{cases} \frac{A-\sigma^4}{q}, & \text{if } j \geq 0 \\ 0, & \text{if } j < 0, \end{cases}\end{aligned}$$

and

$$\begin{aligned}\sigma_{21}(j) &= \mathrm{Cov}\left(\frac{(\varepsilon_{k-q+1-j} + \varepsilon_{k-q+2-j} + \ldots + \varepsilon_{k-j})^2}{q}, \varepsilon_{k}^2\right)\\ &= \begin{cases} 0 & \text{if } j > 0 \\ \frac{A-\sigma^4}{q}, & \text{if } j \leq 0. \end{cases}\end{aligned}$$

In the end, it is easy to show that $\sigma_{22}(j)$ is $A - \sigma^4$ if $j = 0$ and is 0 if $j \neq 0$. Some algebra yields that

$$\sum_{|j| \leq q-1} \Gamma(j) = \begin{bmatrix} A - \sigma^4 + \frac{2(q-1)(2q-1)\sigma^4}{3q} & A - \sigma^4 \\ A - \sigma^4 & A - \sigma^4. \end{bmatrix}$$

That is, by the central limit theorem, we have

$$\frac{1}{\sqrt{(n-1)q+1}}\sum_{k=q}^{nq} Z_k \to_d \mathbb{N}\left(0, \begin{bmatrix} A-\sigma^4+\frac{2(q-1)(2q-1)\sigma^4}{3q} & A-\sigma^4 \\ A-\sigma^4 & A-\sigma^4 \end{bmatrix}\right).$$

The rest is to apply the Delta method, just as in the proof of Theorem 4.2.