Portfolios that Contain Risky Assets 11: Independent, Identically-Distributed Models for Assets

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*
April 27, 2019 version

Portfolios that Contain Risky Assets
Part II: Stochastic Models

Independent, Identically-Distributed Models for Assets

Independent, Identically-Distributed Models for Assets

Investors have long followed the old adage "don't put all your eggs in one basket" by holding diversified portfolios. However, before MPT the value of diversification had not been quantified. Key aspects of MPT are:

1. it uses the return mean as a proxy for reward;
2. it uses volatility as a proxy for risk;
3. it analyzes Markowitz portfolios;
4. it shows diversification can reduce volatility;
5. it identifies the efficient frontier as the place to be.

Independent, Identically-Distributed Models for Assets

The orignial form of MPT did not give guidance to investors about where to be on the efficient frontier. This question was addressed in the 1960's, most notably by William Sharpe, who shared the 1990 Nobel Prize in Economics with Harry Markowitz. We will not present that work here. Rather, we will build stochastic models that can be used in conjunction with MPT to address this question. *By doing so, we will learn that maximizing the return mean is not the best strategy for maximizing reward.*

We begin by building models of one risky asset with a share price history $\{s(d)\}_{d=0}^{D}$. Let $\{r(d)\}_{d=1}^{D}$ be the associated return history. Because each $s(d)$ is positive, each $r(d)$ lies in the interval $(-1, \infty)$.

Independent, Identically-Distributed Models for Assets

An *independent, identically-distributed (IID)* model for this history simply independently draws D random numbers $\{R_d\}_{d=1}^{D}$ from $(-1,\infty)$ in accord with a fixed probability density $q(R)$ over $(-1,\infty)$. This means that $q(R)$ is a nonnegative integrable function such that

$$\int_{-1}^{\infty} q(R)\,\mathrm{d}R = 1\,, \tag{1.1}$$

and that the probability that each R_d takes a value inside any sufficiently nice $A \subset (-1,\infty)$ is given by

$$\Pr\{R_d \in A\} = \int_A q(R)\,\mathrm{d}R\,. \tag{1.2}$$

Here capital letters R_d denote random numbers drawn from $(-1,\infty)$ in accord with the probability density $q(R)$ rather than real return data.

Independent, Identically-Distributed Models for Assets

IID models are the simplest models consistent with the way any portfolio selection theory is used. Such theories have three basic steps.

- Calibrate a model for asset behavior from historical data.
- Use the model to suggest how a set of ideal portfolios might behave.
- Use these suggestions to select the portfolio that optimizes an objective.

This strategy assumes that in the future the market will behave statistically as it did in the past.

This assumption requires the market statistics to be stable relative to its dynamics. But this requires future states to decorrelate from past states. The simplest class of models with this property assumes that future states are independent of past states, which maximizes this decorrelation. These are called Markov models. IID models are the simplest Markov models.

Independent, Identically-Distributed Models for Assets

It is easy to develop more complicated Markov models. For example, we could use a different probability density for each day of the week rather than treating all trading days the same. Because there are usually five trading days per week, Monday through Friday, such a model would require calibrating each of the five densities with one fifth as much data. There would then be greater uncertainty associated with the calibration. Moreover, we then have to figure out how to treat weeks that have less than five trading days due to holidays. Perhaps just the first and last trading days of each week should get their own probability density, no matter on which day of the week they fall.

Before increasing the complexity of a model, we should investigate whether the costs of doing so outweigh the benefits. Specifically, we should investigate whether there is benefit in treating any one trading day of the week differently than the others before building a more complicated model.

Expected Values and Variances

Once we have decided to use an IID model for a particular asset, you might think the next goal is to pick an appropriate probability density $q(R)$. One way to do this is to consider an explicit family of probability densities $q(R;\beta)$ parametrized by β. The values of the parameters β are then calibrated so that a sample $\{R_d\}_{d=1}^D$ drawn from $q(R;\beta)$ mimics certain statistics of observed daily return history $\{r(d)\}_{d=1}^D$. Statisticians call this approach *parametric.*

However, we will take another approach. We will *identify statistical information like the expected value and variance of functions $\psi(R)$ that shed light upon the market and that can be estimated from a sample $\{R_d\}_{d=1}^D$ drawn from $q(R)$.* Ideally this information should be insensitive to details of $q(R)$ within a large class of probability densities. Statisticians call this approach *nonparametric.*

Expected Values and Variances

For any function $\psi : (-1, \infty) \to \mathbb{R}$ the *expected value* of $\Psi = \psi(R)$ is given by

$$\mathrm{Ex}(\Psi) = \int_{-1}^{\infty} \psi(R)\, q(R)\, \mathrm{d}R\,, \tag{2.3}$$

provided that $|\psi(R)|\, q(R)$ is integrable.

Remark. The term "expected value" can be misleading because for most densities $q(R)$ it is not a value that we would expect to see more than other values. For example, if $q(R) = \exp(-1 - R)$ then $\mathrm{Ex}(R) = 0$, but it is clear that values of R close to -1 are over twice as likely than values of R close to 0. More dramatically, if $q(R)$ concentrates around the values $R = -0.50$ and $R = 2.00$ with equal propbability then $\mathrm{Ex}(R) = 0.75$, which is a value that is never seen. However, this terminology is standard, so we have to live with it. *Please keep in mind that an expected value may not be near the values that we should expect to see.*

Expected Values and Variances

The *variance* of $\Psi = \psi(R)$ is given by

$$\begin{aligned}\mathrm{Var}(\Psi) &= \mathrm{Ex}\Big(\big(\psi(R) - \mathrm{Ex}(\Psi)\big)^2\Big) \\ &= \int_{-1}^{\infty} \big(\psi(R) - \mathrm{Ex}(\Psi)\big)^2 \, q(R)\, \mathrm{d}R\,,\end{aligned} \tag{2.4}$$

provided that $|\psi(R)|^2 \, q(R)$ is integrable.

Remark. This term "variance" is clearly better than that of "expected value" because the variance is clearly a quantification of how $\psi(R)$ deviates from $\mathrm{Ex}(\Psi)$. Moreover, it is the most commonly used such measure. However, there are others, so we must always question if its use is appropriate in any situation.

Expected Values and Variances

The most important expected value and variance are those of R itself. These are the return mean μ and return variance ξ, which are obtained from (2.3) and (2.4) by setting $\Psi = \psi(R) = R$, yielding

$$\begin{aligned} \mu &= \mathrm{Ex}(R) = \int_{-1}^{\infty} R\, q(R)\, \mathrm{d}R\,, \\ \xi &= \mathrm{Var}(R) = \mathrm{Ex}\Big((R-\mu)^2\Big) = \int_{-1}^{\infty} (R-\mu)^2\, q(R)\, \mathrm{d}R\,. \end{aligned} \tag{2.5}$$

For these to exist we need to require that $q(R)$ satisfies

$$\int_{-1}^{\infty} R^2 q(R)\, \mathrm{d}R < \infty\,.$$

More restrictive requirements are often imposed so that the expected values and variances of other random variables exist.

Expected Values and Variances

The *standard deviation* of $\Psi = \psi(R)$ is given by

$$\mathrm{St}(\Psi) = \sqrt{\mathrm{Var}(\Psi)}\,, \tag{2.6}$$

provided that $\mathrm{Var}(\Psi)$ exists.

The standard deviation is one measure of how far from $\mathrm{Ex}(\Psi)$ that we can expect the value of any given $\psi(R)$ to be. It arises naturally in the *Chebyshev inequality*, which states that for every $\delta > 1$ we have

$$\mathrm{Pr}\Big\{|\psi(R) - \mathrm{Ex}(\Psi)| \geq \delta\,\mathrm{St}(\Psi)\Big\} \leq \frac{1}{\delta^2}\,. \tag{2.7}$$

Notice that the left-hand side is always less than or equal to 1, so that the condition $\delta > 1$ is required for the bound (2.7) to be meaningful.

Expected Values and Variances

The proof of the Chebyshev inequality (2.7) is simple. We have

$$
\begin{aligned}
\Pr\Big\{ |\psi(R) - \mathrm{Ex}(\Psi)| \geq \delta\, \mathrm{St}(\Psi) \Big\} \\
&= \int_{\{R \in (-1,\infty)\,:\, |\psi(R) - \mathrm{Ex}(\Psi)| \geq \delta\, \mathrm{St}(\Psi)\}} q(R)\, \mathrm{d}R \\
&\leq \int_{-1}^{\infty} \frac{|\psi(R) - \mathrm{Ex}(\Psi)|^2}{\delta^2\, \mathrm{St}(\Psi)^2}\, q(R)\, \mathrm{d}R \\
&= \frac{\mathrm{Var}(\Psi)}{\delta^2\, \mathrm{St}(\Psi)^2} = \frac{1}{\delta^2}\,.
\end{aligned}
$$

□

Remark. The Chebyshev inequality is not sharp, but is often useful.

Expected Value Estimators

Because $q(R)$ is unknown, the expected value of any $\Psi = \psi(R)$ must be estimated from data. Suppose that we draw a sample $\{R_d\}_{d=1}^D$ from the probability density $q(R)$. We claim that for any choice of positive weights $\{w_d\}_{d=1}^D$ such that

$$\sum_{d=1}^{D} w_d = 1\,, \tag{3.8}$$

we can approximate $\mathrm{Ex}(\Psi)$ by the weighted average

$$\widehat{\mathrm{Ex}}(\Psi) = \sum_{d=1}^{D} w_d\,\Psi_d\,, \tag{3.9}$$

where $\Psi_d = \psi(R_d)$. The weighted average (3.9) is the *sample mean* of $\{\Psi_d\}_{d=1}^D$ for the weights $\{w_d\}_{d=1}^D$.

Expected Value Estimators

We will present three facts that make precise the sense in which the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ approximates $\mathrm{Ex}(\Psi)$. They will show that $\widehat{\mathrm{Ex}}(\Psi)$ is more likely to take values closer to $\mathrm{Ex}(\Psi)$ for larger samples $\{R_d\}_{d=1}^{D}$. Therefore we call $\widehat{\mathrm{Ex}}(\Psi)$ an *estimator* of $\mathrm{Ex}(\Psi)$.

The first fact is simply the computation of the expected value of the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.9).

Fact 1.

$$\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \mathrm{Ex}(\Psi) \,. \tag{3.10}$$

This says that $\widehat{\mathrm{Ex}}(\Psi)$ is a so-called *unbiased estimator* of $\mathrm{Ex}(\Psi)$.

Expected Value Estimators

Proof. Because each draw is independent, probability density over $(-1,\infty)^D$ of the sample $\{R_d\}_{d=1}^D$ is

$$q(R_1)\, q(R_2)\, \cdots\, q(R_D)\,.$$

Therefore we have

$$\begin{aligned}
\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) &= \int_{-1}^{\infty} \cdots \int_{-1}^{\infty} \sum_{d=1}^{D} w_d\, \psi(R_d)\, q(R_1) \cdots q(R_D)\, \mathrm{d}R_1 \cdots \mathrm{d}R_D \\
&= \sum_{d=1}^{D} w_d \int_{-1}^{\infty} \psi(R_d)\, q(R_d)\, \mathrm{d}R_d \\
&= \sum_{d=1}^{D} w_d \mathrm{Ex}(\Psi) = \mathrm{Ex}(\Psi)\,.
\end{aligned}$$

Expected Value Estimators

The second fact is simply the computation of the variance of the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.9).

Fact 2.

$$\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \bar{w}_D \,\mathrm{Var}(\Psi)\,, \tag{3.11}$$

where $\bar{w}_D$ is the weighted average of the weights $\{w_d\}_{d=1}^D$ given by

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2\,. \tag{3.12}$$

This fact says that the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ converges to $\mathrm{Ex}(\Psi)$ like $\sqrt{\bar{w}_D}$ as $D \to \infty$. Because $\bar{w}_D = 1/D$ for uniform weights, we see that this rate of convergence is $1/\sqrt{D}$ as $D \to \infty$.

Expected Value Estimators

Remark. The *Cauchy inequality* from multivariable calculus states that

$$\sum_{d=1}^{D} a_d \, b_d \leq \left(\sum_{d=1}^{D} a_d^2 \right)^{\frac{1}{2}} \left(\sum_{d=1}^{D} b_d^2 \right)^{\frac{1}{2}} .$$

By using fact (3.8) that the weights $\{w_d\}_{d=1}^{D}$ sum to 1 and applying the Cauchy inequality to $a_d = 1$ and $b_d = w_d$ we see that

$$1 = \left(\sum_{d=1}^{D} 1 \, w_d \right)^2 \leq \left(\sum_{d=1}^{D} 1^2 \right) \left(\sum_{d=1}^{D} w_d^2 \right) = D \, \bar{w}_D \, .$$

Therefore $1/D \leq \bar{w}_D$ for any choice of weights. Because $\bar{w}_D = 1/D$ for uniform weights, we see that the rate of convergence of $\widehat{\mathrm{Ex}}(\Psi)$ to $\mathrm{Ex}(\Psi)$ is fastest for uniform weights.

Expected Value Estimators

Proof. By Fact 1 we have

$$\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \mathrm{Ex}(\Psi)\,,$$

whereby

$$\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \sum_{d=1}^{D} w_d \Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)\,.$$

By squaring both sides of this equality we obtain

$$\begin{aligned}&\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big)\Big)^2 \\ &= \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} w_{d_1}\, w_{d_2} \Big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\Big) \Big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\Big)\,.\end{aligned}$$

By taking the expected value of this relation we find that

Expected Value Estimators

$$
\begin{aligned}
\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) &= \mathrm{Ex}\left(\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big)\Big)^2\right) \\
&= \mathrm{Ex}\left(\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1}\, w_{d_2}\,\Big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\Big)\Big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\Big)\right) \\
&= \sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1}\, w_{d_2}\,\mathrm{Ex}\left(\Big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\Big)\Big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\Big)\right) \\
&= \sum_{d=1}^{D} w_d^2\,\mathrm{Ex}\left(\Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)^2\right) \\
&= \sum_{d=1}^{D} w_d^2\,\mathrm{Var}(\Psi) = \bar{w}_D\,\mathrm{Var}(\Psi)\,.
\end{aligned}
$$

Expected Value Estimators

Remark. As in the proof of **Fact 1**, here we computed expected values by using the probability density over $(-1, \infty)^D$ given by

$$q(R_1)\, q(R_2)\, \cdots\, q(R_D)\,.$$

The off-diagonal terms in the foregoing double sum vanished because

$$\mathrm{Ex}\Big(\big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\big)\,\big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\big)\Big) = 0 \quad \text{when } d_1 \neq d_2\,,$$

while the diagonal terms reduced to

$$\begin{aligned} &\mathrm{Ex}\Big(\big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\big)\,\big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\big)\Big) \\ &= \mathrm{Ex}\Big(\big(\Psi_{d} - \mathrm{Ex}(\Psi)\big)^2\Big) \quad \text{when } d_1 = d_2 = d\,. \end{aligned}$$

Expected Value Estimators

The third fact is simply the Chebyshev inequality associated with the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.9).

Fact 3. For every $\delta > \sqrt{\bar{w}_D}$ we have

$$\mathrm{Pr}\left\{ \left| \widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi) \right| \geq \delta \, \mathrm{St}(\Psi) \right\} \leq \frac{\bar{w}_D}{\delta^2} \,. \tag{3.13}$$

Remark. The proof of this fact is similar to that of the Chebyshev inequality (2.7). The difference is that here we will integrate over $(-1, \infty)^D$ with probability density

$$q(R_1)\, q(R_2) \, \cdots \, q(R_D) \,,$$

rather than $(-1, \infty)$ with probability density $q(R)$.

Expected Value Estimators

Proof. By Fact 2 we have

$$\begin{aligned}
&\Pr\left\{\left|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\right| \geq \delta\,\mathrm{St}(\Psi)\right\} \\
&\quad = \int \cdots \int_{\left\{\left|\widehat{\mathrm{Ex}}(\Psi)-\mathrm{Ex}(\Psi)\right|\geq\delta\,\mathrm{St}(\Psi)\right\}} q(R_1)\cdots q(R_D)\,\mathrm{d}R_1\cdots\mathrm{d}R_D \\
&\quad \leq \int_{-1}^{\infty}\cdots\int_{-1}^{\infty} \frac{|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)|^2}{\delta^2\,\mathrm{St}(\Psi)^2}\, q(R_1)\cdots q(R_D)\,\mathrm{d}R_1\cdots\mathrm{d}R_D \\
&\quad = \frac{\mathrm{Var}\left(\widehat{\mathrm{Ex}}(\Psi)\right)}{\delta^2\,\mathrm{St}(\Psi)^2} = \frac{\bar{w}_D\,\mathrm{Var}(\Psi)}{\delta^2\,\mathrm{St}(\Psi)^2} = \frac{\bar{w}_D}{\delta^2}\,.
\end{aligned}$$

□

Expected Value Estimators

Remark. The Chebyshev inequality (3.13) with $\Psi = \psi(R) = R$ implies

$$\mathrm{Pr}\left\{\left|\widehat{\mathrm{Ex}}(R) - \mathrm{Ex}(R)\right| < \delta\,\mathrm{St}(R)\right\} > 1 - \frac{\bar{w}_D}{\delta^2}\,.$$

This can be used to quantify the uncertainty in the estimator $\widehat{\mathrm{Ex}}(R)$ of the return mean $\mu = \mathrm{Ex}(R)$ of an asset with standard deviation $\sigma = \mathrm{St}(R)$. For example, if we use uniform weights with $D = 250$ then $\bar{w}_D = \frac{1}{250}$ and:

- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{2}\sigma$ of μ with probability > 0.984;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{5}\sigma$ of μ with probability > 0.900;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{7}\sigma$ of μ with probability > 0.804;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{10}\sigma$ of μ with probability > 0.600;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{15}\sigma$ of μ with probability > 0.100.

Expected Value Estimators

Remark. Fact 3 establishes the *law of large numbers*, which states that the sample means $\widehat{\mathrm{Ex}}(\Psi)$ converge to $\mathrm{Ex}(\Psi)$:

$$\lim_{D\to\infty} \widehat{\mathrm{Ex}}(\Psi) = \mathrm{Ex}(\Psi)\,.$$

More precisely, it establishes the *weak law of large numbers*, which asserts that the sample means *converge in probability*.

There is also the *strong law of large numbers*, which asserts that the sample means *converge almost surely*.

These notions of convergence are covered in advanced probability courses. In practice D is finite, so bounds like the one discussed on the last slide are often more useful than these limits.

Variance Estimators

Because $q(R)$ is unknown, the variance of any $\Psi = \psi(R)$ must also be estimated from data. Suppose that we draw a sample $\{R_d\}_{d=1}^{D}$ from the probability density $q(R)$. We claim that for any choice of positive weights $\{w_d\}_{d=1}^{D}$ such that

$$\sum_{d=1}^{D} w_d = 1\,, \tag{4.14}$$

we can approximate $\mathrm{Var}(\Psi)$ by the weighted average

$$\widehat{\mathrm{Var}}(\Psi) = \frac{1}{1 - \bar{w}_D} \sum_{d=1}^{D} w_d \Big(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\Big)^2\,, \tag{4.15}$$

where $\Psi_d = \psi(R_d)$. This weighted average is the factor $1/(1 - \bar{w}_D)$ times the *sample variance* of $\{\Psi_d\}_{d=1}^{D}$ for the weights $\{w_d\}_{d=1}^{D}$.

Variance Estimators

The factor $1/(1 - \bar{w}_D)$ multiplying the sample variance in (4.15) insures that $\widehat{\mathrm{Var}}(\Psi)$ is an *unbiased estimator* of $\mathrm{Var}(\Psi)$.

Fact 4.

$$\mathrm{Ex}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) = \mathrm{Var}(\Psi)\,. \tag{4.16}$$

Remark. This fact about $\widehat{\mathrm{Var}}(\Psi)$ is the analog of **Fact 1** about $\widehat{\mathrm{Ex}}(\Psi)$.

Remark. There are facts about $\widehat{\mathrm{Var}}(\Psi)$ that are analogs of **Fact 2** and **Fact 3** about $\widehat{\mathrm{Ex}}(\Psi)$. However, we will not state them here. Taken together these facts show that $\widehat{\mathrm{Var}}(\Psi)$ given by (4.15) is an unbiased estimator of $\mathrm{Var}(\Psi)$.

Variance Estimators

Proof. First, verify the identity

$$\widehat{\mathrm{Var}}(\Psi) = \frac{1}{1-\bar{w}_D}\sum_{d=1}^{D} w_d\Big(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\Big)^2$$
$$= \frac{1}{1-\bar{w}_D}\sum_{d=1}^{D} w_d\Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)^2 - \frac{1}{1-\bar{w}_D}\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\Big)^2 . \tag{4.17}$$

By **Fact 2** we have

$$\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \bar{w}_D\,\mathrm{Var}(\Psi) .$$

Variance Estimators

By taking the expected value of (4.17) we confirm that

$$\begin{aligned}
\mathrm{Ex}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) &= \sum_{d=1}^{D} \frac{w_d}{1-\bar{w}_D}\,\mathrm{Ex}\Big(\big(\Psi_d - \mathrm{Ex}(\Psi)\big)^2\Big) \\
&\quad - \frac{1}{1-\bar{w}_D}\,\mathrm{Ex}\Big(\big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\big)^2\Big) \\
&= \sum_{d=1}^{D} \frac{w_d}{1-\bar{w}_D}\,\mathrm{Var}(\Psi) - \frac{1}{1-\bar{w}_D}\,\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) \\
&= \frac{\mathrm{Var}(\Psi)}{1-\bar{w}_D} - \frac{\bar{w}_D\,\mathrm{Var}(\Psi)}{1-\bar{w}_D} = \mathrm{Var}(\Psi)\,.
\end{aligned}$$

□

Checking for Identically Distributed

In an IID model the random numbers $\{R_d\}_{d=1}^D$ are each drawn from $(-1,\infty)$ in accord with the *same* probability density $q(R)$. Therefore if we plot the points $\{(d,R_d)\}_{d=1}^D$ in the dr-plane they will usually be distributed in a way that looks uniform in d. *Therefore if the return history $\{r(d)\}_{d=1}^D$ is mimiced by such a model then the points $\{(d,r(d))\}_{d=1}^D$ plotted in the dr-plane should appear to be distributed in a way that is unifrom in d.*

This will be the case if every subsample of the return history $\{r(d)\}_{d=1}^D$ behaves as if it was drawn from the same probability density. Therefore the question that we must address is how to tell when two samples, $\{r_1(d)\}_{d=1}^{D_1}$ and $\{r_2(d)\}_{d=1}^{D_2}$, might be drawn from the same probability density.

Checking for Identically Distributed

We start with a simpler question. How to compare two probability densities over $(-1, \infty)$, say $p_1(R)$ and $p_2(R)$ where $p_1(R) \geq 0$, $p_2(R) \geq 0$, and

$$\int_{-1}^{\infty} p_1(R)\,\mathrm{d}R = \int_{-1}^{\infty} p_2(R)\,\mathrm{d}R = 1\,.$$

One idea is to compare their distributions $P_1(R)$ and $P_2(R)$, which are

$$P_1(R) = \int_{-1}^{R} p_1(R')\,\mathrm{d}R'\,, \qquad P_2(R) = \int_{-1}^{R} p_2(R')\,\mathrm{d}R'\,.$$

These are nondecreasing functions of R over $(-1, \infty)$ such that

$$\lim_{R\to-1} P_1(R) = \lim_{R\to-1} P_1(R) = 0\,, \qquad \lim_{R\to\infty} P_1(R) = \lim_{R\to\infty} P_1(R) = 1\,.$$

Checking for Identically Distributed

The *Kolmogorov-Smirnov* measure of the closeness of P_1 and P_2 is the sup norm of their difference:

$$\|P_2 - P_1\|_{\mathrm{KS}} = \sup\{|P_2(R) - P_1(R)| \,:\, R \in (-1, \infty)\}\,.$$

The *Kuiper* measure of the closeness of P_1 and P_2 is

$$\begin{aligned}\|P_2 - P_1\|_{\mathrm{Ku}} = &\sup\{P_2(R) - P_1(R) \,:\, R \in (-1, \infty)\} \\ &+ \sup\{P_1(R) - P_2(R) \,:\, R \in (-1, \infty)\}\,.\end{aligned}$$

It can be shown that

$$\|P_2 - P_1\|_{\mathrm{KS}} \leq \|P_2 - P_1\|_{\mathrm{Ku}} \leq 1\,.$$

Checking for Identically Distributed

The *Cramer-von Mises* measure of the closeness of P_1 and P_2 is the L^2-norm of their difference:

$$\|P_2 - P_1\|_{\mathrm{CvM}} = \left(\int_{-1}^{\infty} \big(P_2(R) - P_1(R)\big)^2 \mathrm{d}R \right)^{\frac{1}{2}} .$$

This can clearly be generalized to any L^p-norm with respect to any positive measure over $(-1, \infty)$.

For simplicity we will stick to the Kolmogorov-Smirnov and Kuiper measures.

Checking for Identically Distributed

Now we return to our original question. Given two samples, $\{r_1(d)\}_{d=1}^{D_1}$ and $\{r_2(d)\}_{d=1}^{D_2}$, we construct their so-called *emperical distributions*

$$\widehat{P}_1(R) = \frac{\#\{d \,:\, r_1(d) \leq R\}}{D_1}\,, \qquad \widehat{P}_2(R) = \frac{\#\{d \,:\, r_2(d) \leq R\}}{D_2}\,.$$

Here $\#S$ denotes the number of elements in a set S. These approximate the unknown true distributions P_1 and P_2 because

$$P_1(R) = \mathrm{Pr}\{r_1(d) \leq R\}\,, \qquad P_2(R) = \mathrm{Pr}\{r_2(d) \leq R\}\,.$$

Then the Kolmogorov-Smirnov and Kuiper measures of the difference $\widehat{P}_2 - \widehat{P}_1$ give us a way to quantify the likelihood that samples are drawn from similar distributions.

Checking for Identically Distributed

Because $\widehat{P}_1$ and $\widehat{P}_2$ are step functions, we see that

$$\|\widehat{P}_2 - \widehat{P}_1\|_{\mathrm{KS}} = \max\left\{|\widehat{P}_2(R) - \widehat{P}_1(R)| \,:\, R \in (-1,\infty)\right\}.$$

$$\begin{aligned}\|\widehat{P}_2 - \widehat{P}_1\|_{\mathrm{Ku}} = {} & \max\left\{\widehat{P}_2(R) - \widehat{P}_1(R) \,:\, R \in (-1,\infty)\right\} \\ & + \max\left\{\widehat{P}_1(R) - \widehat{P}_2(R) \,:\, R \in (-1,\infty)\right\}.\end{aligned}$$

Fortunately statisticians have provided software that efficiently computes these values given any two samples $\{r_1(d)\}_{d=1}^{D_1}$ and $\{r_2(d)\}_{d=1}^{D_2}$. These are called respectively the *two-sample KS test* and the *two-sample Kuiper test*.

Checking for Identically Distributed

Finally, given return histories over a year $\{r(d)\}_{d=1}^{D}$, we can split the year into quarters and compare the emperical distribution of each quarter with that of another quarter or with that of the other three quarters combined. The maximum of all such comparisons made is the score for the year. For example, for each year we might define

$$\omega_{\mathrm{KS}} = 1 - \max\left\{ \|\widehat{P}_2 - \widehat{P}_1\|_{\mathrm{KS}} \;:\; \text{all comparisons made} \right\},$$
$$\omega_{\mathrm{Ku}} = 1 - \max\left\{ \|\widehat{P}_2 - \widehat{P}_1\|_{\mathrm{Ku}} \;:\; \text{all comparisons made} \right\}.$$

If we choose to compare quarters with each other then six comparisons are made. If we choose to compare each quarter with the other three quarters combined then four comparisons are made. Notice that $\omega_{\mathrm{Ku}} \leq \omega_{\mathrm{KS}} \leq 1$, and that the distributions are closer when ω_{Ku} is nearer 1.

Checking for Identically Distributed

Remark. These measures can be applied to any risky asset. We might see a difference between the stock of a single company and a broad-based index fund. Similarly, the asset could be a tangent portfolio for a class of portfolios with a given leverage limit, for example, long portfolios. Reacall that given a return history $\{\mathbf{r}(d)\}_{d=0}^{D}$ of N risky assests and a risk-free rate model μ_{rf} the return of the Markowitz portfolio with allocation $\mathbf{f}$ on day d is

$$r(d) = (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{r}(d)^{\mathrm{T}}\mathbf{f}\,.$$

Stationary Autoregression Models

One way to quantify how well a return history $\{r(d)\}_{d=1}^{D}$ is mimicked by an IID model is to fit it to a more complicated model and then measure how far that fit is from an IID model. We illustrate this approach using the family of *stationary autoregression models*. These models have the form

$$R_d = a + b\, R_{d-1} + Z_d \quad \text{for } d = 1, \cdots, D\,, \tag{6.18}$$

where a and b are real numbers, R_0 is a random variable and $\{Z_d\}_{d=1}^{\infty}$ is a sequence of IID random variable with mean zero.

The model is called *stationary* when for every $d \in \{1, \cdots, \infty\}$ the random variable R_d has the same statistical behavior as R_0. We will see that stationarity implies that $|b| < 1$.

Stationary Autoregression Models

Let μ and ξ be the mean and variance of the random variable R_0. Then stationarity implies that

$$\mathrm{Ex}(R_d) = \mu\,, \quad \mathrm{Var}(R_d) = \xi\,, \quad \text{for every } d \in \{0, \cdots, \infty\}\,. \tag{6.19a}$$

Let ξ_d denote the covariance of R_d with R_0, so that

$$\xi_d = \mathrm{Cov}(R_0, R_d) = \mathrm{Ex}\big((R_0 - \mu)\,(R_d - \mu)\big) \quad \text{for every } d \in \{0, \cdots, \infty\}\,. \tag{6.19b}$$

(Notice that $\xi_0 = \xi$.) Then stationarity implies that

$$\mathrm{Cov}(R_d, R_{d'}) = \mathrm{Ex}\big((R_d - \mu)\,(R_{d'} - \mu)\big) = \xi_{|d-d'|}\,, \quad \text{for every } d, d' \in \{0, \cdots, \infty\}\,. \tag{6.19c}$$

Stationary Autoregression Models

Let η be the variance of the IID mean-zero variables Z_d. Then

$$\mathrm{Ex}(Z_d) = 0\,, \quad \mathrm{Var}(Z_d) = \eta\,, \quad \text{for every } d \in \{1, \cdots, \infty\}\,. \tag{6.20a}$$

Because the random variables $\{Z_d\}_{d=1}^{\infty}$ are IID, we have

$$\begin{gathered}\mathrm{Cov}(Z_d, Z_{d'}) = \mathrm{Ex}(Z_d\, Z_{d'}) = 0\,, \\ \text{for every } d, d' \in \{1, \cdots, \infty\} \text{ with } d \neq d'\,.\end{gathered} \tag{6.20b}$$

Because the random variable R_0 is independent of each Z_d, we have

$$\begin{gathered}\mathrm{Cov}(R_0, Z_d) = \mathrm{Ex}((R_0 - \mu)\, Z_d) = 0\,, \\ \text{for every } d \in \{1, \cdots, \infty\}\,.\end{gathered} \tag{6.20c}$$

Stationary Autoregression Models

Because each Z_d has mean zero, by taking expected values in (6.18) while using (6.20) we see that

$$\mu = \mathrm{Ex}(R_d) = a + b\,\mathrm{Ex}(R_{d-1}) + \mathrm{Ex}(Z_d) = a + b\mu\,.$$

Therefore a, b, and μ are related by

$$\mu = a + b\mu\,. \tag{6.21}$$

By using this relation to eliminate a from the form (6.18), we obtain

$$R_d = \mu + b\,(R_{d-1} - \mu) + Z_d \quad \text{for } d = 1, \cdots, \infty\,,$$

which can be recast as

$$R_d - \mu = b\,(R_{d-1} - \mu) + Z_d \quad \text{for } d = 1, \cdots, \infty\,. \tag{6.22}$$

Stationary Autoregression Models

Multiplying (6.22) by $Z_{d'}$ and taking expected values we obtain

$$\mathrm{Ex}\big((R_d - \mu)\, Z_{d'}\big) = b\, \mathrm{Ex}\big((R_{d-1} - \mu)\, Z_{d'}\big) + \mathrm{Ex}\big(Z_d\, Z_{d'}\big)\,, \quad \text{for every } d, d' \in \{1, \cdots, \infty\}\,. \tag{6.23}$$

By using (6.20b) we see from (6.23) that

$$\mathrm{Ex}\big((R_d - \mu)\, Z_{d'}\big) = b\, \mathrm{Ex}\big((R_{d-1} - \mu)\, Z_{d'}\big)\,, \quad \text{for every } d, d' \in \{1, \cdots, \infty\} \text{ with } d < d'\,.$$

Then by using (6.20c) we can prove by induction that

$$\mathrm{Cov}(R_d, Z_{d'}) = \mathrm{Ex}\big((R_d - \mu)\, Z_{d'}\big) = 0\,, \quad \text{for every } d, d' \in \{0, \cdots, \infty\} \text{ with } d < d'\,. \tag{6.24}$$

Stationary Autoregression Models

By squaring (6.22) and taking expected values while using (6.19), (6.20), and (6.24), we see that

$$\begin{aligned}
\xi = \mathrm{Var}(R_d) &= \mathrm{Ex}\Big((R_d - \mu)^2\Big) = \mathrm{Ex}\Big(\big(b\,(R_{d-1} - \mu) + Z_d\big)^2\Big) \\
&= b^2 \mathrm{Ex}\Big((R_{d-1} - \mu)^2\Big) + 2b\,\mathrm{Ex}((R_{d-1} - \mu)\,Z_d) + \mathrm{Ex}\Big(Z_d^2\Big) \\
&= b^2 \mathrm{Var}(R_{d-1}) + \mathrm{Var}(Z_d) = b^2 \xi + \eta\,.
\end{aligned}$$

Therefore b, ξ, and η are related by

$$(1 - b^2)\xi = \eta\,. \tag{6.25}$$

Because the variances ξ and η are positive, we see that

$$b^2 < 1\,, \qquad \eta \leq \xi\,.$$

Notice that if $b = 0$ then $\xi = \eta$ and the stationary autoregression model (6.22) reduces to an IID model.

Stationary Autoregression Models

By multiplying (6.22) by $(R_0 - \mu)$ and taking expected values while using (6.19b) and (6.20c) we see that

$$\begin{aligned}\xi_d &= \mathrm{Ex}\big((R_0 - \mu)\,(R_d - \mu)\big) \\ &= b\,\mathrm{Ex}\big((R_0 - \mu)\,(R_{d-1} - \mu)\big) + \mathrm{Ex}\big((R_0 - \mu)\,Z_d\big) \\ &= b\,\xi_{d-1}\,.\end{aligned}$$

Because $\xi_0 = \xi$, by induction we can show that

$$\xi_d = \xi\, b^d \quad \text{for every } d \in \{1, \cdots, \infty\}\,. \tag{6.26}$$

Because $|b| < 1$, we see that ξ_d decays as d increases.

Stationary Autoregression Models

By setting $d' = d$ in (6.23) while using (6.19a) and (6.24) we obtain

$$\mathrm{Cov}(R_d, Z_d) = \mathrm{Var}(Z_d) = \eta\,, \quad \text{for every } d \in \{1, \cdots, \infty\}\,. \tag{6.27}$$

By using (6.20b) we see from (6.23) that

$$\mathrm{Ex}\big((R_d - \mu)\, Z_{d'}\big) = b\, \mathrm{Ex}\big((R_{d-1} - \mu)\, Z_{d'}\big)\,, \quad \text{for every } d, d' \in \{1, \cdots, \infty\} \text{ with } d' < d\,.$$

Then by using (6.27) we can prove by induction that

$$\mathrm{Cov}(R_d, Z_{d'}) = \mathrm{Ex}\big((R_d - \mu)\, Z_{d'}\big) = \eta\, b^{d-d'}\,, \quad \text{for every } d, d' \in \{1, \cdots, \infty\} \text{ with } d' \le d\,. \tag{6.28}$$

Because $|b| < 1$, we see that $\mathrm{Cov}(R_d, Z_{d'})$ decays as d increases.

Stationary Autoregression Models

The *correlation time* d_c of the stationary autoregression model (6.18) is defined by

$$\frac{1}{d_c} = \log\left(\frac{1}{|b|}\right) , \tag{6.29}$$

so that

$$|\xi_d| = \xi \exp\left(-\frac{d}{d_c}\right) , \quad \text{for every } d \in \{0, \cdots, \infty\} ,$$

and

$$|\mathrm{Cov}(R_d, Z_{d'})| = \eta \exp\left(-\frac{d}{d_c}\right) ,$$

for every $d, d' \in \{1, \cdots, \infty\}$ with $d' \leq d$.

The smaller d_c the closer the stationary autoregression model is to an IID model.

Stationary Autoregression Models

We have seen that a stationary autoregression model in the form (6.18) is specified by three parameters. These can be $a \in \mathbb{R}$, $b \in (-1, 1)$, and $\eta > 0$, in which case μ, ξ, and ξ_1 are given by

$$\mu = \frac{a}{1 - b}\,, \qquad \xi = \frac{\eta}{1 - b^2}\,, \qquad \xi_1 = \frac{\eta\, b}{1 - b^2}\,.$$

Alternatively, they can be $\mu \in \mathbb{R}$, $\xi > 0$, and $\xi_1 \in (-\xi, \xi)$, in which case a, b, and η are given by

$$a = \left(1 - \frac{\xi_1}{\xi}\right)\mu\,, \qquad b = \frac{\xi_1}{\xi}\,, \qquad \eta = \xi - \frac{{\xi_1}^2}{\xi}\,.$$

In the next section we will show how to pick the parameters to best fit a given data set.

Fitting Stationary Autoregression Models

Given a return history $\{r(d)\}_{d=0}^{D}$ and a choice of positive weights $\{w_d\}_{d=1}^{D}$ that sum to 1 we can use least squares to fit a stationary autoregression model of the form (6.18). Specifically, this approach constructs estmators $\hat{a}$ and $\hat{b}$ such

$$(\hat{a}, \hat{b}) = \arg\min \left\{ \sum_{d=1}^{D} w_d |r(d) - a - b\, r(d-1)|^2 \right\} , \tag{7.30}$$

and then construct the estmator $\hat{\eta}$ by

$$\begin{aligned} \hat{\eta} &= \min \left\{ \sum_{d=1}^{D} w_d |r(d) - a - b\, r(d-1)|^2 \right\} \\ &= \sum_{d=1}^{D} w_d |r(d) - \hat{a} - \hat{b}\, r(d-1)|^2 . \end{aligned} \tag{7.31}$$

Fitting Stationary Autoregression Models

It is helpful to define the return mean estimators

$$\hat{m}_0 = \sum_{d=1}^{D} w_d \, r(d) \,, \qquad \hat{m}_1 = \sum_{d=1}^{D} w_d \, r(d-1) \,, \tag{7.32a}$$

the return variance estimators

$$\hat{v}_{00} = \sum_{d=1}^{D} w_d \big(r(d) - \hat{m}_0 \big)^2 \,, \qquad \hat{v}_{11} = \sum_{d=1}^{D} w_d \big(r(d-1) - \hat{m}_1 \big)^2 \,, \tag{7.32b}$$

and the return autocovariance estimator

$$\hat{v}_{10} = \sum_{d=1}^{D} w_d \big(r(d-1) - \hat{m}_0 \big) \big(r(d) - \hat{m}_1 \big) \,. \tag{7.32c}$$

It is also helpful to replace a with $\tilde{a}$ that is defined by

$$a = \hat{m}_0 - b \, \hat{m}_1 + \tilde{a} \,. \tag{7.33}$$

Fitting Stationary Autoregression Models

Then

$$
\begin{aligned}
z(d) &= r(d) - a - b\, r(d-1) \\
&= \big(r(d) - \hat{m}_0\big) - b\,\big(r(d-1) - \hat{m}_1\big) + \tilde{a} \\
&= \tilde{r}_0(d) - b\tilde{r}_1(d) + \tilde{a}\,,
\end{aligned}
$$

where we define

$$
\tilde{r}_0(d) = r(d) - \hat{m}_0\,, \qquad \tilde{r}_1(d) = r(d-1) - \hat{m}_1\,. \tag{7.34}
$$

Therefore

$$
\begin{aligned}
|z(d)|^2 = |\tilde{r}_0(d)|^2 + b^2|\tilde{r}_1(d)|^2 + \tilde{a}^2 \\
- 2b\,\tilde{r}_1(d)\,\tilde{r}_0(d) + 2\tilde{a}\,\tilde{r}_0(d) - 2\tilde{a}b\,\tilde{r}_1(d)\,.
\end{aligned} \tag{7.35}
$$

Fitting Stationary Autoregression Models

It is evident from (7.32) and (7.34) that $\{\tilde{r}_0(d)\}_{d=1}^D$ and $\{\tilde{r}_1(d)\}_{d=1}^D$ satisfy

$$\sum_{d=1}^D w_d \tilde{r}_0(d) = 0\,, \qquad \sum_{d=1}^D w_d \tilde{r}_1(d) = 0\,,$$

$$\sum_{d=1}^D w_d |\tilde{r}_0(d)|^2 = \hat{v}_{00}\,, \quad \sum_{d=1}^D w_d |\tilde{r}_1(d)|^2 = \hat{v}_{11}\,,$$

$$\sum_{d=1}^D w_d \tilde{r}_1(d)\, \tilde{r}_0(d) = \hat{v}_{10}\,.$$

By using these facts we see from (7.35) that

$$\sum_{d=1}^D w_d |z(d)|^2 = \hat{v}_{00} + b^2 \hat{v}_{11} + \tilde{a}^2 - 2b\, \hat{v}_{10}\,.$$

Fitting Stationary Autoregression Models

Because $\hat{v}_{11} > 0$, the foregoing quantity is clearly minimized when

$$\tilde{a} = 0 \,, \qquad b = \frac{\hat{v}_{10}}{\hat{v}_{11}} \,,$$

and that

$$\min\left\{ \sum_{d=1}^{D} w_d |z(d)|^2 \right\} = \hat{v}_{00} - \frac{\hat{v}_{10}^{\,2}}{\hat{v}_{11}} \,.$$

Recalling (7.30), (7.31), and (7.33), this suggests using the estimators

$$\hat{a} = \hat{m}_0 - \frac{\hat{v}_{10}}{\hat{v}_{11}} \hat{m}_1 \,, \qquad \hat{b} = \frac{\hat{v}_{10}}{\hat{v}_{11}} \,, \qquad \hat{\eta} = \hat{v}_{00} - \frac{\hat{v}_{10}^{\,2}}{\hat{v}_{11}} \,. \tag{7.36}$$

Fitting Stationary Autoregression Models

However, the estimators (7.36) given by the least squares fit have a problem. Specifically, the formula for $\hat{b}$ can give values that lie outside of the interval $(-1, 1)$. So rather than use the estimators (7.36), we will use the estimators

$$\hat{a} = \hat{m}_0 - \frac{\hat{v}_{10}}{\hat{v}_{11}}\,\hat{m}_1\,, \qquad \hat{b} = \frac{\hat{v}_{10}}{\sqrt{\hat{v}_{00}\,\hat{v}_{11}}}\,, \qquad \hat{\eta} = \hat{v}_{00} - \frac{\hat{v}_{10}^{\,2}}{\hat{v}_{11}}\,. \tag{7.37}$$

These estimators will satisfy $\hat{b} \in (-1, 1)$ and $\hat{\eta} > 0$ if and only if the *autocovariance matrix* $\widehat{V}$ is positive definite, where

$$\widehat{V} = \begin{pmatrix} \hat{v}_{00} & \hat{v}_{10} \\ \hat{v}_{10} & \hat{v}_{11} \end{pmatrix}\,. \tag{7.38}$$

This condition is always met in practice.

Fitting Stationary Autoregression Models

Notice that the last two estimators in (7.36) satisfy

$$\hat{\eta} = \hat{v}_{00}\left(1 - \hat{b}^2\right) .$$

Because $\hat{v}_{00}$ is the sample variance of $\{r(d)\}_{d=1}^{D}$ while $\hat{\eta}$ is the sample variance of $\{z(d)\}_{d=1}^{D}$, we see that $\hat{b}^2$ is the fraction of the sample variance of $\{r(d)\}_{d=1}^{D}$ that is due to the autoregression. This suggests that a natural measure of how well the history $\{r(d)\}_{d=1}^{D}$ can be mimicked by an IID model is

$$\omega_{\mathrm{ar}} = 1 - \hat{b}^2 = 1 - \frac{\hat{v}_{10}^2}{\hat{v}_{00}\,\hat{v}_{11}} . \tag{7.39}$$

The closer ω_{ar} is to 1, the better the IID model.

Fitting Stationary Autoregression Models

Remark. Given a return history $\{r(d)\}_{d=0}^{D}$ of any market index, we can use the autoregression estimator $\hat{b}$ given by (7.37) to estimate a correlation time for that index. Motivated by formula (6.29), we define $\hat{d}_c$ by

$$\frac{1}{\hat{d}_c} = \log\left(\frac{1}{|\hat{b}|}\right) . \tag{7.40}$$

Because the history has length D, we would like $\hat{d}_c \ll D$ in order to have some confidence in our estimators of the return mean μ and the return variance ξ.

Checking for Independence

In an IID model the random numbers $\{R_d\}_{d=1}^{D}$ are drawn from $(-1, \infty)$ in accord with the probability density $q(R)$ *independent* of each other, there is no correlation of R_d with $R_{d'}$ when $d \neq d'$. We would like to check how well a return history $\{r(d)\}_{d=1}^{D}$ is mimiced by such a model.

Because there is no correlation of R_d with $R_{d'}$ when $d \neq d'$, if we plot the points $\{(R_d, R_{d+c})\}_{d=1}^{D-c}$ in the rr'-plane for any $c > 0$ they will be distributed in accord with the probability density $q(R)q(R')$. *Therefore if the return history $\{r(d)\}_{d=1}^{D}$ is mimiced by such a model then the points $\{(r(d), r(d+c))\}_{d=1}^{D-c}$ plotted in the rr'-plane should appear to be distributed in a way consistant with the probability density $q(r)q(r')$.* Such plots are called *scatter plots.*

We expect that the strongest correlation should be seen when $c = 1$ because the behavior of an asset price on any given trading day often does correlate with its behavior on the previous trading day.

Checking for Independence

Given a return history $\{r(d)\}_{d=0}^{D}$ and a choice of positive weights $\{w_d\}_{d=1}^{D}$ that sum to 1, we define the return mean estimators

$$\hat{m}_0 = \sum_{d=1}^{D} w_d r(d) \,, \qquad \hat{m}_1 = \sum_{d=1}^{D} w_d r(d-1) \,,$$

the return variance estimators

$$\hat{v}_{00} = \sum_{d=1}^{D} w_d \big(r(d) - \hat{m}_0\big)^2 \,, \qquad \hat{v}_{11} = \sum_{d=1}^{D} w_d \big(r(d-1) - \hat{m}_1\big)^2 \,,$$

and the return autocovariance estimator

$$\hat{v}_{10} = \sum_{d=1}^{D} w_d \big(r(d-1) - \hat{m}_0\big)\big(r(d) - \hat{m}_1\big) \,.$$

This is often done with uniform weights $w_d = 1/D$.

Checking for Independence

Consider the 2 × 2 *autocovariance matrix*

$$\widehat{V} = \begin{pmatrix} \hat{v}_{00} & \hat{v}_{10} \\ \hat{v}_{10} & \hat{v}_{11} \end{pmatrix} . \tag{8.41}$$

This matrix is symmetric and is usually positive definite. If the data was drawn from an IID process with mean μ and variance ξ then it can be shown that

$$\mathrm{Ex}\left(\widehat{V}\right) = \xi W , \quad \text{where} \quad W = \begin{pmatrix} 1 - \bar{w} & -\bar{w}_1 \\ -\bar{w}_1 & 1 - \bar{w} \end{pmatrix} , \tag{8.42}$$

with

$$\bar{w} = \sum_{d=1}^{D} w_d^2 , \qquad \bar{w}_1 = \sum_{d=2}^{D} w_d \, w_{d-1} .$$

Checking for Independence

The matrix W is known. For uniform weights $w_d = 1/D$ we have

$$\bar{w} = \frac{1}{D}\,, \qquad \bar{w}_1 = \frac{D-1}{D^2}\,,$$

whereby

$$W = \begin{pmatrix} 1 - \frac{1}{D} & -\frac{D-1}{D^2} \\ -\frac{D-1}{D^2} & 1 - \frac{1}{D} \end{pmatrix}.$$

It can be shown for $D > 1$ that in general we have

$$0 < \bar{w}_1 < \bar{w}\,, \qquad \bar{w} + \bar{w}_1 < 1\,, \tag{8.43}$$

which implies that the symmetric matrix W given by (8.42) is always *diagonally dominant* and thereby is always *positive definite*.

Checking for Independence

The deviation of $\widehat{V}$ given by (8.41) from the form (8.42) measures of how well an IID model mimics the data. For example, its size can be measured with the Frobenius norm, which for any real matrix A is determined by

$$\|A\|_{\mathrm{F}}^2 = \mathrm{tr}\big(A^{\mathrm{T}} A\big) .$$

We first estimate ξ in the form (8.42) to give the best least squares fit with respect to this norm. In other words, we set

$$\hat{\xi} = \arg\min \Big\{ \mathrm{tr}\big((\widehat{V} - \xi W)^2\big) \Big\}$$

Because

$$\mathrm{tr}\big((\widehat{V} - \xi W)^2\big) = \mathrm{tr}\big(\widehat{V}^2\big) - 2\xi \, \mathrm{tr}\big(W \, \widehat{V}\big) + \xi^2 \, \mathrm{tr}\big(W^2\big) ,$$

we see that

$$\hat{\xi} = \frac{\mathrm{tr}\big(W \, \widehat{V}\big)}{\mathrm{tr}\big(W^2\big)} . \tag{8.44}$$

Checking for Independence

When the estimator $\hat{\xi}$ is expressed in terms of the entries of the matrices $\widehat{V}$ and W given by (8.41) and (8.42) we have

$$\hat{\xi} = \frac{(1-\bar{w})\big(\hat{v}_{00}+\hat{v}_{11}\big) - 2\bar{w}_1\hat{v}_{10}}{2\big((1-\bar{w})^2+\bar{w}_1{}^2\big)} .$$

The fact that $\hat{\xi} > 0$ whenever $\widehat{V} \neq 0$ is can be seen directly from (8.44) and the following general fact, the proof of which is left as an exercise.

Fact. If A and B are symmetric matrices of the same size such that A is positive definite, B is nonnegative definite, and $B \neq 0$ then $\mathrm{tr}(AB) > 0$.

Moreover, it is evident from (8.42) and (8.44) that

$$\mathrm{Ex}(\hat{\xi}) = \frac{\mathrm{tr}\big(W\,\mathrm{Ex}(\widehat{V})\big)}{\mathrm{tr}\big(W^2\big)} = \frac{\mathrm{tr}\big(\xi\, W^2\big)}{\mathrm{tr}\big(W^2\big)} = \xi .$$

Therefore $\hat{\xi}$ is an unbiased estimator of ξ.

Checking for Independence

The size of the deviation of $\widehat{V}$ given by (8.41) from the form (8.42) is thereby quantified by

$$\frac{\|\widehat{V} - \hat{\xi} W\|_{\mathrm{F}}^2}{\|\widehat{V}\|_{\mathrm{F}}^2} = 1 - \frac{\mathrm{tr}\big(W\,\widehat{V}\big)^2}{\mathrm{tr}\big(\widehat{V}^2\big)\;\mathrm{tr}\big(W^2\big)}\,.$$

Therfore we defined the measure

$$\omega_{\mathrm{ind}} = \frac{\mathrm{tr}\big(W\,\widehat{V}\big)^2}{\mathrm{tr}\big(\widehat{V}^2\big)\;\mathrm{tr}\big(W^2\big)}\,. \tag{8.45}$$

This is the square of the cosine of the angle between $\widehat{V}$ and W as determined by the Frobenius scalar product. The closer ω_{ind} is to 1, the better an IID model mimics the data.

Checking for Independence

Remark. From (8.45) we can show by using (8.41) and (8.42) that

$$1 - \omega_{\text{ind}} = \delta^2 + \left(1 - \delta^2\right) \cos(\phi)^2 \,,$$

where

$$\delta^2 = \frac{\left(\hat{v}_{00} - \hat{v}_{11}\right)^2}{\left(\hat{v}_{00} - \hat{v}_{11}\right)^2 + \left(\hat{v}_{00} + \hat{v}_{11}\right)^2 + 4\hat{v}_{10}^2} \,,$$

$$\cos(\phi)^2 = \frac{\left(2(1 - \bar{w})\hat{v}_{10} + \hat{w}_1\left(\hat{v}_{00} + \hat{v}_{11}\right)\right)^2}{\left((1 - \bar{w})^2 + \bar{w}_1^2\right)\left(\left(\hat{v}_{00} + \hat{v}_{11}\right)^2 + 4\hat{v}_{10}^2\right)} \,.$$

This shows that ω_{ind} is close to 1 if and only if δ and $\cos(\phi)$ are small. The first condition holds if and only if $\hat{v}_{00}$ and $\hat{v}_{11}$ are relatively close. The second holds if and only if the vectors $(1 - \bar{w}, \hat{w}_1)$ and $(2\hat{v}_{10}, \hat{v}_{00} + \hat{v}_{11})$ are nearly orthogonal.