Portfolios that Contain Risky Assets
7.2. Assessing Independence for IID Models

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*
May 9, 2022 version

Portfolios that Contain Risky Assets
Part II: Probabilistic Models

Portfolios that Contain Risky Assets
Part II: Probabilistic Models

7. Assessing Independent, Identically-Distributed Models

Assessing Independence for IID Models

Introduction

Independent, Identically Distributed (IID) models of returns make two simplifying assumptions.

1. **Independent.** That what happens on day *d* is independent of what has happened in the past.
2. **Identically Distributed.** What happens each day is statistically identical to what happens every other day.

In IID models the random numbers $\{R_d\}_{d=1}^D$ that mimic a return history are each drawn from $(-1, \infty)$ in accord with the *same* probability density.

The question arises as to how well a given return history $\{r(d)\}_{d=1}^D$ is mimiced by such a model. We will present ways by which the validity of each assumption can be assessed.

Introduction

Earlier we examined how to assess the validity of the **identically distributed** assumption. This came down to understanding how likely it is that two different return histories, say $\{r_1(d)\}_{d=1}^{D_1}$ and $\{r_2(d)\}_{d=1}^{D_2}$, might be drawn from the same probability density. We took three approaches:

- graphical,
- comparing means and variances,
- comparing distributions.

Here we will examine how to assess the validity of the **independent** assumption. This comes down to understanding how correlated each $r(d)$ is with earlier values, say with $r(d-1)$. We will take three approaches:

- graphical,
- comparing with an autoregressive model,
- comparing autocovariance matrices.

Stationary Autoregressive Models (Introduction)

Stationary Autoregressive 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 autoregressive models*. These models have the form

$$R_d = a + b\,R_{d-1} + Z_d \quad \text{for } d \in \{1, 2, \cdots\}\,, \tag{2.1}$$

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 variables with mean zero.

Definition. An autoregressive model in the form (2.1) is called *stationary* when the statistical behavior of the random variables is translation invariant in d.

Remark. We will see that stationarity implies that $|b| < 1$.

Stationary Autoregressive Models (η, μ and ξ)

Let $\eta > 0$ be the variance of the IID mean-zero variables Z_d. Then

$$\mathrm{Ex}(Z_d) = 0\,, \quad \mathrm{Vr}(Z_d) = \eta\,, \quad \text{for every } d \in \{1, 2, \cdots\}\,. \tag{2.2a}$$

Because the random variables $\{Z_d\}_{d=1}^{\infty}$ are mean-zero and are IID, we have the covariance formula

$$\begin{gathered} \mathrm{Cv}(Z_d, Z_{d'}) = \mathrm{Ex}(Z_d\, Z_{d'}) = 0\,, \\ \text{for every } d, d' \in \{1, 2, \cdots\} \text{ with } d \neq d'\,. \end{gathered} \tag{2.2b}$$

Let $\mu \in \mathbb{R}$ and $\xi > 0$ be the mean and variance of the random variable R_0. Then stationarity implies that

$$\mathrm{Ex}(R_d) = \mu\,, \quad \mathrm{Vr}(R_d) = \xi\,, \quad \text{for every } d \in \{0, 1, \cdots\}\,. \tag{2.2c}$$

Stationary Autoregressive Models (Main Result)

The mean-variance statistics of stationary autoregressive models in the form (2.1) are specified by just three parameters. Specifically, we will prove the following.

Fact 1. The parameters a, b, μ, ξ, and η satisfy the relations

$$\mu = a + b\,\mu\,, \qquad (1 - b^2)\,\xi = \eta\,, \qquad b^2 < 1\,. \tag{2.3a}$$

If $b \neq 0$ then for every $d, d' \in \{0, 1, \cdots\}$ with $d' \geq 1$ we have the formula

$$\mathrm{Cv}(R_d, Z_{d'}) = \begin{cases} 0 & \text{for } d < d'\,, \\ \eta\, b^{d-d'} & \text{for } d \geq d'\,. \end{cases} \tag{2.3b}$$

If $b \neq 0$ then for every $d, d' \in \{0, 1, \cdots\}$ we have the formula

$$\mathrm{Cv}(R_d, R_{d'}) = \xi\, b^{|d-d'|}\,. \tag{2.3c}$$

Stationary Autoregressive Models (Remarks)

Remarks. Before proving **Fact 1**, we make some remarks about it.

- If $b = 0$ then $\mu = a$ and $\xi = \eta$ and the stationary autoregressive model (2.1) reduces to the IID model with $R_d = a + Z_d$.
- The fact that $\mathrm{Cv}(R_d, Z_{d'})$ and $\mathrm{Cv}(R_d, R_{d'})$ given by formulas (2.3b) and (2.3c) are functions of $d - d'$ is a consequence of the stationarity of the model.
- Because $|b| < 1$, we see that both $\mathrm{Cv}(R_d, Z_{d'})$ and $\mathrm{Cv}(R_d, R_{d'})$ decay as $|d - d'|$ increases.
- The fact that $\mathrm{Cv}(R_d, Z_{d'}) = 0$ for $d < d'$ reflects the fact that R_d is independent of any future $Z_{d'}$.
- The fact that $\mathrm{Cv}(R_d, Z_{d'}) \neq 0$ for $d \geq d'$ reflects the fact that R_d is dependent upon any present or past $Z_{d'}$ when $b \neq 0$.

Stationary Autoregressive Models (Mean Relation)

Proof. By taking expected values of the form (2.1) while using the facts from (2.2) that $\mathrm{Ex}(R_d) = \mathrm{Ex}(R_{d-1}) = \mu$ and $\mathrm{Ex}(Z_d) = 0$ we obtain

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

Therefore a, b, and μ satisfy the first relation in (2.2a), which is

$$\mu = a + b\,\mu\,. \tag{2.4}$$

By using this relation to eliminate a from the form (2.1) we obtain

$$\widetilde{R}_d = b\,\widetilde{R}_{d-1} + Z_d\,, \qquad \text{for } d = 1, 2, \cdots\,, \tag{2.5a}$$

where

$$\widetilde{R}_d = R_d - \mu\,. \tag{2.5b}$$

Stationary Autoregressive Models ($\mathrm{Cv}(R_d, Z_{d'})$)

By multiplying (2.5a) by $Z_{d'}$ and taking expected values we obtain

$$\mathrm{Ex}\left(\widetilde{R}_d\, Z_{d'}\right) = b\,\mathrm{Ex}\left(\widetilde{R}_{d-1}\, Z_{d'}\right) + \mathrm{Ex}(Z_d\, Z_{d'})\,, \quad \text{for every } d, d' \in \{1, 2, \cdots\}\,. \tag{2.6}$$

Because the random variable R_0 is independent of each $Z_{d'}$, we have

$$\mathrm{Ex}\left(\widetilde{R}_0\, Z_{d'}\right) = 0\,, \qquad \text{for every } d' \in \{1, 2, \cdots\}\,. \tag{2.7}$$

Because $\mathrm{Ex}(Z_d\, Z_{d'}) = 0$ for $d \neq d'$ by (2.2b), we see from (2.6) that

$$\mathrm{Ex}\left(\widetilde{R}_d\, Z_{d'}\right) = b\,\mathrm{Ex}\left(\widetilde{R}_{d-1}\, Z_{d'}\right)\,, \qquad \text{for every } d < d'\,.$$

Then an induction argument on d initialized by (2.7) proves that

$$\mathrm{Cv}(R_d, Z_{d'}) = \mathrm{Ex}\left(\widetilde{R}_d\, Z_{d'}\right) = 0\,, \quad \text{for every } d, d' \in \{0, 1 \cdots\} \text{ with } d < d'\,. \tag{2.8}$$

Stationary Autoregressive Models ($\mathrm{Cv}(R_d, Z_{d'})$)

By setting $d = d'$ in (2.6) while using (2.2a) and (2.8) we obtain

$$\mathrm{Ex}\Big(\widetilde{R}_{d'}\, Z_{d'}\Big) = \mathrm{Vr}(Z_{d'}) = \eta\,, \qquad \text{for every } d' \in \{1, 2, \cdots\}\,. \tag{2.9}$$

Because $\mathrm{Ex}(Z_d\, Z_{d'}) = 0$ for $d \neq d'$ by (2.2b), we see from (2.6) that

$$\mathrm{Ex}\Big(\widetilde{R}_d\, Z_{d'}\Big) = b\,\mathrm{Ex}\Big(\widetilde{R}_{d-1}\, Z_{d'}\Big)\,, \qquad \text{for every } d > d'\,.$$

Then an induction argument on d initialized by (2.9) proves that

$$\begin{gathered}\mathrm{Cv}(R_d, Z_{d'}) = \mathrm{Ex}\Big(\widetilde{R}_d\, Z_{d'}\Big) = \eta\, b^{d-d'}\,, \\ \text{for every } d, d' \in \{1, 2, \cdots\} \text{ with } d' \leq d\,.\end{gathered} \tag{2.10}$$

Combining (2.8) with (2.10) proves formula (2.3b).

Stationary Autoregressive Models (Variance Relation)

By squaring (2.5a) and taking expected values while using (2.2) and (2.8), we obtain

$$\begin{aligned}\xi = \mathrm{Vr}(R_d) &= \mathrm{Ex}\Big(\widetilde{R}_d^2\Big) = \mathrm{Ex}\Big(\Big(b\,\widetilde{R}_{d-1} + Z_d\Big)^2\Big) \\ &= b^2\mathrm{Ex}\Big(\widetilde{R}_{d-1}^2\Big) + 2b\,\mathrm{Ex}\Big(\widetilde{R}_{d-1}\,Z_d\Big) + \mathrm{Ex}\Big(Z_d^2\Big) \\ &= b^2\mathrm{Vr}(R_{d-1}) + \mathrm{Vr}(Z_d) = b^2\xi + \eta\,.\end{aligned}$$

Therefore b, ξ, and η are related by

$$(1 - b^2)\,\xi = \eta\,. \tag{2.11a}$$

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

$$b^2 < 1\,, \qquad 0 < \eta \le \xi\,. \tag{2.11b}$$

Combining (2.4) with (2.11) proves relations (2.3a).

Stationary Autoregressive Models ($\mathrm{Cv}(R_d, R_{d'})$)

By multiplying (2.5a) by $\widetilde{R}_{d'}$ and taking expected values we obtain

$$\mathrm{Ex}\Big(\widetilde{R}_d\,\widetilde{R}_{d'}\Big) = b\,\mathrm{Ex}\Big(\widetilde{R}_{d-1}\,\widetilde{R}_{d'}\Big) + \mathrm{Ex}\Big(Z_d\,\widetilde{R}_{d'}\Big)\,, \quad \text{for every } d \in \{1,2,\cdots\} \text{ and } d' \in \{0,1,\cdots\}\,. \tag{2.12}$$

We know from (2.2c) that

$$\mathrm{Ex}\Big(\widetilde{R}_{d'}\,\widetilde{R}_{d'}\Big) = \mathrm{Vr}(R_{d'}) = \xi\,, \qquad \text{for every } d' \in \{0,1,\cdots\}\,. \tag{2.13}$$

Because $\mathrm{Ex}\Big(Z_d\,\widetilde{R}_{d'}\Big) = 0$ for $d > d'$ by (2.8), we see from (2.12) that

$$\mathrm{Ex}\Big(\widetilde{R}_d\,\widetilde{R}_{d'}\Big) = b\,\mathrm{Ex}\Big(\widetilde{R}_{d-1}\,\widetilde{R}_{d'}\Big)\,, \qquad \text{for every } d > d'\,.$$

Then an induction argument on d initialized by (2.13) proves that

$$\mathrm{Cv}(R_d, R_{d'}) = \mathrm{Ex}\Big(\widetilde{R}_d\,\widetilde{R}_{d'}\Big) = \xi\, b^{d-d'}\,, \quad \text{for every } d \in \{0,1,\cdots\} \text{ with } d \geq d'\,. \tag{2.14}$$

Stationary Autoregressive Models $(\mathrm{Cv}(R_d, R_{d'}))$

We know from (2.10) that

$$\mathrm{Ex}\Big(Z_d\,\widetilde{R}_{d'}\Big) = \eta\, b^{d'-d}\,, \qquad \text{for } d \leq d'\,,$$

and from (2.11a) that

$$\eta = \xi\,(1 - b^2)\,.$$

Therefore if $d \leq d'$ then (2.12) becomes

$$\mathrm{Ex}\Big(\widetilde{R}_d\,\widetilde{R}_{d'}\Big) = b\,\mathrm{Ex}\Big(\widetilde{R}_{d-1}\,\widetilde{R}_{d'}\Big) + \xi\,(1 - b^2)\,b^{d'-d}\,, \quad \text{for every } d' \in \{1,2,\cdots\} \text{ and } d \in \{1,2,\cdots,d'\}\,. \tag{2.15}$$

Stationary Autoregressive Models ($\mathrm{Cv}(R_d, R_{d'})$)

Setting $d = d'$ in (2.15) we can use (2.2c) to find that

$$\xi = b\,\mathrm{Ex}\Big(\widetilde{R}_{d-1}\,\widetilde{R}_{d'}\Big) + \xi\,(1-b^2)\,, \qquad \text{for every } d' \in \{1,2,\cdots\}\,.$$

Hence, if $b \neq 0$ we obtain

$$\mathrm{Ex}\Big(\widetilde{R}_{d'-1}\,\widetilde{R}_{d'}\Big) = \xi\,b\,, \qquad \text{for every } d' \in \{1,2,\cdots\}\,. \tag{2.16}$$

Then by using (2.15) we can make an countdown induction argument on d initialized by (2.16) to show that if $b \neq 0$ then

$$\begin{gathered}\mathrm{Cv}(R_d, R_{d'}) = \mathrm{Ex}\Big(\widetilde{R}_d\,\widetilde{R}_{d'}\Big) = \xi\,b^{d'-d}\,,\\ \text{for every } d \in \{0,1,\cdots\} \text{ with } d \leq d'\,.\end{gathered} \tag{2.17}$$

Combining (2.14) with (2.17) proves formula (2.3c). □

Stationary Autoregressive Models (Autoregression Time)

When $b \neq 0$ the *autoregression time* t_{ar} of the stationary autoregressive model (2.1) is defined by

$$\frac{1}{t_{\text{ar}}} = \log\left(\frac{1}{|b|}\right) . \tag{2.18}$$

Then by (2.3b) we have

$$|\text{Cv}(R_d, Z_{d'})| = \eta \, \exp\left(-\frac{d-d'}{t_{\text{ar}}}\right) ,$$
$$\text{for every } d, d' \in \{1, \cdots\} \text{ with } d' \leq d ,$$

and by (2.3c) we have

$$|\text{Cv}(R_d, R_{d'})| = \xi \, \exp\left(-\frac{|d-d'|}{t_{\text{ar}}}\right) , \qquad \text{for every } d, d' \in \{0, 1, \cdots\} ,$$

The stationary autoregressive model is close to an IID model if t_{ar} is small.

Fitting Stationary Autoregressive Models (Introduction)

Fitting Stationary Autoregressive 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 autoregressive model of the form (2.1). Specifically, this approach constructs estmators $\hat{a}$ and $\hat{b}$ such

$$\left(\hat{a}, \hat{b}\right) = \arg\min\left\{\sum_{d=1}^{D} w_d |r(d) - a - b\, r(d-1)|^2\right\}, \tag{3.19}$$

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

$$\begin{aligned} \hat{\eta} &= \min\left\{\sum_{d=1}^{D} w_d |r(d) - a - b\, r(d-1)|^2\right\} \\ &= \sum_{d=1}^{D} w_d |r(d) - \hat{a} - \hat{b}\, r(d-1)|^2 . \end{aligned} \tag{3.20}$$

Fitting Stationary Autoregressive Models (Statistics)

It is helpful to define the return sample means

$$m_0 = \sum_{d=1}^{D} w_d \, r(d) \,, \qquad m_1 = \sum_{d=1}^{D} w_d \, r(d-1) \,, \tag{3.21a}$$

the return sample variances

$$v_{00} = \sum_{d=1}^{D} w_d \big(r(d) - m_0 \big)^2 \,, \qquad v_{11} = \sum_{d=1}^{D} w_d \big(r(d-1) - m_1 \big)^2 \,, \tag{3.21b}$$

and the return sample autocovariance

$$v_{10} = \sum_{d=1}^{D} w_d \big(r(d-1) - m_1 \big) \big(r(d) - m_0 \big) \,. \tag{3.21c}$$

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

$$a = m_0 - b \, m_1 + \tilde{a} \,. \tag{3.22}$$

Fitting Stationary Autoregressive Models (Nugget)

Then our goal is to find (a, b) that minimizes the *nugget*

$$\sum_{d=1}^{D} w_d |z(d)|^2 , \quad \text{where} \quad z(d) = r(d) - a - b\, r(d-1) .$$

We see from (3.21) and (3.22) that

$$\begin{aligned} z(d) &= \big(r(d) - m_0\big) - b\,\big(r(d-1) - m_1\big) + \tilde{a} \\ &= \tilde{r}_0(d) - b\tilde{r}_1(d) + \tilde{a} , \end{aligned}$$

where we define

$$\tilde{r}_0(d) = r(d) - m_0 , \qquad \tilde{r}_1(d) = r(d-1) - m_1 . \tag{3.23}$$

Then

$$\begin{aligned} |z(d)|^2 &= |\tilde{r}_0(d)|^2 + b^2 |\tilde{r}_1(d)|^2 + \tilde{a}^2 \\ &\quad - 2b\, \tilde{r}_1(d)\, \tilde{r}_0(d) + 2\tilde{a}\, \tilde{r}_0(d) - 2\tilde{a}b\, \tilde{r}_1(d) . \end{aligned} \tag{3.24}$$

Fitting Stationary Autoregressive Models (Nugget)

It is evident from (3.21) and (3.23) that $\{\tilde{r}_0(d)\}_{d=1}^{D}$ and $\{\tilde{r}_1(d)\}_{d=1}^{D}$ satisfy

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

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

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

By using these facts we see from (3.24) that our goal now is to find $(\tilde{a}, b)$ that minimizes the nugget

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

Fitting Stationary Autoregressive Models (Minimizer)

Because $v_{11} > 0$, the nugget is clearly minimized when

$$\tilde{a} = 0 \,, \qquad b = \frac{v_{10}}{v_{11}} \,,$$

and that

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

Recalling (3.19), (3.20), and (3.22), this suggests using the estimators

$$\hat{a} = m_0 - \frac{v_{10}}{v_{11}} m_1 \,, \qquad \hat{b} = \frac{v_{10}}{v_{11}} \,, \qquad \hat{\eta} = v_{00} - \frac{v_{10}^2}{v_{11}} \,. \qquad (3.25)$$

However, there is a problem with these estimators. Namely, the formula for $\hat{b}$ can give values that lie outside of the interval $(-1, 1)$.

Fitting Stationary Autoregressive Models (New Estimators)

So rather than use the estimators (3.25) given by the least squares fit, we will use the estimators

$$\hat{a} = m_0 - \frac{v_{10}}{v_{11}}\, m_1\,, \qquad \hat{b} = \frac{v_{10}}{\sqrt{v_{00}\, v_{11}}}\,, \qquad \hat{\eta} = v_{00} - \frac{v_{10}^2}{v_{11}}\,. \tag{3.26}$$

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

$$V = \begin{pmatrix} v_{00} & v_{10} \\ v_{10} & v_{11} \end{pmatrix}\,. \tag{3.27}$$

This condition is always met in practice. If we set $\hat{\xi} = v_{00}$ then

$$\hat{\eta} = \hat{\xi}\left(1 - \hat{b}^2\right)\,, \tag{3.28}$$

which is an analog of the second relation in (2.3a).

Fitting Stationary Autoregressive Models ($\hat{t}_{\mathrm{ar}}$)

Remark. Given a return history $\{r(d)\}_{d=0}^{D}$ of any risky asset, we can use the autoregressive estimator $\hat{b}$ given by (3.26) to estimate a autoregrssion time for that asset when $\hat{b} \neq 0$. In that case, motivated by formula (2.18), we define $\hat{t}_{\mathrm{ar}}$ by

$$\frac{1}{\hat{t}_{\mathrm{ar}}} = \log\left(\frac{1}{|\hat{b}|}\right) . \tag{3.29}$$

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

Autocovariance Matrix (Introduction)

Consider the 2×2 *autocovariance matrix*

$$V = \begin{pmatrix} v_{00} & v_{10} \\ v_{10} & v_{11} \end{pmatrix} . \tag{4.30}$$

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

$$\mathrm{Ex}(V) = \xi\, W , \quad \text{where} \quad W = \begin{pmatrix} 1 - \bar{w} & -\bar{w}_1 \\ -\bar{w}_1 & 1 - \bar{w} \end{pmatrix} , \tag{4.31}$$

with

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

Autocovariance Matrix (W Matrix)

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

$$0 < \bar{w}_1 < \bar{w}\,, \qquad \bar{w} + \bar{w}_1 < 1\,, \tag{4.32}$$

which implies that the symmetric matrix W given by (4.31) is always *diagonally dominant* and thereby is always *positive definite*.

Example. 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 is the positive definite matrix

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

Autocovariance Matrix (Least Squares Fit)

The deviation of V given by (4.30) from the form (4.31) 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

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

It is easily seen that $\|A\|_{\mathrm{F}}^2$ is the sum of the squares of the entries of A.

We can estimate ξ in the form (4.31) to give the best least squares fit with respect to the Frobenius norm. In other words, we set

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

Autocovariance Matrix (ξ Estimator)

Because

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

we see that its minimizer yields the estimator

$$\hat{\xi} = \frac{\mathrm{tr}(W\, V)}{\mathrm{tr}\big(W^2\big)}\,. \tag{4.33}$$

When this estimator $\hat{\xi}$ is expressed in terms of the entries of the matrices V and W given by (4.30) and (4.31) we have

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

Autocovariance Matrix (ξ Estimator)

The fact that $\hat{\xi} > 0$ whenever $V \neq 0$ is can be seen directly from (4.33) and the following general fact, the proof of which is left as an exercise.

Fact 2. 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 $\mathrm{tr}(AB) > 0$. (Hint: Diagonalize B.)

Moreover, it is evident from (4.31) and (4.33) that

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

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

Assessing Independence (Introduction)

Assessing Independence. We will now present three ways to assess how much a given return history $\{r(d)\}_{d=1}^{D}$ that is consistent with the *identical distribution assumption* of an IID model is also consistent with the *independence assumption* of an IID model. More specifically, we will present:

- a graphical assessment,
- an autoregressive assessment,
- an autocovariance assessment.

The first is purely visual, but can be used to build understanding of the data. The other two are analytical. They will yield measures ω^{ar} and ω^{ac} of how consistent the given data is with the independence assumption. As before, these measures will take values in the interval $[0,1]$ with higher values indicating greater consistency with the independence assumption.

Assessing Independence (Graphical)

Graphical Assessment. 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. This means that there is no correlation between R_d and $R_{d'}$ when $d \neq d'$. Because of this, if we *scatter plot* the points $\{(R_d, R_{d+c})\}_{d=1}^{D-c}$ in the rr'-plane for any $c > 0$ then 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 mimicked by such a model then when the points $\{(r(d), r(d+c))\}_{d=1}^{D-c}$ are scatter plotted in the rr'-plane they should appear to be distributed in a way consistant with the probability density $q(r)q(r')$.

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

Assessing Independence (Autoregressive Statistics)

Autoregressive Assessment. 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 sample means

$$m_0 = \sum_{d=1}^{D} w_d \, r(d) \,, \qquad m_1 = \sum_{d=1}^{D} w_d \, r(d-1) \,,$$

the return sample variances

$$v_{00} = \sum_{d=1}^{D} w_d \big(r(d) - m_0 \big)^2 \,, \qquad v_{11} = \sum_{d=1}^{D} w_d \big(r(d-1) - m_1 \big)^2 \,,$$

and the return sample autocovariance

$$v_{10} = \sum_{d=1}^{D} w_d \big(r(d-1) - m_1 \big) \big(r(d) - m_0 \big) \,.$$

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

Assessing Independence (Autoregressive Calibration)

The estimators (3.26) for the autoregressive model of the return history $\{r(d)\}_{d=0}^{D}$ are then given by

$$\hat{a} = m_0 - \frac{v_{10}}{v_{11}}\, m_1 \,, \qquad \hat{b} = \frac{v_{10}}{\sqrt{v_{00}\, v_{11}}} \,, \qquad \hat{\eta} = v_{00} - \frac{v_{10}^2}{v_{11}} \,. \qquad (5.34)$$

Because v_{00} is the sample variance of $\{r(d)\}_{d=1}^{D}$, we can set

$$\hat{\xi} = v_{00} \,.$$

Then the estimators satisfy

$$\hat{\eta} = \left(1 - \hat{b}^2\right) \hat{\xi} \,.$$

Because $\hat{\eta}$ is the sample variance of nugget $\{z(d)\}_{d=1}^{D}$, we see that

- $\hat{b}^2$ is the fraction of $\hat{\xi}$ contributed by the autoregression term;
- $1 - \hat{b}^2$ is the fraction of $\hat{\xi}$ contributed by the nugget term.

Assessing Independence (Autoregressive Metric)

This suggests that a natural metric of how well the history $\{r(d)\}_{d=1}^{D}$ can be mimicked by an IID model is

$$\omega^{\mathrm{ar}} = \hat{b}^2 = \frac{v_{10}^2}{v_{00}\, v_{11}}\,. \tag{5.35}$$

The closer ω^{ar} is to 0, the better the IID model.

Assessing Independence (Autocovariance Metric)

Autocovariance Assessment. The size of the deviation of V given by (4.30) from the form (4.31) with $\xi = \hat{\xi}$ given by (4.33) is quantified by

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

Therfore we defined the metric

$$\omega^{\mathrm{ac}} = 1 - \frac{\mathrm{tr}(W\,V)^2}{\mathrm{tr}(V^2)\;\mathrm{tr}(W^2)}\,. \tag{5.36}$$

This is the square of the sine of the angle between V and W as determined by the Frobenius scalar product. The closer ω^{ac} is to 0, the better an IID model mimics the data.

Assessing Independence (Autocovariance Metric)

Remark. From (5.36) we can show by using (4.30) and (4.31) that

$$\omega^{\mathrm{ac}} = \delta^2 + \left(1 - \delta^2\right)\cos(\phi)^2 \,,$$

where

$$\delta^2 = \frac{(v_{00} - v_{11})^2}{(v_{00} - v_{11})^2 + (v_{00} + v_{11})^2 + 4v_{10}^2} \,,$$

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

This shows that ω^{ac} is near 0 if and only if both δ and $\cos(\phi)$ are small. The first condition holds if and only if v_{00} and v_{11} are relatively close. The second holds if and only if the vectors $(1 - \bar{w}, \bar{w}_1)$ and $(2v_{10}, v_{00} + v_{11})$ are nearly orthogonal.