Modeling Portfolios that Contain Risky Assets

Stochastic Models II: Portfolios with Risky Assets

C. David Levermore
University of Maryland, College Park

Math 420: *Mathematical Modeling*
March 31, 2014 version

Stochastic Models II: Portfolios with Risky Assets

1. IID Models for Markets
2. IID Models for Markowitz Portfolios
3. Inequalities
4. Uncertainty in the Return Rate Estimators

Stochastic Models II: Portfolios with Risky Assets

IID Models for Markets. We now consider a market with N risky assets. Let $\{s_i(d)\}_{d=0}^{D_h}$ be the share price history of asset i. The associated return rate and growth rate histories are $\{r_i(d)\}_{d=1}^{D_h}$ and $\{x_i(d)\}_{d=1}^{D_h}$ where

$$r_i(d) = D\left(\frac{s_i(d)}{s_i(d-1)} - 1\right), \qquad x_i(d) = D \log\left(\frac{s_i(d)}{s_i(d-1)}\right).$$

Because each $s_i(d)$ is positive, each $r_i(d)$ is in $(-D, \infty)$ and each $x_i(d)$ is in $(-\infty, \infty)$. Let $\mathbf{r}(d)$ and $\mathbf{x}(d)$ be the N-vectors

$$\mathbf{r}(d) = \begin{pmatrix} r_1(d) \\ \vdots \\ r_N(d) \end{pmatrix}, \qquad \mathbf{x}(d) = \begin{pmatrix} x_1(d) \\ \vdots \\ x_N(d) \end{pmatrix}.$$

The market return rate and growth rate histories can then be expressed simply as $\{\mathbf{r}(d)\}_{d=1}^{D_h}$ and $\{\mathbf{x}(d)\}_{d=1}^{D_h}$ respectively.

An IID model for this market draws D_h random vectors $\{\mathbf{R}(d)\}_{d=1}^{D_h}$ from a fixed probablity density $q(\mathbf{R})$ over $(-D,\infty)^N$. Such a model is reasonable when the points $\{(d,\mathbf{r}(d))\}_{d=1}^{D_h}$ are distributed uniformly in d. This is hard to visualize when N is not small. You might think a necessary condition for the entire market to have an IID model is that each asset has an IID model. *This can be visualized for each asset by plotting the points $\{(d,r_i(d))\}_{d=1}^{D_h}$ in the dr-plane and seeing if they appear to be distributed uniformly in d.* Similar visual tests based on pairs of assets can be carried out by plotting the points $\{(d,r_i(d),r_j(d))\}_{d=1}^{D_h}$ in $\mathbb{R}^3$ with an interactive 3D graphics package.

Visual tests like those described above often show that funds behave more like IID models than individual stocks or bonds. This means that portfolio balancing strategies based on IID models might work better for portfolios composed largely of funds. This is one reason why some investors prefer investing in funds over investing in individual stocks and bonds.

A better lesson to be drawn from the observation in the last paragraph is that every sufficiently diverse portfolio of assets in a market will behave more like an IID model than many of the individual assets in that market. In other words, IID models for a market can be used to develop portfolio balancing strategies when the portfolios considered are sufficiently diverse, even when the behavior of individual assets in that market may not be well described by the model. This is another reason to prefer holding diverse, broad-based portfolios. *More importantly, this suggests that it is better to apply visual tests like those described above to representative portfolios rather than to individual assets in the market.*

Remark. Such visual tests can only warn you when IID models might not be appropriate for describing the data. There are also statistical tests that can play this role. *There is no visual or statistical test that can insure the validity of using an IID model for a market. However, due to their simplicity, IID models are often used unless there is a good reason not to use them.*

After you have decided to use an IID model for the market, you must gather statistical information about the return rate probability density $q(\mathbf{R})$. The mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Xi}$ of $\mathbf{R}$ are given by

$$\boldsymbol{\mu} = \int \mathbf{R}\, q(\mathbf{R})\, \mathrm{d}\mathbf{R}\,, \qquad \boldsymbol{\Xi} = \int (\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathsf{T}} q(\mathbf{R})\, \mathrm{d}\mathbf{R}\,.$$

Given any sample $\{\mathbf{R}(d)\}_{d=1}^{D_h}$ drawn from $q(\mathbf{R})$, these have the unbiased estimators

$$\hat{\boldsymbol{\mu}} = \sum_{d=1}^{D_h} w(d)\, \mathbf{R}(d)\,, \qquad \hat{\boldsymbol{\Xi}} = \sum_{d=1}^{D_h} \frac{w(d)}{1-\bar{w}} \left(\mathbf{R}(d)-\hat{\boldsymbol{\mu}}\right) \left(\mathbf{R}(d)-\hat{\boldsymbol{\mu}}\right)^{\mathsf{T}}\,.$$

If we assume that such a sample is given by the return rate data $\{\mathbf{r}(d)\}_{d=1}^{D_h}$ then these estimators are given in terms of the vector $\mathbf{m}$ and matrix $\mathbf{V}$ by

$$\hat{\boldsymbol{\mu}} = \mathbf{m}\,, \qquad \hat{\boldsymbol{\Xi}} = D\,\mathbf{V}\,.$$

IID Models for Markowitz Portfolios. Recall that the value of a portfolio that holds a risk-free balance $b_{\mathrm{rf}}(d)$ with return rate μ_{rf} and $n_i(d)$ shares of asset i during trading day d is

$$\Pi(d) = b_{\mathrm{rf}}(d)\left(1 + \tfrac{1}{D}\,\mu_{\mathrm{rf}}\right) + \sum_{i=1}^{N} n_i(d) s_i(d)\,.$$

We will assume that $\Pi(d) > 0$ for every d. Then the return rate $r(d)$ and growth rate $x(d)$ for this portfolio on trading day d are given by

$$r(d) = D\left(\frac{\Pi(d)}{\Pi(d-1)} - 1\right)\,, \qquad x(d) = D\,\log\!\left(\frac{\Pi(d)}{\Pi(d-1)}\right)\,.$$

Recall that the return rate $r(d)$ for the Markowitz portfolio associated with the distribution $\mathbf{f}$ can be expressed in terms of the vector $\mathbf{r}(d)$ as

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

This implies that if the underlying market has an IID model with return rate probability density $q(\mathbf{R})$ then the Markowitz portfolio with distribution $\mathbf{f}$ has the IID model with return rate probability density $q_{\mathbf{f}}(R)$ given by

$$q_{\mathbf{f}}(R) = \int \delta\Big(R - (1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} - \mathbf{R}^{\mathsf{T}}\mathbf{f}\Big)\, q(\mathbf{R})\, \mathsf{d}\mathbf{R}\,.$$

Here $\delta(\,\cdot\,)$ denotes the *Dirac delta distribution*, which can be defined by the property that for every sufficiently nice function $\psi(R)$

$$\int \psi(R)\, \delta\Big(R - (1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} - \mathbf{R}^{\mathsf{T}}\mathbf{f}\Big)\, \mathsf{d}R = \psi\Big((1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} + \mathbf{R}^{\mathsf{T}}\mathbf{f}\Big)\,.$$

Hence, for every sufficiently nice function $\psi(R)$ we have the formula

$$\begin{aligned}
\mathsf{Ex}\Big(\psi(R)\Big) &= \int \psi(R)\, q_{\mathbf{f}}(R)\, \mathsf{d}R \\
&= \iint \psi(R)\, \delta\Big(R - (1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} - \mathbf{R}^{\mathsf{T}}\mathbf{f}\Big)\, q(\mathbf{R})\, \mathsf{d}\mathbf{R}\, \mathsf{d}R \\
&= \int \psi\Big((1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} + \mathbf{R}^{\mathsf{T}}\mathbf{f}\Big)\, q(\mathbf{R})\, \mathsf{d}\mathbf{R}\,.
\end{aligned}$$

We can thereby compute the mean μ and variance ξ of $q_{\mathbf{f}}(R)$ to be

$$\begin{aligned}
\mu = \mathsf{Ex}(R) &= \int \Big((1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} + \mathbf{R}^{\mathsf{T}}\mathbf{f}\Big)\, q(\mathbf{R})\, \mathsf{d}\mathbf{R} \\
&= (1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} \int q(\mathbf{R})\, \mathsf{d}\mathbf{R} + \left(\int \mathbf{R}\, q(\mathbf{R})\, \mathsf{d}\mathbf{R}\right)^{\mathsf{T}} \mathbf{f} \\
&= (1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} + \boldsymbol{\mu}^{\mathsf{T}}\mathbf{f}\,, \\
\xi = \mathsf{Ex}\Big((R-\mu)^2\Big) &= \int \Big((1 - \mathbf{1}^{\mathsf{T}}\mathbf{f})\mu_{\mathsf{rf}} + \mathbf{R}^{\mathsf{T}}\mathbf{f} - \mu\Big)^2 q(\mathbf{R})\, \mathsf{d}\mathbf{R} \\
&= \int \Big(\mathbf{R}^{\mathsf{T}}\mathbf{f} - \boldsymbol{\mu}^{\mathsf{T}}\mathbf{f}\Big)^2 q(\mathbf{R})\, \mathsf{d}\mathbf{R} = \int \mathbf{f}^{\mathsf{T}}(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{f}\, q(\mathbf{R})\, \mathsf{d}\mathbf{R} \\
&= \mathbf{f}^{\mathsf{T}}\left(\int (\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathsf{T}} q(\mathbf{R})\, \mathsf{d}\mathbf{R}\right)\mathbf{f} = \mathbf{f}^{\mathsf{T}}\boldsymbol{\Xi}\,\mathbf{f}\,,
\end{aligned}$$

where we have used the facts that

$$\int q(\mathbf{R})\, \mathsf{d}\mathbf{R} = 1\,, \qquad \int \mathbf{R}\, q(\mathbf{R})\, \mathsf{d}\mathbf{R} = \boldsymbol{\mu}\,,$$

$$\int (\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathsf{T}} q(\mathbf{R})\, \mathsf{d}\mathbf{R} = \boldsymbol{\Xi}\,.$$

If we assume that the return rate history $\{\mathbf{r}(d)\}_{d=1}^{D_h}$ is an IID sample drawn from a probability density $q(\mathbf{R})$ then unbiased estimators of the associated mean $\boldsymbol{\mu}$ and variance Ξ are given in terms of $\mathbf{m}$ and $\mathbf{V}$ by

$$\widehat{\boldsymbol{\mu}} = \mathbf{m}\,, \qquad \widehat{\Xi} = D\,\mathbf{V}\,.$$

Then the Markowitz portfolio with distribution $\mathbf{f}$ has the return rate history $\{r(d)\}_{d=1}^{D_h}$ with

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

Moreover, this history is an IID sample drawn from a probability density $q_{\mathbf{f}}(R)$ and the formulas on the previous page show that the mean μ and variance ξ of $q_{\mathbf{f}}(R)$ have the unbiased estimators

$$\widehat{\mu} = \mu_{\mathsf{rf}}(1 - \mathbf{1}^{\mathsf{T}}\mathbf{f}) + \mathbf{m}^{\mathsf{T}}\mathbf{f}\,, \qquad \widehat{\xi} = D\mathbf{f}^{\mathsf{T}}\mathbf{V}\mathbf{f}\,.$$

Inequalities. In the next section we will use the *Cauchy inequality*, also called the *Cauchy-Schwarz inequality*. Often this inequality is first seen when studying vectors, where for any vectors $\mathbf{u}$ and $\mathbf{v}$ it takes the form

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \, ,$$

where $\| \cdot \|$ denotes the Euclidean norm (length). This inequality will be an equality if and only if $\mathbf{u}$ and $\mathbf{v}$ are proportional. Alternatively, when the vectors are D_h-dimesional it can be written in the form

$$\left(\sum_{d=1}^{D_h} u(d) \, v(d) \right)^2 \leq \left(\sum_{d=1}^{D_h} u(d)^2 \right) \left(\sum_{d=1}^{D_h} v(d)^2 \right) \, .$$

If we take $u(d) = 1$ for every d then the Cauchy inequalty yields

$$\left(\sum_{d=1}^{D_h} v(d) \right)^2 \leq \left(\sum_{d=1}^{D_h} 1^2 \right) \left(\sum_{d=1}^{D_h} v(d)^2 \right) = D_h \sum_{d=1}^{D_h} v(d)^2 \, .$$

We will also use an integral form of the Cauchy inequality, commonly called the *Cauchy inequality*, *Schwarz inequality*, or *Cauchy-Schwarz inequality*. For any real-valued functions $f(z)$ and $g(z)$ this inequality takes the form

$$\left(\int_a^b f(z)\, g(z)\, \mathrm{d}z \right)^2 \leq \left(\int_a^b f(z)^2\, \mathrm{d}z \right) \left(\int_a^b g(z)^2\, \mathrm{d}z \right) ,$$

where $\int_a^b \cdot\, \mathrm{d}z$ denotes the definite integral over a nonempty interval (a, b). The Cauchy inequality will be an equality if and only if $f(z)$ and $g(z)$ are proportional over (a, b). For example, if $p(z)$ is a positive probability density over (a, b) and we take

$$f(z) = z^2 \sqrt{p(z)} \quad \text{and} \quad g(z) = \sqrt{p(z)}\, ,$$

then because these functions are not proportional over any interval (a, b), the Cauchy inequality yields the strict inequality

$$\left(\int_a^b z^2 p(z)\, \mathrm{d}z \right)^2 < \left(\int_a^b z^4 p(z)\, \mathrm{d}z \right) \left(\int_a^b p(z)\, \mathrm{d}z \right) = \int_a^b z^4 p(z)\, \mathrm{d}z\, .$$

In the next section we will also use the *Jensen inequality*. This states that if the function $f(z)$ is convex (concave) over an interval $[a, b]$, the points $\{z(d)\}_{d=1}^{D_h}$ all lie within $[a, b]$, and the nonnegative weights $\{w(d)\}_{d=1}^{D_h}$ sum to one, then

$$f(\bar{z}) \leq \overline{f(z)} \qquad \left(\overline{f(z)} \leq f(\bar{z})\right),$$

where

$$\bar{z} = \sum_{d=1}^{D_h} z(d)\, w(d)\,, \qquad \overline{f(z)} = \sum_{d=1}^{D_h} f(z(d))\, w(d)\,.$$

For example, if we take $f(z) = z^p$ for some $p > 1$, so that $f(z)$ is convex over $[0, \infty)$, and we take $z(d) = w(d)$ for every d then because the points $\{w(d)\}_{d=1}^{D_h}$ all lie within $[0, 1]$, the Jensen inequality yields

$$\bar{w}^p = \left(\sum_{d=1}^{D_h} w(d)^2\right)^p \leq \sum_{d=1}^{D_h} w(d)^{p+1} = \overline{w^p}\,.$$

The Jensen inequality can be proved for the case when $f(z)$ is convex and differentiable over $[a,b]$ by starting from the inequality

$$f(z) \geq f(\bar{z}) + f'(\bar{z})(z - \bar{z}) \qquad \text{for every } z \in [a,b]\,.$$

This inequality simply says that the tangent line to the graph of f at $\bar{z}$ lies below the graph of f over $[a,b]$. By setting $z = z(d)$ in the above inequality, multiplying both sides by $w(d)$, and summing over d we obtain

$$\begin{aligned}\sum_{d=1}^{D_h} f(z(d))\, w(d) &\geq \sum_{d=1}^{D_h} \Big(f(\bar{z}) + f'(\bar{z})(z(d) - \bar{z})\Big)\, w(d) \\ &= f(\bar{z}) \sum_{d=1}^{D_h} w(d) + f'(\bar{z}) \left(\sum_{d=1}^{D_h} \Big(z(d) - \bar{z}\Big)\, w(d) \right) .\end{aligned}$$

The Jensen inequality then follows from the definitions of $\bar{z}$ and $\overline{f(z)}$.

Remark. There is an integral version of the Jensen inequality that we do not give here because we do not need it.

Uncertainity in the Return Rate Estimators. Suppose that $\{R(d)\}_{d=1}^{D_h}$ is an IID sample drawn from a probability density $q_{\mathbf{f}}(R)$ with mean μ and variance ξ. Let $\widehat{\mu}$ and $\widehat{\xi}$ be the unbiased estimators

$$\widehat{\mu} = \sum_{d=1}^{D_h} w(d)R(d) \,, \qquad \widehat{\xi} = \frac{1}{1-\bar{w}} \sum_{d=1}^{D_h} w(d)\Big(R(d) - \widehat{\mu}\Big)^2 .$$

We would like to quantify how close these estimators are to μ and ξ for the Markowitz portfolio given by $\mathbf{f}$.

In the previous lecture we estimated how close $\widehat{\mu}$ is to μ by computing its variance. We found that

$$\mathrm{Var}(\widehat{\mu}) = \mathrm{Ex}\Big((\widehat{\mu}-\mu)^2\Big) = \bar{w}\,\xi \,, \qquad \text{where} \quad \bar{w} = \sum_{d=1}^{D_h} w(d)^2 .$$

This shows that $\widehat{\mu}$ converges to μ like $\sqrt{\bar{w}}$ as $\bar{w} \to 0$.

Remark. The Cauchy inequality implies that

$$1 = \left(\sum_{d=1}^{D_h} 1 \cdot w(d) \right)^2 \leq \left(\sum_{d=1}^{D_h} 1^2 \right) \left(\sum_{d=1}^{D_h} w(d)^2 \right) = D_h \, \bar{w} \, .$$

This shows that for any weighting we have $\bar{w} \geq 1/D_h$. Therefore the variance is smallest for uniform weights when we have $w(d) = 1/D_h$.

Remark. For uniform weights the formula for $\mathsf{Var}(\widehat{\mu})$ reduces to

$$\mathsf{Var}(\widehat{\mu}) = \frac{1}{D_h} \, \xi \, .$$

Therefore $\widehat{\mu}$ converges to μ like $1/\sqrt{D_h}$ as $D_h \to \infty$ for uniform weights.

The above considerations suggest that the uncertainties associated with the unbiased estimator $\widehat{\mu}$ can be measured by

$$\left(\bar{w} \, \widehat{\xi} \right)^{\frac{1}{2}} \, .$$

We can also estimate how close $\widehat{\xi}$ is to ξ by computing its variance. To do this we must assume that the probability density $q_{\mathbf{f}}(R)$ has a finite fourth moment. Let ξ_4 be the centered fourth moment, which is given by

$$\xi_4 = \mathrm{Ex}\Big((R-\mu)^4\Big) = \int_{-D}^{\infty} (R-\mu)^4 q_{\mathbf{f}}(R)\,\mathrm{d}R < \infty\,.$$

Observe that by the strict Cauchy inequality we have

$$\xi_4 = \int_{-D}^{\infty} (R-\mu)^4 q_{\mathbf{f}}(R)\,\mathrm{d}R > \left(\int_{-D}^{\infty} (R-\mu)^2 q_{\mathbf{f}}(R)\,\mathrm{d}R\right)^2 = \xi^2\,.$$

The first step is to let $\tilde{R}(d) = R(d) - \mu$ and express $\widehat{\xi}$ as

$$\begin{aligned}\widehat{\xi} &= \frac{1}{1-\bar{w}}\left(\sum_{d=1}^{D_h} w(d)\tilde{R}(d)^2 - (\widehat{\mu}-\mu)^2\right)\\ &= \frac{1}{1-\bar{w}}\left(\sum_{d=1}^{D_h} w(d)\tilde{R}(d)^2 - \sum_{d_1=1}^{D_h}\sum_{d_2=1}^{D_h} w(d_1)w(d_2)\tilde{R}(d_1)\tilde{R}(d_2)\right)\,.\end{aligned}$$

By squaring this expression and relabeling some indices we obtain

$$\begin{aligned}\hat{\xi}^2 = &\sum_{d=1}^{D_h}\sum_{d'=1}^{D_h} \frac{w(d)w(d')}{(1-\bar{w})^2}\,\tilde{R}(d)^2\tilde{R}(d')^2 \\ &- 2\sum_{d=1}^{D_h}\sum_{d_1=1}^{D_h}\sum_{d_2=1}^{D_h} \frac{w(d)w(d_1)w(d_2)}{(1-\bar{w})^2}\,\tilde{R}(d)^2\tilde{R}(d_1)\tilde{R}(d_2) \\ &+ \sum_{d_1=1}^{D_h}\sum_{d_2=1}^{D_h}\sum_{d_3=1}^{D_h}\sum_{d_4=1}^{D_h}\Bigg(\frac{w(d_1)w(d_2)w(d_3)w(d_4)}{(1-\bar{w})^2} \\ &\qquad \tilde{R}(d_1)\tilde{R}(d_2)\tilde{R}(d_3)\tilde{R}(d_4)\Bigg)\,.\end{aligned}$$

The next step is to compute $\mathsf{Ex}\Big(\hat{\xi}^2\Big)$, which requires us to compute

$$\mathsf{Ex}\Big(\tilde{R}(d)^2\tilde{R}(d')^2\Big)\ , \qquad \mathsf{Ex}\Big(\tilde{R}(d)^2\tilde{R}(d_1)\tilde{R}(d_2)\Big)\ ,$$
$$\mathsf{Ex}\Big(\tilde{R}(d_1)\tilde{R}(d_2)\tilde{R}(d_3)\tilde{R}(d_4)\Big)\ .$$

Let $\delta_{dd'}$ denote the Kronecker delta, which is defined by

$$\delta_{dd'} = \begin{cases} 1 & \text{if } d = d' \,, \\ 0 & \text{if } d \neq d' \,. \end{cases}$$

Because $\tilde{R}(d)$ and $\tilde{R}(d')$ are independent when $d \neq d'$, and because $\mathrm{Ex}\Big(\tilde{R}(d)\Big) = 0$, $\mathrm{Ex}\Big(\tilde{R}(d)^2\Big) = \xi$, and $\mathrm{Ex}\Big(\tilde{R}(d)^4\Big) = \xi_4$, we see that

$$\begin{aligned}
\mathrm{Ex}\Big(\tilde{R}(d)^2 \tilde{R}(d')^2\Big) &= \delta_{dd'}\,\xi_4 + \Big(1 - \delta_{dd'}\Big)\,\xi^2 \,, \\
\mathrm{Ex}\Big(\tilde{R}(d)^2 \tilde{R}(d_1)\tilde{R}(d_2)\Big) &= \delta_{d_1 d_2}\left(\delta_{dd_1}\,\xi_4 + \Big(1 - \delta_{dd_1}\Big)\,\xi^2\right) \,, \\
\mathrm{Ex}\Big(\tilde{R}(d_1)\tilde{R}(d_2)\tilde{R}(d_3)\tilde{R}(d_4)\Big) &= \delta_{d_1 d_2}\,\delta_{d_2 d_3}\,\delta_{d_3 d_4}\,\xi_4 \\
&\quad + \delta_{d_1 d_2}\,\delta_{d_3 d_4}\Big(1 - \delta_{d_1 d_3}\Big)\,\xi^2 \\
&\quad + \delta_{d_1 d_3}\,\delta_{d_4 d_2}\Big(1 - \delta_{d_1 d_4}\Big)\,\xi^2 \\
&\quad + \delta_{d_1 d_4}\,\delta_{d_2 d_3}\Big(1 - \delta_{d_1 d_2}\Big)\,\xi^2 \,.
\end{aligned}$$

Then the expected value of the quantity $\widehat{\xi}^2$ given two slides back is

$$\mathsf{Ex}\left(\widehat{\xi}^2\right) = \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\xi_4 + \frac{1 - 3\bar{w} + 2\overline{w^2} + 3\bar{w}^2 - 3\overline{w^3}}{(1-\bar{w})^2}\,\xi^2\,,$$

where $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ are given by

$$\bar{w} = \sum_{d=1}^{D_h} w(d)^2\,, \qquad \overline{w^2} = \sum_{d=1}^{D_h} w(d)^3\,, \qquad \overline{w^3} = \sum_{d=1}^{D_h} w(d)^4\,.$$

Therefore the variance of $\widehat{\xi}$ is

$$\begin{aligned}\mathsf{Var}\left(\widehat{\xi}\right) &= \mathsf{Ex}\left((\widehat{\xi} - \xi)^2\right) = \mathsf{Ex}\left(\widehat{\xi}^2\right) - \xi^2 \\ &= \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\xi_4 + \frac{-\bar{w} + 2\overline{w^2} + 2\bar{w}^2 - 3\overline{w^3}}{(1-\bar{w})^2}\,\xi^2 \\ &= \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\left(\xi_4 - \xi^2\right) + 2\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\xi^2\,.\end{aligned}$$

Remark. For uniform weights the formula for $\mathsf{Var}(\widehat{\xi})$ reduces to

$$\mathsf{Var}\left(\widehat{\xi}\right) = \frac{1}{D_h}\left(\xi_4 - \xi^2\right) + \frac{2}{D_h(D_h-1)}\,\xi^2\,.$$

Therefore $\widehat{\xi}$ converges to ξ like $1/\sqrt{D_h}$ as $D_h \to \infty$ for uniform weights.

The coefficient in front of $(\xi_4 - \xi^2)$ above is the smallest possible because, by the Cauchy inequality, the general coefficient of $(\xi_4 - \xi^2)$ satisfies

$$\begin{aligned}
\frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2} &= \frac{1}{(1-\bar{w})^2}\sum_{d=1}^{D_h}\Big(1-w(d)\Big)^2 w(d)^2 \\
&\geq \frac{1}{(1-\bar{w})^2}\,\frac{1}{D_h}\left(\sum_{d=1}^{D_h}\Big(1-w(d)\Big)w(d)\right)^2 \\
&= \frac{1}{(1-\bar{w})^2}\,\frac{1}{D_h}\,(1-\bar{w})^2 = \frac{1}{D_h}\,.
\end{aligned}$$

In order to treat cases when the weights are not uniform it is useful to derive an upper bound for $\mathrm{Var}\left(\widehat{\xi}\right)$ *in which the coefficients of* $(\xi_4 - \xi^2)$ *and* ξ^2 *depend on* $\bar{w}$ *but not on* $\overline{w^2}$ *and* $\overline{w^3}$*.* Because the Jensen inequality implies that $\bar{w}^3 \leq \overline{w^3}$, the coefficient of ξ^2 can be bounded as

$$\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2} \leq \frac{\bar{w}^2 - \bar{w}^3}{(1-\bar{w})^2} = \frac{\bar{w}^2}{1-\bar{w}}\,.$$

The coefficient of $(\xi_4 - \xi^2)$ requires more work. It can be checked that $f(z) = z - 2z^2 + z^3$ is concave over $[0, \frac{2}{3}]$. Hence, when the weights $\{w(d)\}_{d=1}^{D_h}$ all lie in $[0, \frac{2}{3}]$ the Jensen inequality with $z(d) = w(d)$ yields

$$\overline{w - 2w^2 + w^3} = \overline{f(w)} \leq f(\bar{w}) = \bar{w} - 2\bar{w}^2 + \bar{w}^3\,.$$

In that case the coefficient of $(\xi_4 - \xi^2)$ can be bounded as

$$\frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2} \leq \frac{\bar{w} - 2\bar{w}^2 + \bar{w}^3}{(1-\bar{w})^2} = \bar{w}\,.$$

Therefore if every $w(d) \le \frac{2}{3}$ then we obtain the upper bound

$$\mathsf{Var}\left(\widehat{\xi}\right) \le \bar{w}\left(\xi_4 - \xi^2\right) + \frac{2\bar{w}^2}{1-\bar{w}}\,\xi^2 \,.$$

This shows that $\widehat{\xi}$ converges to ξ like $\sqrt{\bar{w}}$ as $\bar{w} \to 0$ for arbitrary weights. Moreover, the above inequality is an equality when the weights are uniform.

The above considerations suggest that the uncertainties associated with the unbiased estimator $\widehat{\xi}$ can be measured by

$$\left(\bar{w}\left(\widehat{\xi}_4 - \widehat{\xi}^2\right) + \frac{2\bar{w}^2}{1-\bar{w}}\,\widehat{\xi}^2\right)^{\frac{1}{2}} \,,$$

where we choose to use the (biased) estimator of ξ_4 given by

$$\widehat{\xi}_4 = \frac{1}{(1-\bar{w})^2} \sum_{d=1}^{D_h} w(d)\left(R(d) - \widehat{\mu}\right)^4 .$$