

Multivariable Calculus and Lagrange Multipliers

Gradients, Hessians and multipliers: how calculus finds the best mix of two assets, and what that answer is worth out of sample.

VOLUME I · Calculus for Finance

BUILDS ON LM2 · LM3 · LM4

LEADS TO LM10 · LM32 · LM33 · LM34 · LM43

TIME ≈ 3 h + lab

LEARNING OUTCOMES

Mastery *After this module you should be able to:*

☐	1 compute partial derivatives and total differentials of functions of two or three variables, and interpret them as sensitivities with the other inputs held fixed
☐	2 construct the gradient and Hessian of a function of several variables, and approximate the function with its second-order multivariate Taylor expansion
☐	3 locate and classify the critical points of a function of two variables with the first-order conditions and the 2×2 definiteness test
☐	4 solve equality-constrained problems with Lagrange multipliers, and interpret the multiplier as a shadow price
☐	5 derive the two-asset minimum-variance weight and explain how correlation determines the diversification benefit
☐	6 derive the two-asset tangency weights and Kelly fractions, and relate them to $\Sigma^{-1}(\boldsymbol{\mu} - r_f \mathbf{1})$
☐	7 apply the Karush–Kuhn–Tucker conditions to long-only, policy and leverage constraints, and identify which constraints bind
☐	8 implement numerical gradients, Newton's method and constrained optimization in Python, and distinguish in-sample optimal weights from out-of-sample results

LEARNING MODULE OVERVIEW

- A partial derivative is a sensitivity with every other input frozen. A TWD investor's return on a US-dollar bond is $(1+x)(1+y) - 1$; in 2022 the cross term xy alone cost 00679B holders 3.2 percentage points.
- The gradient points uphill and the Hessian measures curvature. The second-order expansion $h(\mathbf{a} + \Delta) \approx h(\mathbf{a}) + \nabla h \cdot \Delta + \frac{1}{2} \Delta^\top H \Delta$ reproduces the rise in a fixed 0050/00679B mix's variance from 2017–2021 to 2022–2026 almost exactly; the first-order terms explain only 74% of it.
- A critical point of $h(x, y)$ is a local minimum when $h_{xx} > 0$ and $\det H = h_{xx}h_{yy} - h_{xy}^2 > 0$, a local maximum when $h_{xx} < 0$ and $\det H > 0$, and a saddle when $\det H < 0$.
- At a constrained optimum the gradients of objective and constraint are parallel, $\nabla h = \lambda \nabla \psi$. The multiplier λ is a shadow price: the rate at which the optimal value improves as the constraint is relaxed.
- The two-asset minimum-variance weight is $w^* = (\sigma_2^2 - \sigma_{12})/(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$. Estimated on 2017–2021 it put 42.3% in 0050 and 57.7% in 00679B, with a predicted volatility of 9.22%; in 2022–2026 the same mix realized 12.44%.
- Most of the miss came from changed inputs, not from the weight: 0050's volatility rose from 16.4% to 23.4% and the correlation from -0.22 to -0.02 . Because variance is flat at its minimum, a 17-point weight error cost only 0.9 points of volatility.
- The two-asset Kelly fractions $\mathbf{f}^* = \Sigma^{-1}(\boldsymbol{\mu} - r_f \mathbf{1})$ are the tangency portfolio, levered. In-sample estimates of the 00679B fraction swing from $+3.14$ to -4.06 between windows: optimal weights fitted to the past are not recommendations.
- Inequality constraints add the KKT conditions: a binding constraint carries a non-negative multiplier and a slack one a zero multiplier. Newton's method finds the sample's growth-optimal mix in three iterations; fixed-step gradient ascent needs thirty.

1

INTRODUCTION

Every portfolio decision turns more than one dial: how much 0050, how much bond, how much leverage. Single-variable calculus (LM2, LM4) turns one dial at a time; this module turns them together, adds constraints and prices the constraints themselves. Applied to two assets, it yields the minimum-variance portfolio, the tangency portfolio and the Kelly fractions, and it shows why an answer fitted to the past is not the answer you will get.

MOTIVATING CASE How much bond minimizes risk?

A TWD investor holds Yuanta Taiwan 50 (0050) and considers adding Yuanta U.S. Treasury 20+ Year Bond ETF (00679B). The fund tracks the ICE US Treasury 20+ Year Bond Index and is quoted in New Taiwan dollars. Its holdings report lists US Treasuries and US-dollar cash but no currency forwards, so its returns combine US long-bond returns with USD/TWD moves.

From the close of 17 January 2017, 00679B's first trading day, to 8 October 2026, 0050 returned +767.3% with dividends reinvested, at 20.1% annualized volatility. 00679B returned −17.3% at 13.5%. Their daily returns were mildly negatively correlated, −0.11 (Exhibit 1).

Suppose the investor had sized the mix at the end of 2021 with the five years then available. Over 2017–2021 the correlation was −0.22, and the calculus of Section 6 says that 42.3% in 0050 and 57.7% in 00679B minimized volatility, at a predicted 9.22% a year.

Then came 2022: 0050 lost 21.3% and 00679B lost 22.1%. Over 2022–2026 the correlation drifted to −0.02, and the same mix realized 12.44%. Was the calculus wrong, was the weight wrong, or did the inputs change? By the end of Section 6 you will split the 3.2-point miss into its parts.

Exhibit 1: 0050 and 00679B: estimation window, test window and full sample

	Full sample 2017-01 → 2026-10	Estimation window 2017–2021	Test window 2022- 01 → 2026-10
Trading days (daily returns)	2,363	1,213	1,150
0050 total return	+767.3%	+138.0%	+264.4%
00679B total return	−17.3%	+15.6%	−28.5%
0050 annualized volatility	20.1%	16.4%	23.4%
00679B annualized volatility	13.5%	13.5%	13.4%
Correlation of daily returns	−0.11	−0.22	−0.02
Minimum-variance weight in 0050, estimated in the window	32.7%	42.3%	25.3%
Lowest attainable volatility in the window	10.64%	9.22%	11.54%
Volatility of the 42.3% / 57.7% mix	10.91%	9.22%	12.44%

Source: FinMind `TaiwanStockPrice` and `TaiwanStockDividendResult`, retrieved 10 October 2026. Daily total returns between consecutive common trading dates from the base date 17 January 2017. Cash dividends are reinvested on every ex-date in the window (20 for 0050, 37 for 00679B), and 0050 is adjusted for its 4-for-1 split of 18 June 2025. Volatilities (divisor T) are annualized with $A = 243.0543$ trading sessions a year: the sample's 2,363 returns over 3,551/365.25 calendar years. Minimum-variance weights and volatilities use unrounded moments. Mixes are rebalanced daily. Computations: `data/lm07_stats.py` and companion notebook `LM07_lab.ipynb`.

Sections 2–5 extend derivatives, Taylor expansions and optimization to several variables and add equality constraints. Section 6 applies them to two assets and resolves the case, Section 7 adds inequality constraints and Section 8 solves everything in Python. Matrix notation for N assets waits for [LM10](#) and [LM32](#).

NOTE Notation for this module

A generic function h of generic coordinates x, y , instantiated per example, has partial derivatives $h_x = \partial h / \partial x$ and $h_{xy} = \partial^2 h / \partial y \partial x$, gradient ∇h and Hessian H . Constraints read $\psi(\mathbf{x}) = b$ or $\psi_i(\mathbf{x}) \leq b_i$, with multipliers λ (equality) and ν (inequality) and optimal value $J(b)$. Bold letters are vectors: weights $\mathbf{w}$, exposures per unit of capital $\mathbf{f}$ (Kelly fractions), expected returns $\boldsymbol{\mu}$, excess returns $\mathbf{e}$ ($e_i = \mu_i - r_f$), $\mathbf{1} = (1, 1)$, a point $\mathbf{a}$, a direction $\mathbf{u}$, a step Δ . For two assets, σ_1, σ_2 are annualized volatilities, σ_{12} the covariance, ρ the correlation, σ_d^2 the variance of $R_1 - R_2$; V is a portfolio variance, σ_b a volatility budget, g a log growth rate, m, M the smallest and largest curvatures of Σ . Counts: T returns, N assets, iteration k ; i, j index assets or constraints. Moments use divisor T , as in LM3; with 1,150 or more returns, $T - 1$ would change volatilities by under 0.05%.

2 FUNCTIONS OF SEVERAL VARIABLES AND PARTIAL DERIVATIVES

- ☐ compute partial derivatives and total differentials of functions of two or three variables, and interpret them as sensitivities with the other inputs held fixed

A portfolio's volatility depends on two weights, two volatilities and a correlation at once, and a TWD investor's return on a US bond depends on the bond and on the currency.

Functions, surfaces and level curves

DEFINITION Function of two variables

A **function of two variables** assigns to each point (x, y) of a domain $\mathcal{D} \subseteq \mathbb{R}^2$ one number $h(x, y)$. Its **graph** is the surface $z = h(x, y)$. A **level curve** (contour) is the set of points where h takes a fixed value b , $\{(x, y) : h(x, y) = b\}$.

Level curves are the map view of the surface. The central example is the variance of a two-asset portfolio. With weights w_1, w_2 and daily returns R_{1t}, R_{2t} , the portfolio return $R_{pt} = w_1 R_{1t} + w_2 R_{2t}$ deviates from its mean by $w_1(R_{1t} - \bar{R}_1) + w_2(R_{2t} - \bar{R}_2)$. Squaring and averaging gives

$$\begin{aligned} \frac{1}{T} \sum_t (R_{pt} - \bar{R}_p)^2 &= w_1^2 \frac{1}{T} \sum_t (R_{1t} - \bar{R}_1)^2 + 2w_1w_2 \frac{1}{T} \sum_t (R_{1t} - \bar{R}_1)(R_{2t} - \bar{R}_2) \\ &\quad + w_2^2 \frac{1}{T} \sum_t (R_{2t} - \bar{R}_2)^2. \end{aligned}$$

The middle average is the **covariance** of the two returns, positive when they tend to fall on the same side of their means (LM16). Divided by the two standard deviations it gives the **correlation** ρ . It lies between -1 and 1 by the Cauchy-Schwarz inequality $|\sum_t x_t y_t| \leq \sqrt{\sum_t x_t^2} \sqrt{\sum_t y_t^2}$, applied to the deviations (stated without proof; LM8). Annualizing each average with A ,

$$V(w_1, w_2) = w_1^2 \sigma_1^2 + 2w_1w_2 \sigma_{12} + w_2^2 \sigma_2^2, \quad \sigma_{12} = \rho \sigma_1 \sigma_2.$$

The same expansion gives two facts used later. Averaging $(R_{1t} - \bar{R}_1)(R_{pt} - \bar{R}_p)$ instead shows that the covariance of R_1 with the portfolio is $w_1 \sigma_1^2 + w_2 \sigma_{12}$, and weights $(1, -1)$ give $\text{Var}(R_1 - R_2) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$.

When $|\rho| < 1$, the level curves of V are ellipses centred on the all-cash point $(0, 0)$, one for each variance level. The budget rule $w_1 + w_2 = 1$ is a straight line across them. Section 5 shows that the minimum-variance portfolio sits where that line just touches an ellipse, and Exhibit 5 draws the picture.

Partial derivatives

DEFINITION Partial derivative

The **partial derivative** of h with respect to x at (x_0, y_0) is the ordinary derivative of the one-variable function $x \mapsto h(x, y_0)$:

$$h_x(x_0, y_0) = \frac{\partial h}{\partial x}(x_0, y_0) := \lim_{\delta \rightarrow 0} \frac{h(x_0 + \delta, y_0) - h(x_0, y_0)}{\delta}.$$

h_y is defined in the same way with x held at x_0 .

To compute a partial derivative, treat every other variable as a constant and apply the rules of LM2. For the portfolio variance,

$$\frac{\partial V}{\partial w_1} = 2(w_1 \sigma_1^2 + w_2 \sigma_{12}), \quad \frac{\partial V}{\partial w_2} = 2(w_1 \sigma_{12} + w_2 \sigma_2^2).$$

The first bracket is the covariance of asset 1 with the portfolio, found above. So the marginal variance of a position is twice its covariance with the portfolio it joins, a fact LM36 builds into risk budgeting.

Higher-order partial derivatives. Differentiating again gives h_{xx} , h_{yy} and the mixed partials $h_{xy} = (h_x)_y$ and $h_{yx} = (h_y)_x$. For every function in this book the order does not matter.

THEOREM Symmetry of mixed partial derivatives (Schwarz–Clairaut)

If h_{xy} and h_{yx} exist and are continuous near (x_0, y_0) , then $h_{xy}(x_0, y_0) = h_{yx}(x_0, y_0)$. Stated without proof; see Stewart, Clegg and Watson (2021, §14.3).

The total differential and the linear approximation

Move both inputs at once, by Δx and Δy . Going first along x and then along y , and using the one-variable linear approximation of LM2 on each leg, gives

$$h(x_0 + \Delta x, y_0 + \Delta y) \approx h(x_0, y_0) + h_x(x_0, y_0) \Delta x + h_y(x_0, y_0) \Delta y. \quad (7.1)$$

When h has continuous partial derivatives, the error is small relative to the step; when its second partial derivatives are continuous too, the error is of second order in $(\Delta x, \Delta y)$. The linear part, $dh = h_x dx + h_y dy$, is the **total differential**. Dividing by a time step gives the multivariable **chain rule**: if $x = x(t)$ and $y = y(t)$, then $\frac{d}{dt}h(x(t), y(t)) = h_x x'(t) + h_y y'(t)$.

EXAMPLE 1**A US bond in Taiwan dollars: 00679B in 2022**

A TWD investor's return on a US-dollar asset is $h(x, y) = (1 + x)(1 + y) - 1$, where x is the asset's return in USD and y the change in the USD/TWD rate. In 2022, 00679B returned -22.09% in TWD, while USD/TWD rose from 27.690 to 30.708, $+10.90\%$. Back out the USD return implied for the fund, compute the partial derivatives there, and split the TWD return into bond, currency and cross terms.

SOLUTION

The implied USD return solves $1 + x = (1 - 0.2209)/1.1090$: $x = -29.75\%$.

The partial derivatives are $h_x = 1 + y = 1.109$ and $h_y = 1 + x = 0.7025$, and $h_{xy} = h_{yx} = 1$. An extra 1% of bond return in USD adds 1.109% in TWD, because it is converted at a stronger dollar. An extra 1% rise in the dollar adds only 0.70%, because it applies to a bond position already down 30%.

Expanding the product gives the exact split $h = x + y + xy = -29.75\% + 10.90\% - 3.24\% = -22.09\%$. The linear approximation (7.1), $x + y = -18.85\%$, overstates the return by the cross term of 3.24 points: second order, but not small when the moves are 30% and 11%.

EXAMPLE 2**Marginal variance of a 60/40 mix**

With the full-sample estimates for 0050 and 00679B, $\sigma_1^2 = 0.040425$, $\sigma_2^2 = 0.018214$ and $\sigma_{12} = -0.002866$ (so $\sigma_1 = 20.11\%$, $\sigma_2 = 13.50\%$, $\rho = -0.106$), evaluate the variance, volatility and partial

derivatives of V at $\mathbf{w} = (0.6, 0.4)$. Then estimate the change in volatility if (a) one percentage point of 0050 is added and financed with cash, or (b) one point is moved from 00679B to 0050.

SOLUTION

$V = 0.36(0.040425) + 0.16(0.018214) + 0.48(-0.002866) = 0.016092$, so $\sigma_p = 12.69\%$. The partial derivatives are $V_{w_1} = 2(0.6 \cdot 0.040425 - 0.4 \cdot 0.002866) = 0.0462$ and $V_{w_2} = 2(-0.6 \cdot 0.002866 + 0.4 \cdot 0.018214) = 0.0111$.

Because $\sigma_p = \sqrt{V}$, the chain rule gives $\partial\sigma_p/\partial w_i = V_{w_i}/(2\sigma_p)$: 0.182 for 0050 and 0.044 for 00679B.

(a) $\Delta\mathbf{w} = (0.01, 0)$: $\Delta\sigma_p \approx 0.182 \times 0.01 = 0.18$ percentage points (exact: 0.18). (b) $\Delta\mathbf{w} = (0.01, -0.01)$: $\Delta\sigma_p \approx (0.182 - 0.044) \times 0.01 = 0.14$ points (exact: 0.14). The answers differ because the experiments differ: (a) raises total exposure, while (b) stays on the budget line $w_1 + w_2 = 1$.

PITFALL "Holding everything else fixed" is not always possible

In a fully invested portfolio the weights cannot move one at a time: buying 0050 means selling something. So $\partial V/\partial w_1$ describes a trade financed with cash or margin, case (a) above; a budget-neutral switch needs the derivative along $(1, -1)$, case (b), or substitution of the constraint (Sections 5 and 6). Likewise, " $\partial\sigma_p/\partial\rho$ holding volatilities fixed" ignores that in a crisis volatilities and correlations tend to move together.

導讀 偏導數：凍結其他變數的敏感度

偏導數就是「其他條件不變」時的敏感度：只動一個變數，其餘全部凍結。台幣投資美債的報酬是 $(1+x)(1+y)-1$ ，它對匯率的偏導數是 $1+x$ ：2022 年債券本身已跌三成，美元再升 1%，台幣報酬只多 0.70%。全微分把各方向的敏感度加總，是一階近似；變動很大時（債跌三成、美元升一成），交叉項 xy 就有 3.24 個百分點，不能忽略。在投資組合裡「其他條件不變」常常做不到：權重加總必須是 1，加碼 0050 就得減碼另一檔，要看的是沿著 $(1, -1)$ 方向的變化。

KNOWLEDGE CHECK 1

For $h(x, y) = (1 + x)(1 + y) - 1$ at $x = -10\%$, $y = +5\%$, compute h_x and h_y . Use (7.1) to estimate the change in h if y rises a further 2 percentage points, to 7%, and explain why the estimate is exact here.

ANSWER

$h_x = 1 + y = 1.05$ and $h_y = 1 + x = 0.90$. With $\Delta x = 0$ and $\Delta y = 0.02$, (7.1) gives $\Delta h \approx 0.90 \times 0.02 = 1.80$ points. It is exact because h is linear in y when x is fixed: the only second-order term, $\Delta x \Delta y$, vanishes when $\Delta x = 0$.

3

GRADIENT, HESSIAN AND THE MULTIVARIATE TAYLOR EXPANSION

- ☐ construct the gradient and Hessian of a function of several variables, and approximate the function with its second-order multivariate Taylor expansion

Directional derivatives and the gradient

Pick a direction, given by a unit vector $\mathbf{u} = (u_1, u_2)$, and walk from $\mathbf{a}$ along it; the height of the surface along the walk is $z(t) := h(\mathbf{a} + t\mathbf{u})$. By the chain rule, its slope at the start, the **directional derivative**, is

$$\partial_{\mathbf{u}} h(\mathbf{a}) := z'(0) = h_x(\mathbf{a}) u_1 + h_y(\mathbf{a}) u_2.$$

DEFINITION Gradient

The **gradient** of h at $\mathbf{a}$ is the vector of its partial derivatives, $\nabla h(\mathbf{a}) := (h_x(\mathbf{a}), h_y(\mathbf{a}))$. Every directional derivative is a dot product with it: $\partial_{\mathbf{u}} h = \nabla h \cdot \mathbf{u}$.

Two geometric facts follow. First, by the Cauchy-Schwarz inequality $\nabla h \cdot \mathbf{u} \leq \|\nabla h\| \|\mathbf{u}\| = \|\nabla h\|$, with equality only when $\mathbf{u}$ points along ∇h . So **the gradient points in the direction of steepest ascent, and its length is the steepest slope**.

Second, if a point $\mathbf{x}(t)$ moves along a level curve, then $h(\mathbf{x}(t)) = b$, and differentiating gives $\nabla h \cdot \mathbf{x}'(t) = 0$. So **the gradient is perpendicular to the level curves**.

For a step Δ that is not a unit vector, $\nabla h \cdot \Delta$ is the first-order change in h ; case (b) of Example 2 used exactly this, with $\Delta = (0.01, -0.01)$.

The Hessian

DEFINITION Hessian

The **Hessian** of h at $\mathbf{a}$ is the matrix of its second partial derivatives,

$$H(\mathbf{a}) := \begin{pmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{pmatrix}(\mathbf{a}),$$

symmetric when the Schwarz–Clairaut theorem applies.

The Hessian measures curvature in every direction at once. Differentiating $z(t) = h(\mathbf{a} + t\mathbf{u})$ twice with the chain rule gives

$$z''(0) = h_{xx}u_1^2 + 2h_{xy}u_1u_2 + h_{yy}u_2^2 =: \mathbf{u}^\top H \mathbf{u},$$

the curvature of h along $\mathbf{u}$. For the portfolio variance the Hessian is constant, $H = 2 \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = 2\Sigma$: twice the **covariance matrix** Σ , the object [LM10](#) studies.

The second-order Taylor expansion

The one-variable Taylor polynomial of LM3, applied to $z(t) = h(\mathbf{a} + t\Delta)$ at $t = 1$, gives $z(1) \approx z(0) + z'(0) + \frac{1}{2}z''(0)$. Substituting the two derivatives just computed:

$$h(\mathbf{a} + \Delta) \approx h(\mathbf{a}) + h_x\Delta_1 + h_y\Delta_2 + \frac{1}{2}(h_{xx}\Delta_1^2 + 2h_{xy}\Delta_1\Delta_2 + h_{yy}\Delta_2^2) = h(\mathbf{a}) + \nabla h \cdot \Delta + \frac{1}{2}\Delta^\top H \Delta, \quad (7.2)$$

with all derivatives evaluated at $\mathbf{a}$. If h has continuous third partial derivatives, the Lagrange remainder of LM3 (3.5), applied to z , puts the error at $z'''(\xi)/6$ for some $\xi \in (0, 1)$: third order in Δ . With three variables the pattern is the same, with three first derivatives and a 3×3 Hessian. A polynomial of degree two is reproduced exactly by (7.2), and one of degree three once the third-order terms are added.

A Taylor attribution of the 42/58 mix's extra risk

Hold the weights of the case fixed at $w_1 = 42.317\%$ in 0050 and $w_2 = 57.683\%$ in 00679B. Their variance is then a function of three inputs,

$$V(\sigma_1, \sigma_2, \rho) = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2.$$

Between the estimation and test windows the inputs moved from (16.37%, 13.55%, -0.221) to (23.40%, 13.43%, -0.022). The partial derivatives at the starting point, from the unrounded estimates, are $V_{\sigma_1} = 2w_1^2\sigma_1 + 2w_1w_2\rho\sigma_2 = 0.043994$, $V_{\sigma_2} = 2w_2^2\sigma_2 + 2w_1w_2\rho\sigma_1 = 0.072465$ and $V_\rho = 2w_1w_2\sigma_1\sigma_2 = 0.010827$. Multiplying each by its input's change gives the first-order terms of Exhibit 2.

The first-order approximation explains 51.70 of the 69.67 %² rise in variance, or 74%. The Hessian terms supply the rest, slightly overshooting. The two largest are $\frac{1}{2}V_{\sigma_1\sigma_1}(\Delta\sigma_1)^2 = 8.86$, with $V_{\sigma_1\sigma_1} = 2w_1^2$, and the cross term $V_{\sigma_1\rho}\Delta\sigma_1\Delta\rho = 9.28$, with $V_{\sigma_1\rho} = 2w_1w_2\sigma_2$. A higher correlation hurts more when 0050 is more volatile, and only a second-order model can see that. Because V is a polynomial of degree three in its inputs, the single third-order term $2w_1w_2\Delta\sigma_1\Delta\sigma_2\Delta\rho$, about -0.08 , makes the expansion exact.

Exhibit 2: Taylor attribution of the variance of the 42.3% / 57.7% mix, 2017–2021 → 2022–2026

Input or term	Change	Partial derivative	ΔV (% ²)
σ_1 (0050): 16.37% → 23.40%	+0.070325	0.043994	+30.94
σ_2 (00679B): 13.55% → 13.43%	−0.001140	0.072465	−0.83
ρ : −0.221 → −0.022	+0.199414	0.010827	+21.59
First-order approximation $\nabla V \cdot \Delta$			+51.70
Second-order approximation, adding $\frac{1}{2}\Delta^\top H \Delta$			+69.74
Exact change (degree-three polynomial)			+69.67

Note: variance in squared percentage points (%²); it rose from 85.1 to 154.8, so volatility rose from 9.22% to 12.44%. Changes and derivatives in decimal units. Computed by `data/lm07_stats.py`.

To first order, 0050's higher volatility accounts for 44% of the rise in variance and the higher correlation for 31%. The second-order terms add 26%, half of it from the two moving together.

A two-variable Taylor model of portfolio growth

The most important use of (7.2) in this book is the growth rate of a levered portfolio. Hold a fraction f_1 of capital in 0050 and f_2 in 00679B, and finance the rest at an illustrative risk-free rate $r_f = 1.5\%$ a year, r_f/A a day. The daily return is $R_p = r_f/A + f_1X_1 + f_2X_2$, with excess returns $X_i = R_i - r_f/A$. The quantity a Kelly investor maximizes (Kelly 1956; LM4) is the average log growth, annualized:

$$g(f_1, f_2) := A \cdot \frac{1}{T} \sum_{t=1}^T \ln(1 + r_f/A + f_1X_{1t} + f_2X_{2t}).$$

Differentiate under the sum: $g_{f_i} = A \cdot \frac{1}{T} \sum_t X_{it}/(1 + R_{pt})$ and $g_{f_i f_j} = -A \cdot \frac{1}{T} \sum_t X_{it}X_{jt}/(1 + R_{pt})^2$. At $\mathbf{f} = \mathbf{0}$ these are the mean excess returns and minus the mean products, each divided by a power of $1 + r_f/A$.

On the data, $\nabla g(\mathbf{0}) = (0.2275, -0.0255)$, which is (e_1, e_2) up to the factor $1/(1 + r_f/A)$. The Hessian $H(\mathbf{0})$ has entries $(-0.0406, 0.0029, -0.0182)$, against $-\sigma_1^2 = -0.0404$, $-\sigma_{12} = 0.0029$ and $-\sigma_2^2 = -0.0182$. The small gap in the first entry is the squared daily mean, smaller than the variance by a factor of order $1/A$. Dropping it and the factors $1 + r_f/A$, (7.2) becomes

$$g(f_1, f_2) \approx r_f + f_1e_1 + f_2e_2 - \frac{1}{2}(f_1^2\sigma_1^2 + 2f_1f_2\sigma_{12} + f_2^2\sigma_2^2). \quad (7.3)$$

The same model follows from expanding $\ln(1 + R_p) \approx R_p - \frac{1}{2}R_p^2$ day by day and averaging, as in LM3 (3.8). It neglects the squared daily mean, $\frac{1}{2}A\bar{R}_p^2$, and the terms of third and higher order in the daily portfolio return. In matrix form (7.3) reads $g \approx r_f + \mathbf{f}^\top \mathbf{e} - \frac{1}{2}\mathbf{f}^\top \Sigma \mathbf{f}$, the starting point of [LM43](#).

Exhibit 3: Exact sample log growth versus the quadratic model (7.3), 0050 and 00679B, 2017–2026

Portfolio	f_1 (0050)	f_2 (00679B)	Exact g	Model (7.3)	Model – exact
0050 alone	1.00	0.00	22.2200%	22.2297%	+1.0 bp
00679B alone	0.00	1.00	–1.9569%	–1.9573%	0.0 bp
Full-sample minimum-variance mix	0.33	0.67	6.6705%	6.6722%	+0.2 bp
Twice 0050, financed at r_f	2.00	0.00	38.8741%	38.9168%	+4.3 bp
Half Kelly	2.80	–0.26	49.5495%	49.6964%	+14.7 bp
Full Kelly of the model (Section 6)	5.59	–0.52	63.7424%	65.7619%	+202.0 bp

Note: annual log growth; $r_f = 1.5\%$ (illustrative); daily rebalancing, no costs. "Exact" is the sample average of $\ln(1 + R_p)$ times A . Growth is evaluated at unrounded exposures.

The model is excellent for unlevered mixes and degrades as leverage grows. For 0050 alone the 1.0 bp error is the squared mean's 1.2 bp less 0.2 bp of higher-order terms. At full Kelly the model overstates growth by 2.02 points a year: 0.35 from the dropped squared daily mean and 1.67 from the third- and higher-order terms of $\ln(1 + R_p)$, which grow with the daily moves. The fourth-order term alone costs 1.87 points; the third adds back 0.49, and fifth and higher orders cost 0.29.

導讀 梯度、海森矩陣與多變數泰勒展開

梯度 ∇h 指向函數上升最快的方向，長度就是最陡的坡度，而且永遠和等高線垂直。海森矩陣 H 描述各個方向的彎曲程度。多變數泰勒展開把 LM3 的「高度+斜率+曲率」推廣到多個變數： $h(\mathbf{a} + \Delta) \approx h(\mathbf{a}) + \nabla h \cdot \Delta + \frac{1}{2}\Delta^\top H \Delta$ 。拆解 42/58 組合的變異數為何上升：一階項只解釋 74%，加上二階項（尤其是 0050 波動與相關係數「一起」上升的交叉項）幾乎完全吻合。同一個展開也給出凱利成長率的二次近似：槓桿越高越不準，滿凱利時每年高估約 2 個百分點，其中 0.35 個百分點來自被捨去的日均報酬平方，其餘來自三階以上的高階項。

KNOWLEDGE CHECK 2

At $\mathbf{f} = \mathbf{0}$ the growth function has gradient $\nabla g = (0.2275, -0.0255)$. In which unit direction does g rise fastest, and at what rate? What is the rate of change along the budget-neutral unit direction $(1, -1)/\sqrt{2}$?

ANSWER

Steepest ascent is along $\nabla g / \|\nabla g\| = (0.9938, -0.1114)$, at rate $\|\nabla g\| = \sqrt{0.2275^2 + 0.0255^2} = 0.2289$. Along $(1, -1)/\sqrt{2}$ the rate is $(0.2275 + 0.0255)/\sqrt{2} = 0.1789$, smaller, as Cauchy-Schwarz requires.

4

UNCONSTRAINED OPTIMIZATION IN TWO VARIABLES

- ☐ locate and classify the critical points of a function of two variables with the first-order conditions and the 2×2 definiteness test

LM4 found optima of one-variable functions with two conditions: a zero derivative and the sign of the second derivative. Both carry over to two variables, the second with more care, because a surface can curve up in one direction and down in another.

First-order conditions

If h is differentiable and has a local maximum or minimum at an interior point $\mathbf{a}$, then every one-variable slice through $\mathbf{a}$ has zero slope there. In particular $h_x(\mathbf{a}) = h_y(\mathbf{a}) = 0$, that is, $\nabla h(\mathbf{a}) = \mathbf{0}$. Points where the gradient vanishes are **critical** (or stationary) **points**. As in one variable, the condition is necessary, not sufficient.

Second-order conditions: the 2×2 test

At a critical point the linear term of (7.2) vanishes, so for small steps $h(\mathbf{a} + \Delta) - h(\mathbf{a}) \approx \frac{1}{2} \Delta^\top H \Delta$, and the sign of this quadratic form decides the shape. If $h_{xx} \neq 0$, completing the square gives

$$\Delta^\top H \Delta = h_{xx} \left(\Delta_1 + \frac{h_{xy}}{h_{xx}} \Delta_2 \right)^2 + \frac{\det H}{h_{xx}} \Delta_2^2, \quad \det H := h_{xx} h_{yy} - h_{xy}^2. \quad (7.4)$$

If $\det H > 0$, both terms have the sign of h_{xx} , and the form is non-zero for every $\Delta \neq \mathbf{0}$. If $\det H < 0$, the two terms have opposite signs, and the form takes both signs. When h has continuous second partial derivatives, the remainder of (7.2) is small next to $\|\Delta\|^2$ near $\mathbf{a}$ and cannot overturn a definite form (stated without proof; Stewart, Clegg and Watson 2021, §14.7). When $\det H = 0$ the test is silent and higher-order terms decide: $x^2 + y^4$ has a minimum at the origin and $x^2 + y^3$ does not, with the same Hessian there. Exhibit 4 collects the cases. In matrix language, H is **positive definite** exactly when $h_{xx} > 0$ and $\det H > 0$, the 2×2 case of a criterion LM10 extends to N assets.

Exhibit 4: The second-derivative test at a critical point of $h(x, y)$

Condition	Quadratic form $\Delta^\top H \Delta$	Shape	Example in this module
$\det H > 0$ and $h_{xx} > 0$	positive definite	local minimum (bowl)	portfolio variance $V(w_1, w_2)$ with $ \rho < 1$
$\det H > 0$ and $h_{xx} < 0$	negative definite	local maximum (dome)	the growth model (7.3)
$\det H < 0$	indefinite	saddle point	$x^3 - 3xy + y^3$ at the origin (Example 3)
$\det H = 0$	semidefinite or degenerate	test inconclusive	V with $\rho = \pm 1$: a valley floor

Note: $\det H$ at the critical point; the test classifies local behaviour only.

Global optima and convexity

A local test says nothing about the rest of the surface. One condition does: if $\Delta^\top H \Delta \geq 0$ for every Δ at every point (the Hessian is positive semidefinite everywhere), h is **convex**, and any critical point is a global minimum (stated without proof; [LM34](#) develops convexity). Quadratic functions have a constant Hessian, so for them the local test is global.

Portfolio variance has $H = 2\Sigma$ with $\det H = 4\sigma_1^2\sigma_2^2(1 - \rho^2) \geq 0$: it is convex, and strictly convex when $|\rho| < 1$. The growth model (7.3) has $H = -\Sigma$ and is concave.

EXAMPLE 3

A saddle and a local minimum

Find and classify the critical points of $h(x, y) = x^3 - 3xy + y^3$.

SOLUTION

$h_x = 3x^2 - 3y = 0$ gives $y = x^2$, and $h_y = 3y^2 - 3x = 0$ gives $x = y^2$. Hence $x = x^4$, so $x = 0$ or $x = 1$, and the critical points are $(0, 0)$ and $(1, 1)$. The second derivatives are $h_{xx} = 6x$, $h_{yy} = 6y$ and $h_{xy} = -3$.

At $(0, 0)$: $\det H = 0 \cdot 0 - 9 = -9 < 0$, a saddle. Indeed $h(t, -t) = 3t^2 > 0$ while $h(t, t) = 2t^3 - 3t^2 < 0$ for small $t \neq 0$.

At $(1, 1)$: $h_{xx} = 6 > 0$ and $\det H = 36 - 9 = 27 > 0$, a local minimum with $h(1, 1) = -1$. It is not global, because $h(x, 0) = x^3 \rightarrow -\infty$ as $x \rightarrow -\infty$. Without convexity, a local test cannot certify a global optimum.

PITFALL When $\det H = 0$: perfectly correlated assets

With $|\rho| = 1$ the variance Hessian 2Σ has $\det H = 0$. The minimum is then a whole line of weights (a valley floor) or a riskless long-short hedge. The growth model (7.3) is then unbounded unless $SR_1 = SR_2$ (for $\rho = +1$) or $SR_1 = -SR_2$ (for $\rho = -1$), where $SR_i = e_i/\sigma_i$ is asset i 's annualized Sharpe ratio. Otherwise a riskless spread with a positive excess return can be levered without limit.

Real assets are never exactly collinear, but correlations near ± 1 make $\det H$ tiny. Optimizers then return huge, unstable long-short weights; [LM9](#) and [LM11](#) treat this ill-conditioning.

KNOWLEDGE CHECK 3

Classify the critical point $(0, 0)$ of $h(x, y) = x^2 + 3xy + 2y^2$. Every coefficient is positive; is the origin a minimum?

ANSWER

$h_{xx} = 2, h_{yy} = 4, h_{xy} = 3$, so $\det H = 8 - 9 = -1 < 0$: a saddle, not a minimum. For example $h(3t, -2t) = 9t^2 - 18t^2 + 8t^2 = -t^2 < 0 < h(t, 0) = t^2$. Positive coefficients do not make a quadratic form positive definite; the cross term must satisfy $h_{xy}^2 < h_{xx}h_{yy}$.

5 EQUALITY CONSTRAINTS: LAGRANGE MULTIPLIERS

- ☐ solve equality-constrained problems with Lagrange multipliers, and interpret the multiplier as a shadow price

Most portfolio problems come with constraints: the weights must sum to one, the volatility must hit a target, an order must be completed. When a constraint can be solved for one variable, substitution reduces the problem to LM4. Lagrange's method works without solving the constraint, and it delivers a bonus: the price of the constraint.

The tangency condition

Consider optimizing $h(x, y)$ subject to $\psi(x, y) = b$, and walk along the constraint curve. At a constrained optimum the slope of h along the curve must be zero, so ∇h is perpendicular to the curve there. The constraint curve is a level curve of ψ , so $\nabla \psi$ is perpendicular to it too (Section 3). Two vectors perpendicular to the same curve are parallel: at the optimum, the level curve of h just touches the constraint curve. Exhibit 5 shows this for the minimum-variance portfolio.

Algebraically, suppose $\psi_y \neq 0$ at the optimum. Near it the constraint defines y as a function of x (the implicit function theorem, stated without proof), and differentiating $\psi(x, y(x)) = b$ gives $\psi_x + \psi_y y'(x) = 0$, so $y' = -\psi_x/\psi_y$. The one-variable function $F(x) = h(x, y(x))$ has a critical point at the optimum:

$$F'(x) = h_x + h_y y'(x) = h_x - h_y \frac{\psi_x}{\psi_y} = 0.$$

Writing $\lambda := h_y/\psi_y$, this says $h_x = \lambda\psi_x$, and $h_y = \lambda\psi_y$ holds by definition of λ .

THEOREM Lagrange multiplier condition

Let h and ψ have continuous partial derivatives, and let (x^*, y^*) be a local maximum or minimum of h subject to $\psi(x, y) = b$, with $\nabla\psi(x^*, y^*) \neq \mathbf{0}$. Then there is a number λ^* with

$$\nabla h(x^*, y^*) = \lambda^* \nabla\psi(x^*, y^*), \quad \psi(x^*, y^*) = b. \quad (7.5)$$

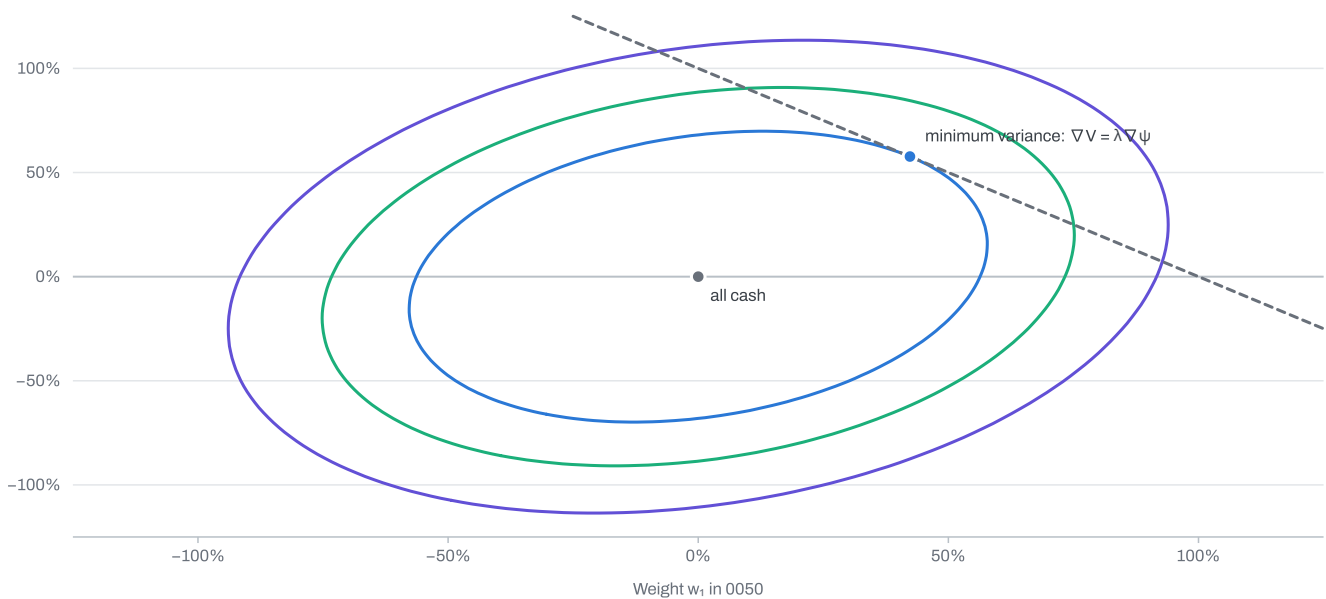
Equivalently, (x^*, y^*, λ^*) is a critical point of the **Lagrangian** $\mathcal{L}(x, y, \lambda) := h(x, y) - \lambda(\psi(x, y) - b)$.

The three conditions $\mathcal{L}_x = \mathcal{L}_y = \mathcal{L}_\lambda = 0$ give candidates; whether a candidate is a maximum, a minimum or neither needs a second-order or a global argument. A convex objective (for a minimum) or a concave one (for a maximum) with a linear constraint makes a candidate the global answer, as in the minimum-variance problems of Section 6; Section 7 states the general result. A curved constraint can produce several candidates: Example 5 has two.

Exhibit 5: Level curves of portfolio variance and the budget line, 2017–2021 estimates

— $\sigma_p = 9.22\%$ (touches the budget line) — $\sigma_p = 12\%$ — $\sigma_p = 15\%$ --- Budget line $w_1 + w_2 = 1$

Weight w_2 in 00679B (2017–2021 estimates)



Note: ellipses are sets of weights with equal volatility; the dashed line is the budget $\psi(\mathbf{w}) = w_1 + w_2 = 1$. The smallest ellipse that meets the line touches it at the minimum-variance weights (42.3%, 57.7%). There ∇V , perpendicular to the ellipse, is parallel to $\nabla\psi = (1, 1)$, perpendicular to the line.

The multiplier is a shadow price

RESULT Shadow price (envelope theorem)

Let $J(b) := h(x^*(b), y^*(b))$ be the optimal value when the constraint level is b , and suppose the solution moves smoothly with b . Then

$$J'(b) = \lambda^*(b). \quad (7.6)$$

Proof. By the chain rule and (7.5), $J'(b) = h_x x^{*'} + h_y y^{*'} = \lambda^* (\psi_x x^{*'} + \psi_y y^{*'})$. Differentiating the identity $\psi(x^*(b), y^*(b)) = b$ shows that the bracket equals 1. ■

This is the envelope theorem of LM4 (4.16) applied to the Lagrangian, whose partial derivative in b is λ . So $\Delta J \approx \lambda^* \Delta b$, with an error of second order in Δb . The units of λ^* are units of objective per unit of constraint: NT\$ per contract, return per unit of risk, variance per unit of capital.

EXAMPLE 4

Splitting a TX order between two sessions

A trader must buy $Q = 100$ TX contracts, x in the day session and y after hours. Model the temporary-impact cost as $C = k_d x^2 + k_n y^2$ with illustrative coefficients $k_d = 2$ and $k_n = 5$ (NT\$ per contract squared), the night book being thinner. Find the cost-minimizing split and interpret the multiplier.

SOLUTION

$\mathcal{L} = 2x^2 + 5y^2 - \lambda(x + y - 100)$. The conditions $\mathcal{L}_x = 4x - \lambda = 0$, $\mathcal{L}_y = 10y - \lambda = 0$ and $x + y = 100$ give $4x = 10y = \lambda$, so $x^* = 71.43$, $y^* = 28.57$ and $\lambda^* = 285.7$, in NT\$ per contract.

The multiplier is the **marginal cost** of the last contract, and the optimum equalizes it across sessions: if one session were cheaper at the margin, moving a contract there would cut the total.

Shadow price: the optimal cost is $C^*(Q) = \frac{k_d k_n}{k_d + k_n} Q^2 = \frac{10}{7} Q^2$, so $C^*(100) = 14,286$ (NT\$) and $C^{*'}(100) = \frac{20}{7} \cdot 100 = 285.7 = \lambda^*$, as (7.6) requires. The 101st contract adds NT\$287.1 exactly; the difference from λ^* is the second-order term $\frac{10}{7} \approx 1.4$.

In whole contracts a 71/29 split costs NT\$14,287, essentially optimal; an equal split costs NT\$17,500, 22.5% more.

EXAMPLE 5

A risk budget across two strategies

A desk runs a TX trend strategy (expected excess return $e_1 = 8\%$, volatility $\sigma_1 = 16\%$) and a futures-spot basis strategy ($e_2 = 3\%$, $\sigma_2 = 4\%$), with correlation $\rho = 0.2$ — illustrative figures.

Exposures f_1, f_2 per unit of capital can be levered. Maximize expected excess return subject to a volatility budget $\sigma_b = 10\%$, and interpret the multiplier.

SOLUTION

Maximize $f_1 e_1 + f_2 e_2$ subject to $\psi(\mathbf{f}) := f_1^2 \sigma_1^2 + 2f_1 f_2 \sigma_{12} + f_2^2 \sigma_2^2 = \sigma_b^2$, with $\sigma_{12} = 0.2 \times 0.16 \times 0.04 = 0.00128$. The Lagrangian conditions are

$$e_1 = 2\lambda(\sigma_1^2 f_1 + \sigma_{12} f_2), \quad e_2 = 2\lambda(\sigma_{12} f_1 + \sigma_2^2 f_2).$$

This is a 2×2 linear system in (f_1, f_2) ; eliminating one unknown (Cramer's rule, LM9) gives $f_1 \propto \sigma_2^2 e_1 - \sigma_{12} e_2$ and $f_2 \propto \sigma_1^2 e_2 - \sigma_{12} e_1$, a ratio of 1 to 7.43. The constraint is an ellipse, so scaling to 10% volatility gives two candidates, $\pm(0.274, 2.038)$, with $\lambda = \pm 4.15$. The minus sign is the minimum, an excess return of -8.31% . The maximum is $\mathbf{f} = (0.274, 2.038)$, with $\lambda > 0$: the low-volatility basis trade is levered about two times. Posed as $\psi(\mathbf{f}) \leq \sigma_b^2$, the problem rules out the second candidate by the sign condition of Section 7.

The expected excess return is 8.31%, against 5.0% for the trend strategy alone at 10% volatility and 7.5% for the basis strategy alone. It equals $\sigma_b \cdot \text{SR}^*$, where $\text{SR}^* = 0.831$ is the best attainable Sharpe ratio, derived in Section 6.

The multiplier on the variance budget is $\lambda^* = \text{SR}^* / (2\sigma_b) = 4.15$. By (7.6), the optimal excess return E^* rises by 4.15 per unit of variance budget. In volatility units $dE^* / d\sigma_b = \text{SR}^*$, so raising the budget from 10% to 11% adds 0.83 percentage points. **The shadow price of risk is the best attainable Sharpe ratio.**

導讀 拉格朗日乘數：限制式的影子價格

在限制條件下求極值，最適點一定落在目標函數的等高線與限制曲線「相切」之處，此時兩者的梯度平行： $\nabla h = \lambda \nabla \psi$ 。乘數 λ 不只是解題工具，它是「影子價格」：把限制放寬一單位，最適值大約改變 λ ，誤差是二階的。拆單例子裡， λ 是最後一口台指期的邊際成本，日盤與夜盤的邊際成本必須相等，否則把一口移到便宜的時段就能省錢。風險預算例子裡，以波動度計的乘數就是可達到的最高夏普比率。看到乘數，先問它的單位，再問放寬限制值不值得。

KNOWLEDGE CHECK 4

In Example 4, night-session liquidity improves to $k_n = 3$ (NT\$ per contract squared). Find the new split and multiplier, and the approximate cost of a 101st contract.

ANSWER

Now $4x = 6y = \lambda$ with $x + y = 100$: $x^* = 60$, $y^* = 40$ and $\lambda^* = 240$, in NT\$ per contract. The 101st contract costs about NT\$240 (exactly $\frac{6}{5}(101^2 - 100^2) = 241.2$, since $C^*(Q) = \frac{2.3}{5}Q^2$).

6

TWO-ASSET PORTFOLIOS BY CALCULUS

- ☐ derive the two-asset minimum-variance weight and explain how correlation determines the diversification benefit
- ☐ derive the two-asset tangency weights and Kelly fractions, and relate them to $\Sigma^{-1}(\mu - r_f \mathbf{1})$

This section applies Sections 2–5 to two assets, in the tradition of Markowitz (1952) and of Merton's (1972) closed forms: the minimum-variance weight, in and out of sample, then the tangency portfolio and the Kelly fractions.

Minimum variance, by substitution and by Lagrange

Substitution. With w in asset 1 and $1 - w$ in asset 2, the variance is a one-variable function,

$$V(w) = w^2\sigma_1^2 + (1 - w)^2\sigma_2^2 + 2w(1 - w)\sigma_{12}.$$

Its derivative is $V'(w) = 2w\sigma_1^2 - 2(1 - w)\sigma_2^2 + 2(1 - 2w)\sigma_{12}$. Setting it to zero:

$$w^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}. \quad (7.7)$$

The second derivative is $V''(w) = 2\sigma_d^2$, where $\sigma_d^2 := \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} = \text{Var}(R_1 - R_2)$ (Section 2). It is positive unless the two returns differ by a constant, so w^* is the global minimum. Substituting w^* back:

$$\sigma_{\min}^2 = V(w^*) = \frac{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}{\sigma_d^2} = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \quad (7.8)$$

Because V is a quadratic in w with second derivative $2\sigma_d^2$, it can be written exactly around its minimum:

$$V(w) = \sigma_{\min}^2 + \sigma_d^2(w - w^*)^2. \quad (7.9)$$

Lagrange. Minimize $V(w_1, w_2)$ subject to $w_1 + w_2 = 1$. With $\mathcal{L} = V - \lambda(w_1 + w_2 - 1)$, the conditions are

$$2(w_1\sigma_1^2 + w_2\sigma_{12}) = \lambda, \quad 2(w_1\sigma_{12} + w_2\sigma_2^2) = \lambda.$$

Subtracting gives $w_1(\sigma_1^2 - \sigma_{12}) = w_2(\sigma_2^2 - \sigma_{12})$, which with $w_1 + w_2 = 1$ is (7.7) again. The two conditions say more: at the minimum, **each asset has the same covariance with the portfolio**.

Multiply the first condition by w_1 , the second by w_2 and add: $2V = \lambda(w_1 + w_2) = \lambda$, so $\lambda^* = 2\sigma_{\min}^2$. That is the shadow price (7.6) of the budget: investing b units in the same proportions gives variance $b^2\sigma_{\min}^2$, whose derivative at $b = 1$ is $2\sigma_{\min}^2$.

Correlation and diversification

Formula (7.8) shows how correlation sets the benefit of diversification.

- $\rho = +1$. Then $\sigma_{\min}^2 = 0$ only through a short position: $w^* = -\sigma_2/(\sigma_1 - \sigma_2)$ when $\sigma_1 > \sigma_2$. Long-only, volatility is linear in w , and there is no diversification.
- $\rho = 0$. Then $w^* = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$: each asset is weighted by its inverse variance.
- $\rho = -1$. Then $w^* = \sigma_2/(\sigma_1 + \sigma_2)$ and $\sigma_{\min} = 0$: a perfect hedge without shorting.

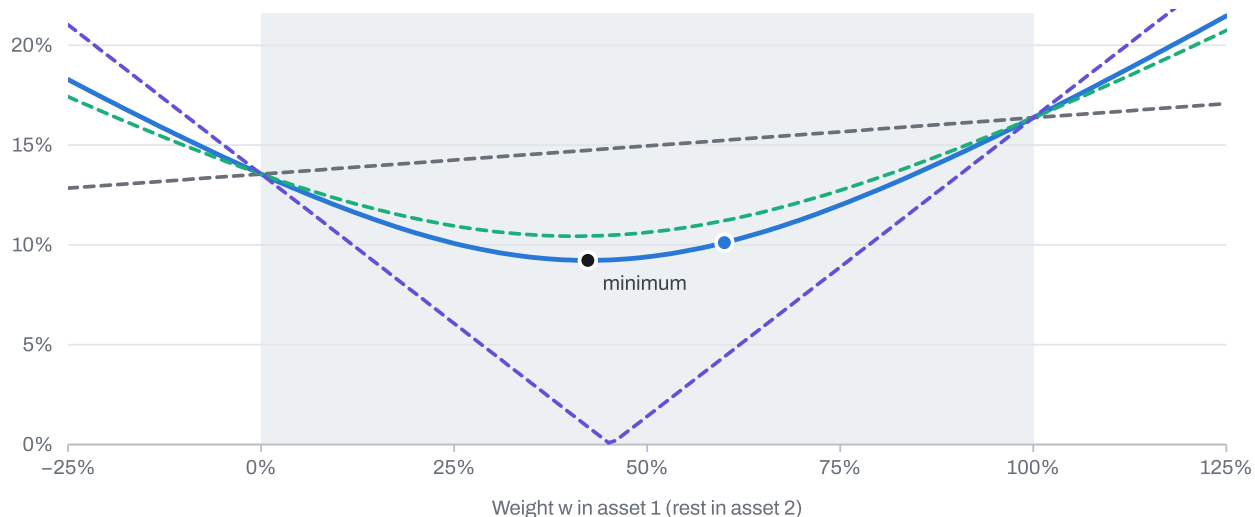
The minimum-variance portfolio shorts the riskier asset when the numerator of (7.7) turns negative, $\sigma_2^2 < \sigma_{12}$, that is, when $\rho > \sigma_2/\sigma_1$ (asset 1 being the riskier one). For 0050 and 00679B the ratio is $13.50/20.11 = 0.67$, far above any correlation in the sample, so the minimum is interior.

The envelope theorem of LM4 (4.16) prices a change in ρ without differentiating (7.8): because $V'(w^*) = 0$, $d\sigma_{\min}^2/d\rho = \partial V/\partial\rho = 2w_1^*w_2^*\sigma_1\sigma_2$ at w^* . With the 2017–2021 estimates this is 0.0108, so a correlation 0.1 higher raises $\sigma_{\min}^2$ by 0.00108 (exact to that precision) and $\sigma_{\min}$ from 9.22% to 9.79%.

Two-asset explorer: weight, correlation and the minimum-variance point

INTERACTIVE

— $\sigma_p(w)$ at $\rho = -0.22$ --- $\rho = +1$ - - - $\rho = 0$ - - - $\rho = -1$



Inputs $\sigma_1 \cdot \sigma_2 \cdot \rho$

16.4% · 13.5% · -0.22

σ_p at $w = 60\%$

10.11%

Weighted average of σ

15.24%

Minimum-variance w^*

42.3%

Minimum σ_p

9.22%

Multiplier $\lambda = 2\sigma^2 \min$

0.0170

Long-only optimum

interior, slack

The portfolio holds w in asset 1 and $1 - w$ in asset 2, rebalanced to these weights. Coloured curve: $\sigma_p(w)$ at the chosen ρ ; dashed curves: the same two assets at $\rho = +1, 0$ and -1 . The shaded band is the long-only region $0 \leq w \leq 1$; when the minimum lies outside it, the long-only optimum is the nearest corner (Section 7).

Back to the case: in sample and out of sample

EXAMPLE 6

The minimum-variance mix estimated at the end of 2021

Over 2017–2021 the annualized estimates are $\sigma_1^2 = 0.026800$ (0050), $\sigma_2^2 = 0.018353$ (00679B) and $\sigma_{12} = -0.004908$, so $\sigma_1 = 16.37\%$, $\sigma_2 = 13.55\%$ and $\rho = -0.221$. Compute w^* , $\sigma_{\min}$ and λ^* .

SOLUTION

The numerator of (7.7) is $\sigma_2^2 - \sigma_{12} = 0.023261$ and the denominator $\sigma_d^2 = 0.054969$, so $w^* = 42.3\%$ in 0050 and 57.7% in 00679B.

From (7.8), $\sigma_{\min}^2 = (0.026800 \times 0.018353 - 0.004908^2)/0.054969 = 0.008510$, so $\sigma_{\min} = 9.22\%$, and $\lambda^* = 2\sigma_{\min}^2 = 0.0170$.

The in-sample calculation is correct: the 42/58 mix's realized volatility over 2017–2021 is exactly 9.22%, because the sample variance of a fixed mix is V at the sample moments (Section 2).

The calculation was right; the inputs changed, as Exhibit 6 shows year by year. In 2022 both funds fell by about 21–22% while their daily correlation was -0.06 : a correlation near zero does not prevent a joint loss in a year when both assets drift down.

From 2023 the trailing one-year correlation moved between -0.29 (April 2025) and $+0.25$ (December 2023) (Exhibit 7). At lower frequencies the correlation turned positive: between the two windows it rose from -0.29 to $+0.21$ for monthly returns and from -0.15 to $+0.09$ for weekly ones.

Exhibit 6: 0050 and 00679B by calendar year

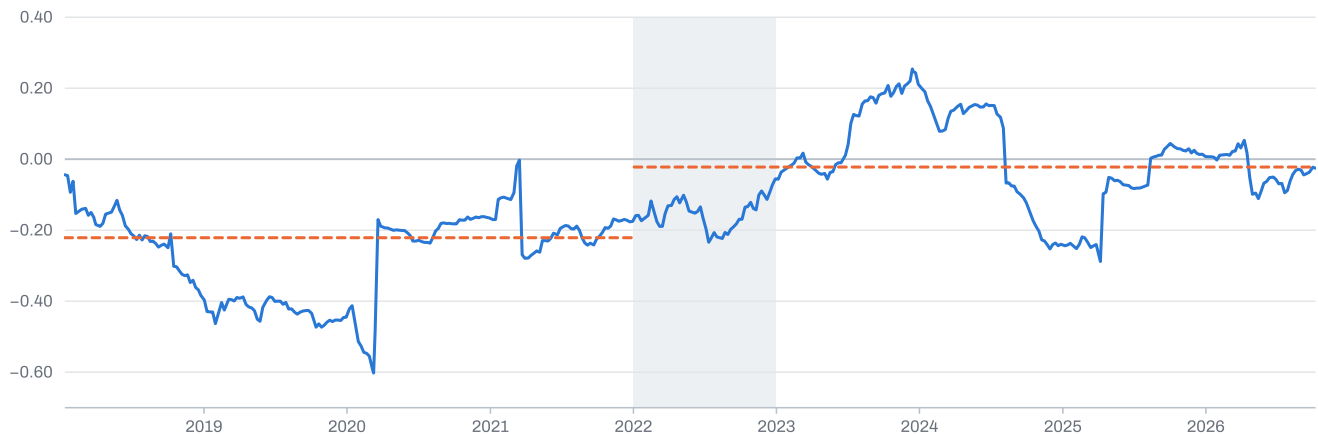
Year	Days	0050 return	00679B return	0050 volatility	00679B volatility	Correlation	Minimum-variance weight in 0050
2017 (from 17 Jan)	235	+17.3%	−0.2%	9.3%	8.4%	−0.04	45.2%
2018	247	−5.0%	+0.7%	16.3%	9.0%	−0.39	30.0%
2019	242	+33.5%	+13.7%	12.2%	12.2%	−0.43	50.1%
2020	245	+31.1%	+8.1%	22.8%	21.3%	−0.17	47.1%
2021	244	+22.0%	−6.4%	17.6%	12.5%	−0.17	36.1%
2022	246	−21.3%	−22.1%	21.6%	16.1%	−0.06	36.4%
2023	239	+27.4%	+3.1%	13.9%	12.6%	+0.20	43.8%
2024	242	+48.7%	−2.9%	24.6%	11.0%	−0.24	21.8%
2025	238	+36.9%	−0.6%	24.5%	15.7%	+0.01	28.9%
2026 (to 8 Oct)	185	+78.7%	−7.7%	30.6%	9.6%	+0.06	7.5%

Note: days are common trading dates (2025 omits 11–17 June, when 0050 was halted for its split); total returns with dividends reinvested; volatilities annualized with A ; correlations of daily returns; the last column applies (7.7) to each year's own unrounded estimates. Source: FinMind, retrieved 10 October 2026; [data/lm07_stats.py](#).

Exhibit 7: Trailing one-year correlation of daily returns, 0050 and 00679B

— Trailing 243-day correlation - - - Full-window correlation (2017–2021, 2022–2026)

Correlation of daily returns, trailing 243 trading days



Note: correlation over the trailing 243 trading days; the shaded band is calendar 2022. The dashed line marks the correlation over each full window, -0.22 for 2017–2021 and -0.02 for 2022–2026.

Identity (7.9) now splits the miss exactly, and Exhibit 8 draws it. With the test-window inputs, the lowest attainable variance was $\sigma_{\min}^2 = 133.15 \%^2$ at $w^* = 25.3\%$, and the coefficient of the quadratic term was $\sigma_d^2 = 0.07419$, half the curvature $V'' = 0.1484$. The in-sample weight was 0.17067 (17.07 points) too high, which adds $0.07419 \times 0.17067^2 = 21.61 \%^2$. Together, $133.15 + 21.61 = 154.76 \%^2$, the 12.44% volatility the 42/58 mix actually realized.

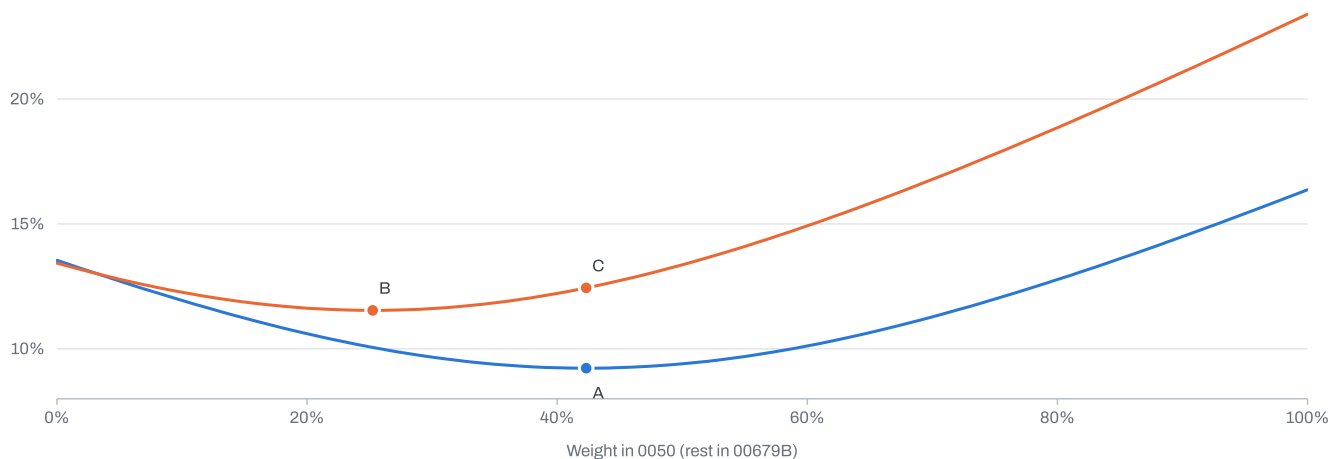
RESULT Resolution of the case

The 3.2-point volatility miss had two sources. The minimum itself moved, from 9.2% to 11.5%, because 0050 became more volatile and the correlation rose toward zero. The weight error added only 0.9 points, because by (7.9) the penalty is quadratic in the weight error: on the test-window curve, a 5-point error costs 0.08 points of volatility and a 10-point error 0.32. The realized 12.44% was still below 00679B's 13.43% and far below 0050's 23.40%, so diversification survived, only smaller than estimated.

Exhibit 8: Portfolio volatility against the weight in 0050, estimated in and out of sample

— Estimated on 2017–2021 — Realized in 2022–2026

Annualized volatility of the mix



Note: volatility of a daily-rebalanced mix from (7.8)–(7.9) with each window's moments. A: in-sample minimum, 9.22% at $w = 42.3\%$. B: ex-post minimum, 11.54% at $w = 25.3\%$. C: the in-sample weight in the test window, 12.44%.

Minimum variance ignores expected returns, which mattered here: over the test window's 4.77 calendar years the 42/58 mix compounded at 8.67% a year and 0050 at 31.12%, while 00679B lost 6.78% a year.

The tangency portfolio

The portfolio with the highest Sharpe ratio, $(\mu_p - r_f)/\sigma_p$, is the **tangency portfolio**: on a chart of expected return against volatility, it is where a line from the risk-free rate touches the set of risky portfolios (LM33).

Example 5 already solved this problem. Maximizing expected excess return for a fixed volatility gave exposures proportional to $(\sigma_2^2 e_1 - \sigma_{12} e_2, \sigma_1^2 e_2 - \sigma_{12} e_1)$, whatever the budget. Scaling exposures by a positive constant leaves the Sharpe ratio unchanged, so the same direction maximizes it. Normalizing it to sum to one gives the tangency weights

$$w_1^{\text{tan}} = \frac{\sigma_2^2 e_1 - \sigma_{12} e_2}{\sigma_2^2 e_1 + \sigma_1^2 e_2 - \sigma_{12}(e_1 + e_2)}, \quad w_2^{\text{tan}} = 1 - w_1^{\text{tan}}. \quad (7.10)$$

The highest Sharpe ratio follows by evaluating the ratio on that direction. With $\det \Sigma = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2$, it is $\text{SR}^{*2} = (\sigma_2^2 e_1^2 - 2\sigma_{12} e_1 e_2 + \sigma_1^2 e_2^2) / \det \Sigma$. Dividing numerator and denominator by $\sigma_1^2 \sigma_2^2$:

$$\text{SR}^{*2} = \frac{\text{SR}_1^2 + \text{SR}_2^2 - 2\rho \text{SR}_1 \text{SR}_2}{1 - \rho^2}. \quad (7.11)$$

Subtracting SR_1^2 from (7.11) leaves $(\text{SR}_2 - \rho \text{SR}_1)^2 / (1 - \rho^2) \geq 0$. So adding asset 2 raises the best Sharpe ratio unless $\text{SR}_2 = \rho \text{SR}_1$, which is exactly when its weight in (7.10) is zero. In Example 5, $\text{SR}^{*2} = (0.25 + 0.5625 - 0.15) / 0.96 = 0.690$, so $\text{SR}^* = 0.831$.

When the denominator of (7.10) is near zero the tangency weights explode, as the test-window estimates in Exhibit 9 show.

The two-asset Kelly fractions

The Kelly investor maximizes the growth rate. With the quadratic model (7.3), the first-order conditions are

$$g_{f_1} = e_1 - (\sigma_1^2 f_1 + \sigma_{12} f_2) = 0, \quad g_{f_2} = e_2 - (\sigma_{12} f_1 + \sigma_2^2 f_2) = 0,$$

a 2×2 linear system. Eliminating one unknown (Cramer's rule, [LM9](#)) gives

$$f_1^* = \frac{\sigma_2^2 e_1 - \sigma_{12} e_2}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}, \quad f_2^* = \frac{\sigma_1^2 e_2 - \sigma_{12} e_1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}. \quad (7.12)$$

The Hessian of (7.3) is $-\Sigma$, with $g_{f_1 f_1} = -\sigma_1^2 < 0$ and $\det(-\Sigma) = \sigma_1^2 \sigma_2^2 (1 - \rho^2) > 0$ whenever $|\rho| < 1$: by Exhibit 4, a maximum, and global because (7.3) is concave. Three readings of (7.12) connect it to the rest of the book.

- Matrix form.** A 2×2 matrix times a vector is the vector of each row's dot product with it. The matrix $\Sigma^{-1} = \frac{1}{\det \Sigma} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}$ undoes multiplication by Σ ([LM9](#)), and $\Sigma^{-1} \mathbf{e}$ is exactly (7.12). So $\mathbf{f}^* = \Sigma^{-1}(\boldsymbol{\mu} - r_f \mathbf{1})$, the multi-asset Kelly rule of [LM43](#).
- Kelly is tangency, levered.** The numerators of (7.12) and (7.10) are the same. Hence $\mathbf{f}^* = L^* \mathbf{w}^{\text{tan}}$, with total exposure $L^* = f_1^* + f_2^*$. The Kelly investor holds the tangency portfolio and chooses only how much of it to hold.
- Growth at the optimum.** At $\mathbf{f}^*$ the conditions give $\mathbf{f}^{*\top} \Sigma \mathbf{f}^* = \mathbf{f}^{*\top} \mathbf{e} = \text{SR}^{*2}$, so (7.3) becomes $g^* = r_f + \frac{1}{2} \text{SR}^{*2}$. With one asset, $f^* = e/\sigma^2 = (\mu - r_f)/\sigma^2$: [LM4](#) derived μ/σ^2 with no financing cost, and [LM42](#) adds r_f .

EXAMPLE 7

Kelly fractions for 0050 and 00679B, full sample

Use the full-sample estimates of Example 2, arithmetic means $\mu_1 = 24.251\%$ and $\mu_2 = -1.047\%$, and an illustrative $r_f = 1.5\%$, so $e_1 = 22.751\%$ and $e_2 = -2.547\%$. Compute the Kelly fractions (7.12), the tangency weights, SR^* and the growth rate of the quadratic model, and compare with the exact sample growth.

SOLUTION

$\det \Sigma = 0.040425 \times 0.018214 - 0.002866^2 = 7.281 \times 10^{-4}$. The numerators are $0.018214(0.22751) - (-0.002866)(-0.02547) = 0.0040709$ and $0.040425(-0.02547) - (-0.002866)(0.22751) = -0.0003776$. So $f_1^* = 5.59$ and $f_2^* = -0.52$: in-sample, the Kelly investor would borrow to hold 5.6 times capital in 0050 and short half a unit of 00679B.

Total exposure is $L^* = 5.07$. The tangency weights (7.10) divide each numerator by their sum, 0.0036933 : $\mathbf{w}^{\text{tan}} = (110.2\%, -10.2\%)$. With $\text{SR}_1 = 1.132$, $\text{SR}_2 = -0.189$ and $\rho = -0.106$, (7.11) gives $\text{SR}^* = 1.134$, hardly above 0050's own 1.132: in this sample the bond adds almost nothing to the best Sharpe ratio.

The model's growth rate is $g^* = 1.5\% + \frac{1}{2}(1.134)^2 = 65.8\%$ a year. The exact sample growth at $\mathbf{f}^*$ is lower, 63.7% (Exhibit 3). The exact optimum, found numerically in Section 8, is (5.32, -0.26) with 64.0%. On 7 April 2025 alone, the position at $\mathbf{f}^*$ would have lost 56.9% of capital, and the exact optimum 53.7%.

Exhibit 9: Kelly fractions and tangency weights estimated on three windows

Window	e_1 (0050)	e_2 (00679B)	σ_1	σ_2	ρ	f_1^*	f_2^*	$L^* =$ $f_1^* + f_2^*$	Tangency weights	SR^*
2017– 2021	17.2%	+2.3%	16.4%	13.5%	-0.22	7.00	+3.14	10.14	69% / +31%	1.13
2022– 2026	28.6%	-7.7%	23.4%	13.4%	-0.02	5.17	-4.06	1.11	466% / -366%	1.34
2017– 2026	22.8%	-2.5%	20.1%	13.5%	-0.11	5.59	-0.52	5.07	110% / -10%	1.13

Note: e_i are arithmetic annualized means minus an illustrative $r_f = 1.5\%$; fractions from (7.12), tangency weights from (7.10) and SR^* from (7.11), all computed from unrounded estimates. In-sample estimates, shown to illustrate instability, not as recommendations.

PITFALL Optimal weights fitted to the past are not recommendations

Exhibit 9 shows how fragile the inputs are. The 00679B fraction flips from 3.1 times long to 4.1 times short between windows, and the tangency weights go from a sensible 69/31 to an absurd 466/-366. Expected returns are the culprit. The standard error of a mean annual return, its typical estimation error, is about $\sigma/\sqrt{\text{years of data}}$ (LM20). Over the 4.95 calendar years of 2017–2021 that is 7.36 points for 0050, 39% of its 18.7% mean, and 6.09 points for 00679B, more than its 3.8% mean.

Chopra and Ziemba (1993) found errors in means far more damaging to optimal portfolios than errors in variances or covariances. Minimum-variance weights avoid means altogether, yet Exhibit 6 shows that variances and correlations move too. LM35 (estimation error and shrinkage) and LM45 (Kelly under uncertainty) take up the repair.

導讀 最小變異數、切點組合與凱利比例

兩資產最小變異數權重 $w^* = (\sigma_2^2 - \sigma_{12})/(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$ 只用到波動度與共變異數，不需要預期報酬。用 2017–2021 的資料算出 0050 占 42.3%、預測波動 9.22%；2022–2026 實際卻是 12.44%。差距主要來自「輸入值變了」：0050 波動從 16.4% 升到 23.4%，相關係數從 -0.22 升到 -0.02（月報酬更從 -0.29 翻正為 +0.21）；權重估錯 17 個百分點只多付 0.9 個百分點的波動，因為變異數曲線在最低點附近很平坦。切點組合與凱利比例還需要預期報酬，估計誤差大得多：00679B 的凱利比例在不同樣本期間從 +3.14 變成 -4.06。凱利就是「加了槓桿的切點組合」，但樣本內算出的最適解不是投資建議。

KNOWLEDGE CHECK 5

Two assets have the same volatility σ and correlation $\rho < 1$. Show that the minimum-variance weights are 50/50 and that $\sigma_{\min}^2 = \sigma^2(1 + \rho)/2$. What happens as $\rho \rightarrow -1$ and as $\rho \rightarrow 1$?

ANSWER

With $\sigma_1 = \sigma_2 = \sigma$ and $\rho < 1$, (7.7) gives $w^* = (\sigma^2 - \rho\sigma^2)/(2\sigma^2 - 2\rho\sigma^2) = 1/2$. Then $V = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{2}\rho\sigma^2 = \sigma^2(1 + \rho)/2$. As $\rho \rightarrow -1$ the risk vanishes (a perfect hedge); as $\rho \rightarrow 1$ it tends to σ^2 (no diversification). At $\rho = 0$ the volatility falls by the factor $1/\sqrt{2}$.

7

INEQUALITY CONSTRAINTS: THE KKT CONDITIONS

- ☐ apply the Karush–Kuhn–Tucker conditions to long-only, policy and leverage constraints, and identify which constraints bind

Real mandates are full of inequalities: no short sales, at least 60% in equities, exposure at most twice capital. LM4 handled one inequality in one variable: a maximum on an interval is either interior, with $h' = 0$, or at an endpoint where h' points outward. The Karush–Kuhn–Tucker conditions, found by Karush (1939) and again by Kuhn and Tucker (1951), extend that logic to several variables and constraints; Kjeldsen (2000) tells the history, and [LM34](#) applies the conditions to larger portfolios.

Slack or binding

At the optimum each inequality constraint is either **slack**, and could be deleted without changing the answer, or **binding**, and behaves like an equality with a multiplier. The multiplier's sign must say that the constraint pushes in the right direction: for a maximization with constraints $\psi_i(\mathbf{x}) \leq b_i$, relaxing a constraint can only help, so its shadow price cannot be negative.

THEOREM Karush–Kuhn–Tucker (KKT) conditions

Let h and every ψ_i have continuous partial derivatives. Maximize $h(\mathbf{x})$ subject to $\psi_i(\mathbf{x}) \leq b_i$, one inequality for each i . If $\mathbf{x}^*$ is a local maximum and the gradients of the binding constraints are linearly independent (for two of them: not parallel), there are multipliers ν_i with

$$\nabla h(\mathbf{x}^*) = \sum_i \nu_i \nabla \psi_i(\mathbf{x}^*), \quad \psi_i(\mathbf{x}^*) \leq b_i, \quad \nu_i \geq 0, \quad \nu_i (b_i - \psi_i(\mathbf{x}^*)) = 0. \quad (7.13)$$

If h is concave and every ψ_i is convex, these conditions are also sufficient for a global maximum. Stated without proof; see Nocedal and Wright (2006, Theorem 12.1) for necessity and Boyd and Vandenberghe (2004, §5.5.3) for sufficiency. For a minimization, apply (7.13) to $-h$. An equality constraint enters as in Section 5, with a multiplier of either sign, and must be linear for sufficiency.

The four parts are **stationarity**, **feasibility**, **dual feasibility** and **complementary slackness**. Stationarity says that the uphill direction ∇h is a non-negative combination of the binding constraints' outward normals, so every direction that would raise h leaves the feasible set. Complementary slackness says that each constraint either binds or has a zero multiplier. As in (7.6), ν_i is the rate at which the optimal value rises when b_i is relaxed.

The practical method is **enumeration**. Guess which constraints bind, solve stationarity together with those constraints as equalities, and keep the candidate only if it is feasible and every multiplier is non-negative. With two variables and three constraints there are at most eight binding sets to try.

EXAMPLE 8

Kelly with a leverage cap and no short sales

Maximize the growth model (7.3) with the full-sample estimates of Example 7 subject to $f_1 + f_2 \leq 2$, $f_1 \geq 0$ and $f_2 \geq 0$. The first constraint, illustrative, caps net (total) exposure at twice capital; it equals gross exposure $|f_1| + |f_2|$ only when both positions are long. Identify the binding constraints and interpret the multipliers.

SOLUTION

The unconstrained optimum $(5.59, -0.52)$ violates the cap and the no-short rule. Of the eight possible binding sets, the one with all three constraints is impossible ($f_1 = f_2 = 0$ breaks $f_1 + f_2 = 2$), which leaves seven; Exhibit 10 solves each. Only one passes every test: the cap and $f_2 \geq 0$ bind, at $\mathbf{f} = (2, 0)$.

Write the constraints as $f_1 + f_2 \leq 2$, $-f_1 \leq 0$ and $-f_2 \leq 0$, with multipliers ν_c , ν_1 and ν_2 .

Stationarity reads $g_{f_1} = \nu_c - \nu_1$ and $g_{f_2} = \nu_c - \nu_2$, with $\nu_1 = 0$ because $f_1 \geq 0$ is slack at $f_1 = 2$.

At $(2, 0)$, $g_{f_1} = e_1 - 2\sigma_1^2 = 0.22751 - 0.08085 = 0.1467$, so $\nu_c = 0.1467$. Next, $g_{f_2} = e_2 - 2\sigma_{12} = -0.02547 + 0.005732 = -0.0197$, so $\nu_2 = 0.1467 + 0.0197 = 0.1664$. Both multipliers are non-negative.

In-sample, at the margin each unit of exposure is worth 14.7 points of model growth a year, the figure to compare with a margin lender's spread over r_f . That is a first-order value: exactly, raising the cap to 2.1 adds 1.45 points, and to 3 adds 12.64.

ν_2 needs care. Relaxing $f_2 \geq 0$ to $f_2 \geq -0.01$ lets the investor short 1% of 00679B and, because the cap limits net exposure, hold 2.01 in 0050. Re-solving gives $(2.01, -0.01)$ and 0.166 points more growth, as ν_2 predicts. The extra 0050 adds 0.1465 and the short 0.0196: since $g_{f_2} = -0.0197 < 0$, selling 00679B raises growth. Under a gross cap $|f_1| + |f_2| \leq 2$ the short would use up cap instead: $(1.99, -0.01)$ loses 0.127 points, and $(2, 0)$ stays optimal.

With the exact growth function instead of (7.3), the optimum is the same and the cap's multiplier is 0.146 (Section 8).

Exhibit 10: KKT candidates for the capped, long-only Kelly problem, full sample

Binding constraints	Candidate (f_1, f_2)	Multipliers	Feasible	All multipliers ≥ 0	Model growth
none	(5.59, -0.52)	—	no: breaks the cap, shorts 00679B	yes (none)	65.76%
cap only	(4.58, -2.58)	$\nu_c = 0.035$	no: shorts 00679B	yes	60.42%
$f_2 \geq 0$ only	(5.63, 0.00)	$\nu_2 = 0.009$	no: breaks the cap	yes	65.52%
$f_1 \geq 0$ only	(0.00, -1.40)	$\nu_1 = -0.224$	no: shorts 00679B	no	3.28%
cap and $f_2 \geq 0$	(2.00, 0.00)	$\nu_c = 0.147, \nu_2 = 0.166$	yes	yes	38.92%
cap and $f_1 \geq 0$	(0.00, 2.00)	$\nu_c = -0.062, \nu_1 = -0.295$	yes	no	-7.24%
$f_1 \geq 0$ and $f_2 \geq 0$	(0.00, 0.00)	$\nu_1 = -0.228, \nu_2 = 0.025$	yes	no	1.50%

Note: model (7.3) with full-sample estimates and $r_f = 1.5\%$ (illustrative). The cap is on net exposure $f_1 + f_2$; "cap only" has gross exposure $4.585 + 2.585 = 7.17$. The optimum is the only candidate that is feasible with non-negative multipliers.

The same logic covers long-only minimum variance. When $\rho > \sigma_2/\sigma_1$, (7.7) puts a negative weight on the riskier asset; with no short sales the optimum is the corner that holds only the safer asset, and the multiplier is the slope of V there. The widget in Section 6 reports it, and Practice Problem 9 works an example.

PITFALL Checking only some of the KKT conditions

Every candidate in Exhibit 10 satisfies stationarity. Four are infeasible, and three of those beat the optimum's growth because they ignore a constraint. Two feasible candidates, $(0, 2)$ and $(0, 0)$, carry a negative multiplier: relaxing that constraint would *lower* growth, so it is not what holds the solution back. Feasibility alone leaves three candidates and the sign test alone four; only $(2, 0)$ passes both. Check all four conditions of (7.13), and state which constraints bind: the binding set is what changes when the mandate changes.

導讀 KKT 條件：限制式是緊還是鬆

不等式限制（不可放空、最低股票比重、曝險上限）在最適點只有兩種狀態：沒碰到（鬆弛），就當它不存在，乘數為 0；碰到了（緊），就當成等式處理，乘數必須 ≥ 0 ，表示這個限制確實擋住了你想去的方向。這就是互補鬆弛性：乘數與鬆弛量至少有一個是 0。實務解法是列舉哪些限制是緊的，對每個候選解檢查四個條件，缺一不可。槓桿上限的乘數是邊際上「多一單位曝險能多賺多少成長率」，樣本內約每年 14.7 個百分點，可以直接拿來和融資利差比較；這是一階估計，上限從 2 放寬到 3 實際多 12.64 個百分點。要注意例子裡的上限管的是淨曝險 $f_1 + f_2$ ：放空 00679B 會騰出額度加碼 0050；若上限管的是總曝險 $|f_1| + |f_2|$ ，放空反而會占用額度。

KNOWLEDGE CHECK 6

An investment policy, illustrative here, requires at least 60% in 0050. With the full-sample estimates of Example 2, does this floor bind for the minimum-variance mix? If so, find its multiplier and the cost of the policy.

ANSWER

Yes: (7.7) gives $w_1^* = 32.7\%$, so the floor binds at $\mathbf{w} = (0.6, 0.4)$. For the maximization of $-V$, with λ on the budget and ν on $-w_1 \leq -0.6$, stationarity reads $-V_{w_1} = \lambda - \nu$ and $-V_{w_2} = \lambda$. Hence $\nu = V_{w_1} - V_{w_2} = 0.0462 - 0.0111 = 0.0351 \geq 0$, the slope of V along the switch direction $(1, -1)$ of Example 2. A 59% floor would cut the variance by about $0.01\nu = 0.000351$ (exactly 0.000344). Against the unconstrained minimum of 10.64%, the policy costs 2.05 points of volatility (12.69%).

8 MULTIVARIABLE OPTIMIZATION IN PYTHON

- ☐ implement numerical gradients, Newton's method and constrained optimization in Python, and distinguish in-sample optimal weights from out-of-sample results

Closed forms such as (7.7) and (7.12) need quadratic problems in two variables; the exact growth function is not quadratic, and real problems have more assets and constraints. Three numerical tools cover most cases: finite-difference derivatives, Newton's method and a constrained solver.

Numerical gradients and Hessians

Central differences carry over from LM2 coordinate by coordinate, with an error of order δ^2 for a step δ . For a mixed partial derivative, difference in both coordinates (the lab applies this to the Hessian):

$$h_{xy} \approx \frac{h(x + \delta, y + \delta) - h(x + \delta, y - \delta) - h(x - \delta, y + \delta) + h(x - \delta, y - \delta)}{4\delta^2}.$$

The code below builds the exact growth function $g(f_1, f_2)$ of Section 3 from the lab's daily returns and checks its gradient at $\mathbf{f} = \mathbf{0}$ against the mean excess returns.

PYTHON

```
import numpy as np

# from Part 1 of the lab: daily total returns R1, R2 (0050, 00679B)
# and A = T/years ≈ 243.0543
rf = 0.015                                # illustrative r_f
rd = rf / A                                # daily financing rate r_f/A
X = np.column_stack([R1, R2]) - rd         # daily excess returns, T x 2

def g(f):
    """Annualized sample log growth of exposures f, financed at r_f."""
    return A * np.log1p(rd + X @ f).mean()

def num_grad(fun, x, d=1e-5):
    e = np.eye(len(x))
    return np.array([(fun(x + d * e[i]) - fun(x - d * e[i])) / (2 * d)
                      for i in range(len(x))])

print(num_grad(g, np.zeros(2)).round(4))    # [ 0.2275 -0.0255] ~ (e1, e2)
```

Gradient ascent and Newton's method

Gradient ascent repeats $\mathbf{f}_{k+1} = \mathbf{f}_k + \alpha \nabla g(\mathbf{f}_k)$: step uphill, re-measure the slope, repeat. For the quadratic model (7.3), whose Hessian is $-\Sigma$, let m and M be the smallest and largest curvatures $\mathbf{u}^\top \Sigma \mathbf{u}$ over unit directions: the roots of $x^2 - (\sigma_1^2 + \sigma_2^2)x + \det \Sigma = 0$, the eigenvalues of Σ (LM11). The best fixed step is $\alpha = 2/(m + M)$, and it shrinks the distance to the optimum at least by the factor $(\kappa - 1)/(\kappa + 1)$ per step, where $\kappa = M/m$ is the **condition number** (stated without proof; see Nocedal and Wright 2006). For the full-sample Σ , $m = 0.01785$ and $M = 0.04079$, so $\kappa = 2.29$, $\alpha = 34.1$ and the factor is 0.39.

Near its optimum the exact growth function is more curved, with curvatures 0.0198 and 0.0493. The same step therefore contracts the distance by $|1 - 34.1 \times 0.0493| = 0.68$ per step, and the growth shortfall by about $0.68^2 = 0.46$.

Newton's method maximizes the second-order Taylor model (7.2) of g at the current point. Setting the model's gradient with respect to the step Δ to zero gives the 2×2 linear system $H(\mathbf{f}_k)\Delta = -\nabla g(\mathbf{f}_k)$, solved as in Section 6, and $\mathbf{f}_{k+1} = \mathbf{f}_k + \Delta$: the two-variable version of LM4's iteration. On a quadratic it lands on the optimum in one step. On the exact growth function, the first step from $\mathbf{0}$ lands at (5.5621, -0.5156), near the model optimum (5.59, -0.52), because at $\mathbf{0}$ the Taylor model is essentially (7.3). Two more steps reach (5.3186, -0.2641).

PYTHON

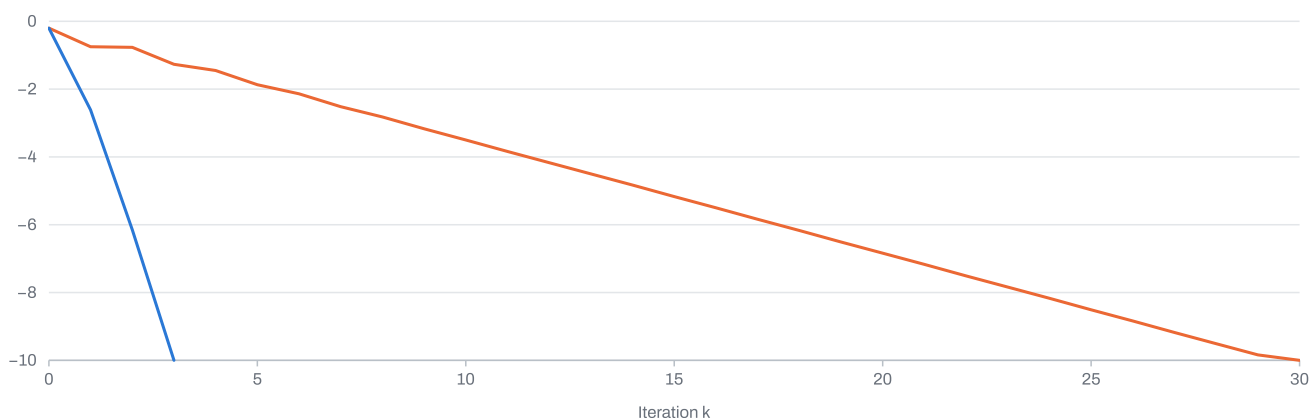
```
def grad_hess(f):
    """Exact gradient and Hessian of g:
    A*mean(X/(1+R_p)) and -A*mean(X X'/(1+R_p)^2)."""
    w = 1 / (1 + rd + X @ f)
    return A * (X * w[:, None]).mean(0), -A * (X.T * w**2) @ X / len(X)

f = np.zeros(2)
for k in range(1, 5):
    gr, H = grad_hess(f)
    f = f + np.linalg.solve(H, -gr)    # Newton step: H Delta = -grad g
    print(k, f.round(4), f"{g(f):.4%}")
# 1 [ 5.5621 -0.5156] 63.7851%
# 2 [ 5.3231 -0.2675] 64.0296%
# 3 [ 5.3186 -0.2641] 64.0296%
# 4 [ 5.3186 -0.2641] 64.0296%
```

Exhibit 11: Growth shortfall by iteration: gradient ascent versus Newton's method

— Gradient ascent, fixed step — Newton's method

Growth shortfall $g^* - g(f_k)$, $\log_{10}$ scale (-6 means 10^{-6})



Note: shortfall $g^* - g(f_k)$ of the exact sample growth from its maximum, 64.03% a year, starting at $f_0 = \mathbf{0}$; gradient ascent uses the fixed step $\alpha = 2/(m + M) \approx 34.1$. Each series ends at its first shortfall below 10^{-10} , drawn on the bottom edge: far smaller shortfalls approach the rounding error of the computed growth rate (about 10^{-16}) and vary from machine to machine.

Newton needs two iterations to bring the shortfall below 10^{-6} and three to go below 10^{-10} ; gradient ascent needs 18 and 30 (Exhibit 11). Newton's shortfall falls from 2.4×10^{-3} to 7.0×10^{-7} and then below 10^{-10} : the number of correct digits roughly doubles at each step. The price is the Hessian: N^2 second derivatives and a linear solve per step, cheap for two assets and expensive for two thousand. Gradient methods dominate large problems for that reason, though they struggle when κ runs into the thousands.

Constrained problems with `scipy.optimize.minimize`

For constraints, `scipy.optimize.minimize` with `method="SLSQP"` (sequential least-squares quadratic programming; Kraft 1988) handles bounds, equalities and inequalities. It solves a sequence of quadratic models, each a small KKT problem. Verify a multiplier the way (7.6) defines it: re-solve with the constraint relaxed and divide the change in the optimum by the change in the constraint. Recent SciPy versions also report the multipliers in the result object.

PYTHON

```
from scipy.optimize import minimize

def kelly_capped(cap):
    """Maximize exact sample growth subject to  $f_1 + f_2 \leq \text{cap}$  and  $f \geq 0$ ."""
    res = minimize(lambda f: -g(f), x0=[0.5, 0.5], method="SLSQP",
                   bounds=[(0, None), (0, None)],
                   constraints=[{"type": "ineq",
                                "fun": lambda f: cap - f.sum()}])

    return res.x, -res.fun

f_cap, g_cap = kelly_capped(2.0)
print(f_cap.round(4), f"{g_cap:.4%}")    # [2. 0.] 38.8741%
nu_cap = (kelly_capped(2.01)[1] - kelly_capped(1.99)[1]) / 0.02
print(f"shadow price of the cap: {nu_cap:.4f}")    # 0.1461
```

The exact solution agrees with Example 8: the cap and the no-short rule bind, and the cap's shadow price is 0.146 against the model's 0.147.

IN PYTHON Companion lab

The Polars-first notebook `LM07_lab.ipynb` builds the total-return series, with all 37 00679B ex-dates and the 0050 split. It reproduces the data exhibits and the examples, and solves the computational practice problems.

The notebook reads FinMind files that you download once with your own token. From the repository root, run `uv run python data/fetch_finmind.py`, then `uv run python data/make_lab_data.py`. FinMind's licence does not allow the book to redistribute the data. After that, the notebook runs offline.

KNOWLEDGE CHECK 7

Why does Newton's method reach the maximum of the model (7.3) in one step from any starting point, but need three steps on the exact growth function?

ANSWER

Each Newton step maximizes the second-order Taylor model at the current point. For (7.3) that model is the function, because (7.3) is quadratic, so the first step lands on $\mathbf{f}^*$. The exact growth function has third- and higher-order terms, so each step lands only near the optimum, with a new error of second order in the old one.

SUMMARY

- Partial derivatives are sensitivities with the other inputs frozen; the total differential (7.1) adds them to first order. For 00679B in 2022 it missed a cross term of -3.24 points.
- The gradient points uphill, perpendicular to level curves; the Hessian measures curvature. The second-order expansion (7.2) explained nearly all of the rise in the 42/58 mix's variance, first-order terms 74%.
- The quadratic growth model (7.3) is accurate to about 1 bp unlevered. At full Kelly it overstates growth by 2.02 points: 0.35 from the dropped squared daily mean, 1.67 from higher-order terms.
- Interior optima have $\nabla h = \mathbf{0}$; the 2×2 test (7.4) classifies them through h_{xx} and $\det H$, and convexity makes a local minimum global.
- At a constrained optimum $\nabla h = \lambda \nabla \psi$ (7.5), and the multiplier is the shadow price $J'(b) = \lambda$ (7.6): a marginal cost, or the best Sharpe ratio for a volatility budget.
- The minimum-variance weight (7.7) needs no expected returns, and $\lambda^* = 2\sigma_{\min}^2$. Fitted on 2017–2021 it predicted 9.22% volatility; 2022–2026 delivered 12.44%, mostly because the inputs changed. By (7.9), the 17-point weight error cost only 0.9 points.
- Kelly fractions (7.12) are the tangency portfolio (7.10) levered by L^* , with $g^* = r_f + \frac{1}{2}\text{SR}^{*2}$ from (7.11). Their estimates swing violently between windows: in-sample optima are not recommendations.
- The KKT conditions (7.13) add non-negative multipliers and complementary slackness. Enumerating binding sets solves small problems, Newton's method and SLSQP larger ones, and bumping a constraint checks its multiplier.

PRACTICE PROBLEMS

1. For $h(x, y) = x^2y + 3xy^2$, the mixed partial derivative h_{xy} at $(1, 2)$ is:
A. 4
B. 14

2. A TWD investor holds a USD bond fund. Over a quarter the fund returns -5% in USD and USD/TWD rises 3% . Compute the TWD return exactly, its linear approximation and the cross term. Then use h_y to estimate the TWD return had USD/TWD risen 4% instead, and compare with the exact value.
3. The two-period geometric mean return is $h(x, y) = \sqrt{(1+x)(1+y)} - 1$. Derive its second-order Taylor expansion about $(0, 0)$, evaluate it at $x = 20\%$, $y = -10\%$, and compare with the exact value. Relate the quadratic term to the volatility drag of LM3.
4. Find and classify all critical points of $h(x, y) = x^3 + y^3 - 3x - 12y + 20$.
5. At the minimum-variance portfolio of two assets the budget multiplier is $\lambda^* = 0.02$. If the investor scales the budget from 1 to 1.05, keeping the proportions, the minimum variance rises by approximately:
 - A. 0.0010
 - B. 0.0020
 - C. 0.0005
6. Asset 1 has volatility 30% and asset 2 has 15% . For which correlation does the unconstrained minimum-variance portfolio short asset 1?
 - A. 0.30
 - B. 0.45
 - C. 0.60
7. Two assets have expected excess returns $e_1 = 6\%$ and $e_2 = 2\%$, volatilities $\sigma_1 = 20\%$ and $\sigma_2 = 8\%$, and correlation $\rho = 0.25$. Compute the tangency weights and the tangency Sharpe ratio, and compare it with each asset's Sharpe ratio.
8. For the assets of Problem 7, compute the Kelly fractions from the growth model (7.3) and the total exposure. Verify that $\mathbf{f}^* = L^* \mathbf{w}^{\text{tan}}$, and compute the model growth rate g^* with $r_f = 2\%$.
9. A broad index fund has $\sigma_1 = 20\%$ and a leveraged fund on a related index has $\sigma_2 = 40\%$, with $\rho = 0.95$ (illustrative). (a) Find the unconstrained minimum-variance weights and volatility. (b) Find the long-only optimum and its KKT multiplier. (c) Interpret the multiplier.
10. Maximize the growth model (7.3) for $e_1 = 10\%$, $e_2 = 4\%$, $\sigma_1 = 20\%$, $\sigma_2 = 10\%$, $\rho = 0$, subject to $f_1 + f_2 \leq 2$, $f_1 \geq 0$, $f_2 \geq 0$. Report the binding constraints and the multipliers, and compare the excess growth with the unconstrained optimum.
11. Which statement about in-sample optimization is correct?

- A. The in-sample minimum-variance weight minimizes future volatility.
- B. Near the minimum, a weight error raises variance in proportion to the square of the error.
- C. Kelly fractions estimated from ten years of daily data are accurate to within ± 0.1 .

12. (Python) With the lab data, estimate the minimum-variance weight (7.7) on each calendar year from 2017 to 2025 and hold that weight, rebalanced daily, through the following year. In how many of the nine following years did it produce lower volatility than a 50/50 mix, also rebalanced daily? Compare the simple averages of the nine annual volatilities with those of the 50/50 mix and of each year's own (ex-post) minimum-variance weight.

SOLUTIONS

1. B is correct. $h_x = 2xy + 3y^2$, so $h_{xy} = 2x + 6y = 2 + 12 = 14$; $h_{yx} = 2x + 6y$ as well. A is $h_{xx} = 2y = 4$, and C is $h_x(1, 2) = 2xy + 3y^2 = 16$, the first derivative rather than the mixed one.
2. Exactly, $0.95 \times 1.03 - 1 = -2.15\%$. The linear approximation is $x + y = -2.00\%$ and the cross term $xy = -0.15\%$. With $h_y = 1 + x = 0.95$, an extra 1 point of currency gain adds 0.95 points, giving $-2.15\% + 0.95\% = -1.20\%$; exactly, $0.95 \times 1.04 - 1 = -1.20\%$. The estimate is exact because h is linear in y for fixed x .
3. At $(0, 0)$: $h = 0$, $h_x = h_y = \frac{1}{2}$, $h_{xx} = h_{yy} = -\frac{1}{4}$ and $h_{xy} = \frac{1}{4}$. By (7.2), $h \approx \frac{1}{2}(x + y) - \frac{1}{8}(x^2 - 2xy + y^2) = \frac{1}{2}(x + y) - \frac{1}{8}(x - y)^2$. At $(0.20, -0.10)$: $0.05 - 0.01125 = 3.875\%$, against the exact $\sqrt{1.08} - 1 = 3.923\%$, an error of 4.8 bp from third- and higher-order terms. The two returns have mean $\frac{1}{2}(x + y)$ and population standard deviation $s = |x - y|/2 = 15\%$, so the quadratic term $-\frac{1}{8}(x - y)^2 = -s^2/2$ is exactly the volatility drag of LM3 (3.9).
4. $h_x = 3x^2 - 3 = 0$ gives $x = \pm 1$; $h_y = 3y^2 - 12 = 0$ gives $y = \pm 2$. With $h_{xx} = 6x$, $h_{yy} = 6y$ and $h_{xy} = 0$, $\det H = 36xy$. At $(1, 2)$: $\det H = 72 > 0$, $h_{xx} = 6 > 0$, a local minimum with $h = 2$. At $(-1, -2)$: $\det H = 72 > 0$, $h_{xx} = -6 < 0$, a local maximum with $h = 38$. At $(1, -2)$ and $(-1, 2)$: $\det H = -72 < 0$, saddles, with $h = 34$ and $h = 6$.
5. A is correct. By (7.6), the minimum variance rises by about $\lambda^* \Delta b = 0.02 \times 0.05 = 0.0010$. Since $\lambda^* = 2\sigma_{\min}^2$, $\sigma_{\min}^2 = 0.01$, and exactly $(1.05^2 - 1) \times 0.01 = 0.001025$; the difference is second order. B takes λ^* for $\sigma_{\min}^2$; C drops the 2 in $\lambda^* = 2\sigma_{\min}^2$.
6. C is correct. The numerator of (7.7) for asset 1 is negative when $\rho > \sigma_2/\sigma_1 = 0.5$. Only 0.60 exceeds the threshold.
7. $\sigma_{12} = 0.25 \times 0.20 \times 0.08 = 0.004$. The numerators of (7.10) are $\sigma_2^2 e_1 - \sigma_{12} e_2 = 0.000384 - 0.00008 = 0.000304$ and $\sigma_1^2 e_2 - \sigma_{12} e_1 = 0.0008 - 0.00024 = 0.00056$, which sum to 0.000864. So $\mathbf{w}^{\tan} = (35.2\%, 64.8\%)$. With $\text{SR}_1 = 0.30$ and $\text{SR}_2 = 0.25$, (7.11) gives $\text{SR}^* = (0.09 + 0.0625 - 0.0375)/0.9375 = 0.1227$ and $\text{SR}^* = 0.350$, above both.
8. $\det \Sigma = 0.04 \times 0.0064 - 0.004^2 = 0.00024$. From (7.12), $f_1^* = 0.000304/0.00024 = 1.267$ and $f_2^* = 0.00056/0.00024 = 2.333$, so $L^* = 3.6$. Then $L^* \mathbf{w}^{\tan} = 3.6 \times (0.3519, 0.6481) = (1.267, 2.333)$, as

required. The model growth is $g^* = 0.02 + \frac{1}{2}(0.1227) = 8.13\%$.

9. (a) $\sigma_{12} = 0.95 \times 0.20 \times 0.40 = 0.076$ and $\sigma_d^2 = 0.04 + 0.16 - 0.152 = 0.048$, so $w_1^* = (0.16 - 0.076)/0.048 = 175\%$ and $w_2^* = -75\%$: short the leveraged fund. $\sigma_{\min}^2 = (0.04 \times 0.16 - 0.076^2)/0.048 = 0.013$, so $\sigma_{\min} = 11.40\%$. (b) Long-only, the constraint $w_1 \leq 1$ binds and the optimum holds only the index fund, $\sigma_p = 20\%$. Here $V'(1) = 2\sigma_1^2 - 2\sigma_{12} = 0.08 - 0.152 = -0.072$, so the multiplier is $\nu = 0.072 \geq 0$. (c) Allowing a 1% short in the leveraged fund ($w_1 = 1.01$) would cut the variance by about 0.00072 (exactly 0.000715). The long-only rule is what stops the investor from hedging the index fund with the leveraged one.
10. Unconstrained, $\mathbf{f}^* = (e_1/\sigma_1^2, e_2/\sigma_2^2) = (2.5, 4)$ when $\rho = 0$, with total exposure 6.5: the cap binds. With only the cap binding, stationarity gives $e_1 - \sigma_1^2 f_1 = e_2 - \sigma_2^2 f_2 = \nu_c$ and $f_1 + f_2 = 2$, so $0.10 - 0.04f_1 = 0.04 - 0.01(2 - f_1)$, $f_1 = 1.6$ and $f_2 = 0.4$. Then $\nu_c = 0.10 - 0.064 = 0.036 \geq 0$, both exposures are positive, and the sign constraints are slack. The excess growth is $f_1 e_1 + f_2 e_2 - \frac{1}{2}(f_1^2 \sigma_1^2 + f_2^2 \sigma_2^2) = 12.40\%$, against 20.50% unconstrained. At the margin the cap is worth 3.6 points per unit; raising it from 2 to 3 adds exactly 3.20 points.
11. **B** is correct, by (7.9): $V(w) - \sigma_{\min}^2 = \sigma_d^2(w - w^*)^2$. **A** is false, as the case shows. **C** is false. With 16% volatility and ten years of data, the mean annual return has a standard error of about $16\%/\sqrt{10} \approx 5$ points. Dividing by $\sigma^2 = 0.0256$ turns that into an error of about ± 2 in a single asset's Kelly fraction.
12. The previous year's weight beat the 50/50 mix in 6 of the 9 years, 2018–2026. Average volatility was 10.84% with the previous year's weight, 11.64% for 50/50 and 10.07% for each year's ex-post best weight, so estimation captured about half of the gain that perfect foresight would have delivered. It lost to 50/50 in 2019, 2023 and, by 0.002 points (14.253% against 14.251%), 2020.

GLOSSARY

Term	中文	Meaning
Covariance	共變異數	Average product of two returns' deviations from their means.
Mixed partial derivative	混合偏導數	h_{xy} ; equal to h_{yx} under continuity.
Gradient	梯度	Vector of partial derivatives; direction of steepest ascent.
Hessian	海森矩陣	Matrix of second partial derivatives; curvature in every direction.
Positive definite	正定	$\Delta^\top H \Delta > 0$ for every $\Delta \neq \mathbf{0}$.
Convex function	凸函數	Positive semidefinite Hessian everywhere; local minima are global.
Lagrangian	拉格朗日函數	$\mathcal{L} = h - \lambda(\psi - b)$; its critical points give constrained optima.
Shadow price	影子價格	Rate at which the optimal value improves as a constraint is relaxed, (7.6).
Envelope theorem	包絡定理	An optimal value's derivative in a parameter equals the Lagrangian's partial derivative there.
KKT conditions	KKT 條件	Stationarity, feasibility, non-negative multipliers, complementary slackness, (7.13).
Binding (active) constraint	緊限制式 (有效限制)	Inequality that holds with equality at the optimum.
Minimum-variance portfolio	最小變異數投資組合	Fully invested mix with the lowest variance, (7.7).
Tangency portfolio	切點投資組合	Fully invested mix with the highest Sharpe ratio, (7.10).
Kelly fractions	凱利比例	Growth-maximizing exposures, (7.12).
In-sample / out-of-sample	樣本內／樣本外	Data used to estimate a rule versus data used to test it.

REFERENCES

- Markowitz, H. (1952). "Portfolio Selection." *Journal of Finance* 7 (1): 77–91.

- Merton, R. C. (1972). "An Analytic Derivation of the Efficient Portfolio Frontier." *Journal of Financial and Quantitative Analysis* 7 (4): 1851–1872.
- Kelly, J. L., Jr. (1956). "A New Interpretation of Information Rate." *Bell System Technical Journal* 35 (4): 917–926.
- Karush, W. (1939). "Minima of Functions of Several Variables with Inequalities as Side Conditions." M.Sc. thesis, Department of Mathematics, University of Chicago.
- Kuhn, H. W., and A. W. Tucker (1951). "Nonlinear Programming." In J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 481–492. Berkeley: University of California Press.
- Kjeldsen, T. H. (2000). "A Contextualized Historical Analysis of the Kuhn–Tucker Theorem in Nonlinear Programming: The Impact of World War II." *Historia Mathematica* 27: 331–361.
- Boyd, S., and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press, chapter 5 (§5.5 optimality conditions).
- Nocedal, J., and S. J. Wright (2006). *Numerical Optimization*, 2nd ed. New York: Springer.
- Kraft, D. (1988). "A Software Package for Sequential Quadratic Programming." Technical Report DFVLR-FB 88-28, DLR German Aerospace Center.
- Chopra, V. K., and W. T. Ziemba (1993). "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." *Journal of Portfolio Management* 19: 6–11.
- Stewart, J., D. K. Clegg, and S. Watson (2021). *Calculus: Early Transcendentals*, 9th ed. Cengage, chapter 14 (partial derivatives, gradients, extreme values and Lagrange multipliers).
- Yuanta Securities Investment Trust. 00679B fund information (tracking index ICE US Treasury 20+ Year Bond Index; listing 17 January 2017) and holdings report, retrieved 10 October 2026: yuantaetfs.com, 00679B.
- 潘姿羽, 中央社 (30 December 2022). "新台幣全年重貶 9.83% 創 25 年最大跌幅，專家點出兩原因." TechNews 科技新報 (USD/TWD 30.708 at end-2022, up 3.018 on the year): finance.technews.tw.
- SciPy 1.16.0 Release Notes, `scipy.optimize` (SLSQP ported to C; constraint multipliers returned): docs.scipy.org.
- FinMind open data: `TaiwanStockPrice`, `TaiwanStockDividendResult` (0050, 00679B), retrieved 10 October 2026.