

Optimization in One Variable

Find the top, prove it is the top, and measure how little its exact location matters.

VOLUME I · Calculus for Finance

BUILDS ON LM2 · LM3

LEADS TO LM6 · LM7 · LM20 · LM41 · LM42

TIME $\approx$ 2.5 h + lab

LEARNING OUTCOMES

Mastery

After this module you should be able to:

☐	1 identify the critical points of a differentiable function and classify them with the first- and second-derivative tests
☐	2 use concavity to show that a point where the derivative vanishes is the global maximum, and prove that the empirical log-growth function of a levered position is strictly concave
☐	3 solve maximization problems on a closed interval, recognising corner solutions created by leverage caps and long-only rules, and interpret the derivative at a corner as a shadow price
☐	4 derive the growth-optimal position of a binary bet, $f^* = p/a - q/b$, and distinguish position size from risk size
☐	5 apply bisection and Newton's method to equations without closed-form solutions, such as a yield to maturity and the empirical Kelly condition, and compare their rates of convergence
☐	6 evaluate how sensitive an optimum and its value are to sizing errors and parameter changes, using the flat-top approximation and the envelope theorem
☐	7 estimate the in-sample growth-optimal leverage of a return series in Python, and explain why it is not a recommendation

LEARNING MODULE OVERVIEW

- At an interior maximum of a differentiable function the derivative is zero; a negative second derivative confirms a strict local maximum. A zero second derivative decides nothing.
- Concavity turns a zero of the derivative into the global maximum. The empirical growth function $\bar{g}(f) = \frac{1}{T} \sum_t \ln(1 + fR_t)$ is strictly concave, so the empirical Kelly leverage exists and is unique whenever a sample has gains and losses.
- On a closed interval the optimum sits at a critical point or at an endpoint. For a concave objective it is the unconstrained optimum clipped to the interval; the derivative at a binding corner is the constraint's shadow price.
- A bet that gains b with probability p and loses a otherwise has the exact Kelly position $f^* = p/a - q/b$. For +2%/-1% at $p = 0.5$ that is 25 times equity, or 25% of equity at risk per trade; keeping 25× while p falls to 0.415 makes long-run growth zero.
- Bisection gains one binary digit per step; Newton's method roughly doubles the number of correct digits. On 0050's Kelly condition, bisection needs 44 steps to shrink the bracket below 10^{-12} ; Newton needs 4. Newton's first step from zero is exactly the quadratic rule $\hat{\mu}/(s^2 + \hat{\mu}^2)$.
- In-sample, constant leverage of 5.42× maximized 0050's growth over 2014–2026: 58.84% a year in log terms against 19.31% unlevered, with a 93.3% drawdown on the way.
- Near the optimum the growth curve is flat: in the quadratic model cf^* keeps $1 - (1 - c)^2$ of the maximal growth rate. Half Kelly keeps 75% (74.2% on 0050) at half the volatility; 0050 loses all growth at 1.80 times Kelly.
- The envelope theorem prices a parameter change without re-optimizing. At the empirical Kelly point each unit of financing rate costs exactly $f^* - 1$ units of growth: 4.42 percentage points a year per point of rate on 0050.

1

INTRODUCTION

Optimization chooses the best value of a decision variable: how many lots to buy, which leverage to hold, what yield prices a bond. In one variable the tools are elementary: a zero derivative at the top, the second derivative to tell tops from bottoms, and the endpoints. A sound answer settles three questions: is the critical point the global top, does a constraint bind, and how much does the exact location matter? This module settles all three for the sizing problem every leveraged trader faces.

MOTIVATING CASE Five point four times 0050

Take 0050's 2,902 daily total returns from 3 November 2014 to 8 October 2026. Suppose you had held a constant leverage f : f units of 0050 for each unit of equity, rebalanced daily and financed at zero cost. Which f makes wealth grow fastest?

Wealth after T days is $\prod_t (1 + fR_t)$, so the question is which f maximizes the sample mean of $\ln(1 + fR_t)$. The answer is **5.42**. At that leverage one unit of equity would have grown to about 1,122, against 10.0 for 0050 itself.

Before acting on it, read the rest of Exhibit 1. The same path fell 93.3% from peak to trough, and on 7 April 2025 alone it lost 54.2%. Half the leverage would have kept about three-quarters of the growth rate.

The textbook rule $f \approx \mu/\sigma^2$ gives 5.68: close, but not equal. And the mean return is estimated with enough error that its 95% confidence band maps into anything from 2.7× to 7.6×.

Which of these numbers, if any, should guide a decision? By Section 7 you will derive each of them. You will see why the optimum is unique, and why its value is robust while its location is not. And you will see why the in-sample 5.42× describes the past rather than recommends a position.

Exhibit 1: Constant-leverage 0050, 3 November 2014 – 8 October 2026

Leverage f	Annual log growth	Growth of 1 unit	Annualized volatility	Maximum drawdown	Worst day (7 Apr 2025)
1 (0050 itself)	19.31%	10.0	19.3%	−33.9%	−10.0%
2 (exact daily 2×)	34.87%	64.2	38.6%	−57.9%	−20.0%
2.71 (half Kelly)	43.63%	182.8	52.3%	−70.1%	−27.1%
5.42 (full Kelly)	58.84%	1,122.5	104.7%	−93.3%	−54.2%
9.74 (zero growth)	0.00%	1.0	188.0%	−99.99%	−97.4%

Note: constant leverage rebalanced daily, with zero financing and trading costs. Mean daily return $\hat{\mu} = 0.08712\%$ and population standard deviation $s = 1.2385\%$, so $\hat{\mu}/s^2 = 5.68$. Annual log growth is the mean of $\ln(1 + fR_t)$ times $A = 243.11$ trading days a year (2,902 returns in 11.937 years); volatility is $f s \sqrt{A}$. Rows 3–5 use the unrounded leverages 2.7107, 5.4215 and 9.7365. Source: FinMind `TaiwanStockPrice` and `TaiwanStockDividendResult`, retrieved 10 October 2026; total returns with reinvested dividends, adjusted for the 4-for-1 split of 18 June 2025, on the trading days common to 0050 and 00631L as in LM3. Computations: `data/lm04_stats.py` and the companion notebook `LM04_lab.ipynb`.

Sections 2–4 build the tools: first- and second-order conditions, concavity and global optima, and constraints that create corner solutions. Section 5 derives the binary Kelly bet exactly. Section 6 solves

equations with no closed form — the empirical Kelly condition and a bond's yield — by bisection and Newton's method. Section 7 measures the sensitivity of an optimum, introduces the envelope theorem and resolves the case; Section 8 does it all in Python.

NOTE Notation in this module

f is the **leverage**, units of the risky asset per unit of equity: a return R becomes fR on equity, with free financing unless stated. $g(f) := \mathbb{E}[\ln(1 + fR)]$ is the expected log growth per period, $\bar{g}(f)$ its sample average, f^* the maximizer, $g^* := g(f^*)$, and $c := f/f^*$ the fraction of full Kelly. For a sample of T returns, $\hat{\mu}$ and s are the sample mean and standard deviation (s divides by T). m_2 , m_3 and m_4 are the means of R^2 , R^3 and R^4 , so $m_2 = \sigma^2 + \mu^2$ in a model and $s^2 + \hat{\mu}^2$ in a sample.

Generic functions are $h(x)$ and $u(x)$ of a real variable x , generic intervals $[\alpha, \beta]$ and domains $\mathcal{D}$. A star marks a solution: x^* is a critical point, a maximizer or a root. A binary bet (Section 5) wins with probability p , $q = 1 - p$, and gains b or loses a per unit staked.

Counts: T returns; N trades with j wins (Section 5). Indices: t for days or payment dates, n for the steps of an iterative method, k for the order of a derivative.

2 FIRST- AND SECOND-ORDER CONDITIONS

- ☐ identify the critical points of a differentiable function and classify them with the first- and second-derivative tests

Optimization starts with a function of one decision variable, such as growth against leverage, and asks where it is largest. Calculus answers in two steps. The first-order condition locates the candidates; the second-order condition sorts them.

Local and global extrema

DEFINITION Local and global maximum

A point x^* of the domain $\mathcal{D} \subseteq \mathbb{R}$ of a function h is a **global maximum** of h on $\mathcal{D}$ if $h(x^*) \geq h(x)$ for every $x \in \mathcal{D}$. It is a **local maximum** if $h(x^*) \geq h(x)$ for every $x \in \mathcal{D}$ in some open interval around x^* . A maximum is **strict** if the inequality is strict for $x \neq x^*$. Minima reverse the inequalities; maxima and minima together are **extrema**.

Every global maximum is a local one, but not conversely. A trader cares about the global maximum, while calculus sees only local information: the slope and curvature at a point. Section 3 shows when local information is enough.

The first-order condition

THEOREM Fermat's theorem (first-order condition)

If h has a local extremum at an interior point x^* of its domain and h is differentiable at x^* , then $h'(x^*) = 0$.

Proof. Let x^* be a local maximum. For small $\Delta > 0$, $h(x^* + \Delta) - h(x^*) \leq 0$, so the right difference quotient $[h(x^* + \Delta) - h(x^*)]/\Delta$ is ≤ 0 and its limit gives $h'(x^*) \leq 0$. For small $\Delta < 0$ the numerator is still ≤ 0 but the denominator is negative, so the left quotient is ≥ 0 and $h'(x^*) \geq 0$. Hence $h'(x^*) = 0$; a minimum reverses the signs. ■

DEFINITION Critical point

An interior point x^* of the domain at which $h'(x^*) = 0$, or at which $h'(x^*)$ does not exist, is a **critical point** of h .

Three warnings follow from the proof. The condition is necessary, not sufficient: $h(x) = x^3$ has $h'(0) = 0$ and no extremum. It applies only at interior points; at an endpoint the one-sided argument yields an inequality instead (Section 4). And a kink, such as $|x|$ at 0, is a candidate even though no derivative exists there.

The second-order condition

THEOREM Second-derivative test

Let $h'(x^*) = 0$ and let h'' be continuous near x^* . If $h''(x^*) < 0$, then x^* is a strict local maximum; if $h''(x^*) > 0$, a strict local minimum. If $h''(x^*) = 0$, the test is inconclusive.

Proof. Taylor's theorem with the Lagrange remainder (LM3, eq. (3.5)) gives $h(x^* + \Delta) = h(x^*) + h'(x^*)\Delta + \frac{1}{2}h''(\xi)\Delta^2$ for some ξ between x^* and $x^* + \Delta$. The linear term vanishes by the first-order condition. If $h''(x^*) < 0$, continuity keeps $h'' < 0$ on a small interval around x^* , so $h(x^* + \Delta) < h(x^*)$ for every small $\Delta \neq 0$. The case $h''(x^*) > 0$ is symmetric. ■

When $h''(x^*) = 0$, look at higher derivatives. If the first non-zero derivative at x^* has order k , a Taylor expansion to order k fixes the sign of $h(x^* + \Delta) - h(x^*)$. For even k the point is an extremum, a maximum if $h^{(k)}(x^*) < 0$; for odd k the graph crosses its tangent, with no extremum. Exhibit 2 collects the cases.

Exhibit 2: Classifying a critical point

Condition at x^*	Conclusion	Example at $x^* = 0$
$h'(x^*) = 0, h''(x^*) < 0$	strict local maximum	$-x^2$
$h'(x^*) = 0, h''(x^*) > 0$	strict local minimum	x^2
$h' = h'' = 0, h''' \neq 0$ at x^*	no extremum (inflection point)	x^3
$h' = h'' = h''' = 0, h^{(4)} < 0$ at x^*	strict local maximum	$-x^4$
$h'(x^*)$ does not exist	compare the signs of h' on each side	$- x $ (maximum)

Note: the first-derivative test — h' changes sign from + to – at x^* — finds a local maximum without h'' and covers the last row.

Growth as a function of leverage

Now the decision this module is about. A position of f units of an asset per unit of equity turns a return R into fR ; one period multiplies wealth by $1 + fR$. Wealth compounds, so the natural objective is the expected log growth per period,

$$g(f) := \mathbb{E}[\ln(1 + fR)]. \quad (4.1)$$

Section 5 shows why maximizing g maximizes long-run wealth; here only its shape matters. Expand $\ln(1 + x) \approx x - x^2/2$ (LM3, eq. (3.3)) with $x = fR$, take expectations, and use $\mathbb{E}[R^2] = \sigma^2 + \mu^2$ (the population form of LM3's identity (3.7)):

$$g(f) \approx f\mu - \frac{f^2}{2}(\sigma^2 + \mu^2). \quad (4.2)$$

The neglected terms are of third order in fR . This quadratic model is a downward parabola. Its first-order condition $\mu - f(\sigma^2 + \mu^2) = 0$ gives

$$f_q = \frac{\mu}{\sigma^2 + \mu^2} \approx \frac{\mu}{\sigma^2}, \quad (4.3)$$

and $g'' = -(\sigma^2 + \mu^2) < 0$ confirms a strict maximum. The second step drops μ^2 , which is smaller than σ^2 by the factor $(\mu/\sigma)^2$; for 0050's daily returns $(\hat{\mu}/s)^2 = 0.0049$. The **Sharpe ratio** $SR := (\mu - r_f)/\sigma$ is the mean return in excess of a risk-free rate r_f , per unit of standard deviation (LM38 studies its estimation). With $r_f = 0$, as here, it is μ/σ per period.

At the optimum $g(f_q) \approx \mu^2/(2(\sigma^2 + \mu^2)) \approx \frac{1}{2}SR^2$: maximal growth is about half the squared Sharpe ratio per period. There the drag term $\frac{1}{2}f_q^2(\sigma^2 + \mu^2)$ is exactly half the levered mean $f_q\mu$, so at the optimum, volatility drag consumes half the levered arithmetic return. On 0050 at the exact optimum, a levered arithmetic mean of 114.83% a year becomes log growth of 58.84%: 51% is kept. LM42 derives the continuous-time versions, $f^* = (\mu - r_f)/\sigma^2$ and $g^* = r_f + SR^2/2$.

EXAMPLE 1

The quadratic rule for 0050 and for a +2% / −1% bet

- (1) Over 2014–2026, 0050's mean daily total return was $\hat{\mu} = 0.08712\%$, with population standard deviation $s = 1.2385\%$. Apply (4.3), and explain why it overshoots the exact optimum of Section 6.
- (2) A trading rule gains 2% or loses 1% of the position with equal probability. Apply (4.3) and compare with the exact optimum of Section 5, 25×.

SOLUTION

(1) The mean square is $m_2 = s^2 + \hat{\mu}^2 = 1.5415 \times 10^{-4}$, so $f_q = 0.0008712/0.00015415 = 5.652$; dropping $\hat{\mu}^2$ gives $\hat{\mu}/s^2 = 5.680$. The predicted maximal growth is $\hat{\mu}^2/(2m_2) = 0.2462\%$ a day, or 59.85% a year with $A = 243.11$. The exact in-sample optimum (Section 6) is 5.42, with 58.84% a year. The rule's leverage is 4% too high, while the growth it predicts is only 1.0 percentage point too high.

The overshoot comes from fat tails. Two more terms of $\ln(1+x)$ add $f^2 m_3 - f^3 m_4$ to the first-order condition. For normal returns $m_3 \approx 3\mu\sigma^2$ and $m_4 \approx 3\sigma^4$, up to terms smaller by the factor $(\mu/\sigma)^2$, so at $f = \mu/\sigma^2$ the two terms cancel.

0050's daily returns have an **excess kurtosis** of 8.12 (LM19 develops this measure of tail weight). The mean of $(R - \hat{\mu})^4$, divided by s^4 , exceeds the normal value 3 by that much. 0050's m_4 is 3.7 times that of normal returns with the same mean and variance, so the quartic term wins and pulls the optimum down. The two $\pm 10\%$ days are not the main cause: without them the rule still overshoots, 5.92 against an exact 5.74, three-quarters of the full-sample gap.

(2) Here $\mu = 0.5\%$ and $\mathbb{E}[R^2] = \frac{1}{2}(0.02^2 + 0.01^2) = 0.00025$, so $f_q = 20$ with predicted growth $\mu^2/(2\mathbb{E}[R^2]) = 5.00\%$ per trade, against the exact 25× and 5.889%. The rule fails because at $f = 20$ the levered outcomes are +40% and −20%. That is far outside the range $|x| \leq 0.1$ where $x - x^2/2$ is accurate (see the pitfall on large returns in LM3).

PITFALL A zero derivative is not a maximum

Setting $h'(x) = 0$ finds every critical point where h is differentiable: maxima, minima and inflection points alike. The tests of Exhibit 2, or concavity on the whole domain (Section 3), tell them apart. Check one of them before calling a critical point optimal.

導讀 先找平的地方，再看往哪邊彎

可微函數在區間內部的極值點，切線一定是水平的（一階條件 $h'(x^*) = 0$ ），但切線水平不保證是山頂：可能是谷底，也可能像 x^3 在 0 那樣只是反曲點。二階條件看彎曲方向， $h'' < 0$ 向下彎就是局部最大。凱利的二次近似 $f\mu - \frac{f^2}{2}(\sigma^2 + \mu^2)$ 是開口向下的拋物線，所以一階條件給出的 $f \approx \mu/\sigma^2$ 就是它的最大值。但它只在 fR 不大時準確：+2%/-1% 的賭局開 20 倍，結果變成 +40%/-20%，二次近似就偏離真正的 25 倍。0050 的二次近似高估約 4%，主因是日報酬的厚尾（超額峰態 8.12）；去掉那兩天 $\pm 10\%$ ，高估仍剩約四分之三。

KNOWLEDGE CHECK 1

Find and classify the critical points of $h(x) = x^3 - 6x^2 + 9x + 1$.

ANSWER

$h'(x) = 3x^2 - 12x + 9 = 3(x-1)(x-3)$, so the critical points are $x = 1$ and $x = 3$. With $h''(x) = 6x - 12$: $h''(1) = -6 < 0$, a strict local maximum with $h(1) = 5$; $h''(3) = 6 > 0$, a strict local minimum with $h(3) = 1$. Neither is global on $\mathbb{R}$, because $h(x) \rightarrow \pm\infty$ as $x \rightarrow \pm\infty$.

3 CONCAVITY AND GLOBAL OPTIMA

- ☐ use concavity to show that a point where the derivative vanishes is the global maximum, and prove that the empirical log-growth function of a levered position is strictly concave

The second-derivative test is local: it compares x^* only with its neighbours. A sizing decision needs a global answer — no leverage anywhere does better. Concavity converts local information into that global guarantee.

DEFINITION Concave and convex functions

A function h is **concave** on an interval I if, for all $x_1, x_2 \in I$ and $w \in [0, 1]$,

$$h(wx_1 + (1-w)x_2) \geq wh(x_1) + (1-w)h(x_2),$$

that is, every chord lies on or below the graph. It is **strictly concave** if the inequality is strict whenever $x_1 \neq x_2$ and $0 < w < 1$. A function h is **convex** (strictly convex) if $-h$ is concave (strictly concave).

For twice-differentiable functions there is a working test, stated here without proof (Simon and Blume 1994, chapter 21). On an open interval, h is concave exactly when $h'' \leq 0$ throughout, and $h'' < 0$ implies strict concavity. The converse of the last statement fails: $-x^4$ is strictly concave although its second derivative vanishes at 0.

THEOREM A critical point of a differentiable concave function is a global maximum

Let h be differentiable and concave on an interval I , and let x^* be an interior point of I with $h'(x^*) = 0$. Then x^* is a global maximum of h on I . If h is strictly concave, x^* is the only maximizer.

Proof (when $h'' \leq 0$ on I). For any $x \in I$, Taylor's theorem (3.5) gives $h(x) = h(x^*) + h'(x^*)(x - x^*) + \frac{1}{2}h''(\xi)(x - x^*)^2 \leq h(x^*)$, because $h'(x^*) = 0$ and $h''(\xi) \leq 0$. If $h'' < 0$ on I , the inequality is strict for $x \neq x^*$. ■ The Deep Dive proves the theorem without second derivatives.

One corollary is used constantly. A differentiable, strictly concave function has at most one critical point, and if it has one, that point is the global maximizer. Finding the optimum then becomes a pure root-finding problem, the subject of Section 6.

DEEP DIVE A concave function lies below its tangent lines optional · [click to expand](#)

Let h be concave and differentiable on I , with $x_1, x \in I$. For $w \in (0, 1]$, concavity with weight w on x gives $h(x_1 + w(x - x_1)) \geq (1 - w)h(x_1) + wh(x)$. Subtract $h(x_1)$ and divide by $w > 0$:

$$\frac{h(x_1 + w(x - x_1)) - h(x_1)}{w} \geq h(x) - h(x_1).$$

As $w \rightarrow 0^+$ the left side tends to $h'(x_1)(x - x_1)$. Hence $h(x) \leq h(x_1) + h'(x_1)(x - x_1)$ for every $x \in I$: the graph lies below each of its tangent lines. If $h'(x_1) = 0$, this reads $h(x) \leq h(x_1)$, which proves the theorem with no assumption on h'' . ■

The log-growth function is strictly concave

Apply this to the sample version of (4.1). For returns $R_1, \dots, R_T$, the **empirical growth function** is

$$\bar{g}(f) := \frac{1}{T} \sum_{t=1}^T \ln(1 + fR_t), \quad (4.4)$$

defined on the set $\mathcal{D}$ of leverages with $1 + fR_t > 0$ for every t . If the sample contains both gains and losses, $\mathcal{D}$ is the open interval $(-1/R_{\max}, 1/|R_{\min}|)$. Differentiating term by term,

$$\bar{g}'(f) = \frac{1}{T} \sum_{t=1}^T \frac{R_t}{1 + fR_t}, \quad (4.5)$$

$$\bar{g}''(f) = -\frac{1}{T} \sum_{t=1}^T \frac{R_t^2}{(1 + f R_t)^2} < 0. \quad (4.6)$$

RESULT Existence and uniqueness of the empirical Kelly leverage

Suppose a sample contains at least one gain and one loss. Then $\bar{g}$ is strictly concave on $\mathcal{D} = (-1/R_{\max}, 1/|R_{\min}|)$ and has exactly one maximizer $f^* \in \mathcal{D}$, the root of $\bar{g}'(f) = 0$.

Uniqueness follows from (4.6), since $\bar{g}'$ is continuous and strictly decreasing on $\mathcal{D}$. For existence, let f rise towards $1/|R_{\min}|$: the worst day's term $R_{\min}/(1 + f R_{\min})$ tends to $-\infty$, so $\bar{g}' \rightarrow -\infty$. As f falls towards $-1/R_{\max}$, the best day's term tends to $+\infty$, and so does $\bar{g}'$.

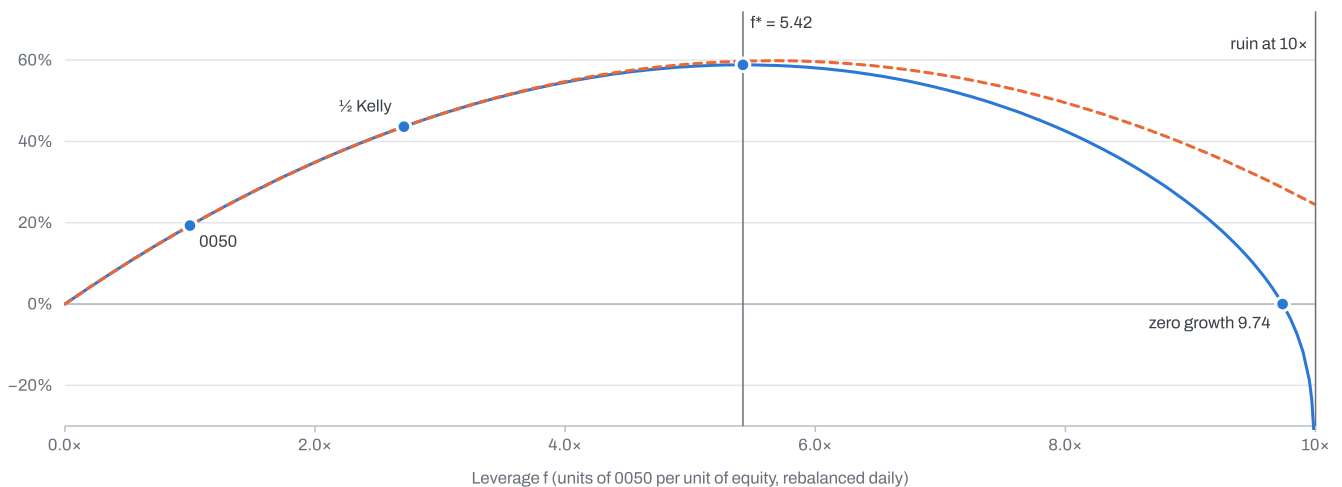
The **intermediate value theorem** says that a continuous function that is positive at one point and negative at another is zero somewhere between them. It is stated here without proof (Stewart, Clegg and Watson 2021, chapter 2). So $\bar{g}'$ crosses zero, and exactly once. The same edges show that $\bar{g}(f) \rightarrow -\infty$: near $1/|R_{\min}|$ the worst day all but wipes out the account.

For 0050 the largest daily loss and gain were both exactly 10%: -10.00% on 7 April 2025 and $+10.00\%$ on 31 July 2026. Ten per cent is Taiwan's daily price limit, in force since 1 June 2015 (it was 7% before). So $\mathcal{D} = (-10, 10)$, and at $10\times$ a single limit-down day ends the experiment. Exhibit 3 plots $\bar{g}$ for 0050 against the quadratic model.

Exhibit 3: Annual log growth of constant-leverage 0050, November 2014 – October 2026

— Exact: mean of $\ln(1 + fR)$ - - - Quadratic model: $f \times \text{sample mean} - f^2 m_2/2$

Annual log growth rate



Note: the exact curve is $A \bar{g}(f)$ from (4.4), with $A = 243.11$; the dashed curve is $A(f\hat{\mu} - \frac{1}{2}f^2 m_2)$ from (4.2). The exact curve peaks at $f^* = 5.42$ (58.84% a year), returns to zero at 9.74 and falls to $-\infty$ at 10. The parabola peaks at 5.65 and would return to zero at 11.30, beyond the ruin point.

EXAMPLE 2

Where the parabola misleads

A familiar rule says growth returns to zero at about twice the Kelly leverage. Check it on 0050, and explain the gap using the worst day in the sample.

SOLUTION

The quadratic model's zero is $2f_q = 2\hat{\mu}/m_2 = 11.30$, outside $\mathcal{D}$. The exact zero-growth leverage, the root of $\bar{g}(f) = 0$ above f^* , is 9.74, or 1.80 times f^* . It has no closed form; Section 6 shows how to find it.

The largest daily losses explain the gap. At $f = 9.737$, 7 April 2025 alone multiplies equity by $1 - 0.9737 = 0.0263$, a log return of -3.64 . That single day cancels the gains of the other 2,901 days.

At f^* the same day costs -0.78 log points out of a total of 7.02. The logarithm punishes a large loss far more than it rewards an equal gain: $\ln(1 - 0.54) = -0.78$, but $\ln(1 + 0.54) = +0.43$. Fat tails therefore pull the growth curve down faster than any parabola as leverage approaches the ruin point.

PITFALL Local optima of a non-concave objective

"Local equals global" is a property of the objective, not of the search. A backtest's Sharpe ratio is usually bumpy in its parameters; a grid search returns the highest bump, which is largely noise (LM39). Kelly sizing is well posed because the logarithm of an affine function of f is concave. Treat the optimum of a bumpy objective as a hypothesis.

導讀 凹函數：局部最大就是全域最大

凹函數（concave，圖形向下彎、弦在圖形下方）有個關鍵性質：可微的凹函數只要找到切線水平的點，它就是全域最大；嚴格凹時還是唯一的。平均對數成長 $\bar{g}(f)$ 是許多 $\ln(1 + fR_t)$ 的平均，整體嚴格凹，所以只要樣本有漲有跌，凱利最適槓桿一定存在，而且只有一個。0050 單日最大漲跌都是 10%（台股漲跌幅限制），槓桿只能落在 -10 到 10 倍之間；越靠近 10 倍，一根跌停就讓帳戶歸零，曲線掉得比拋物線快得多。用語提醒：本書 concave = 凹函數、convex = 凸函數；有些微積分課本說的「凹口向上」其實是 convex。

KNOWLEDGE CHECK 2

Show that the growth function of a binary bet, $h(f) = p \ln(1 + fb) + q \ln(1 - fa)$ with $0 < p < 1$, $q = 1 - p$ and $a, b > 0$, is strictly concave on $(-1/b, 1/a)$.

ANSWER

$h'(f) = \frac{pb}{1+fb} - \frac{qa}{1-fa}$ and $h''(f) = -\frac{pb^2}{(1+fb)^2} - \frac{qa^2}{(1-fa)^2}$. Both terms are negative on the interval, so $h'' < 0$ and h is strictly concave. Any critical point is therefore the unique global maximum, which Section 5 computes.

4

CONSTRAINTS AND CORNER SOLUTIONS

- ☐ solve maximization problems on a closed interval, recognising corner solutions created by leverage caps and long-only rules, and interpret the derivative at a corner as a shadow price

Real positions face limits. A cash account cannot borrow, many mandates cannot sell short, and brokers cap margin. Each limit confines f to an interval, and the optimum may then sit on the boundary, where the derivative need not be zero.

THEOREM Extreme value theorem

A continuous function on a closed, bounded interval $[\alpha, \beta]$ attains a maximum and a minimum on it.

The theorem is stated without proof (Stewart, Clegg and Watson 2021, chapter 4). It turns optimization on $[\alpha, \beta]$ into a finite search, the **closed-interval method**. Evaluate h at its critical points in (α, β) and at both endpoints, and take the largest value; the method needs continuity only, not concavity.

First-order conditions at a corner

At an endpoint the argument of Fermat's theorem runs on one side only, so it delivers an inequality. If a differentiable h attains its maximum on $[\alpha, \beta]$ at β , every left difference quotient $[h(\beta - \Delta) - h(\beta)]/(-\Delta)$ is ≥ 0 , so $h'(\beta) \geq 0$. Likewise a maximum at α requires $h'(\alpha) \leq 0$. Collecting the three cases,

$$\text{maximum at } x^* \in [\alpha, \beta] \implies \begin{cases} h'(\alpha) \leq 0 & \text{if } x^* = \alpha, \\ h'(x^*) = 0 & \text{if } \alpha < x^* < \beta, \\ h'(\beta) \geq 0 & \text{if } x^* = \beta. \end{cases}$$

These are the one-variable **Karush–Kuhn–Tucker (KKT) conditions**, which LM7 extends with Lagrange multipliers. For a concave h they are also sufficient, by the tangent-line inequality of the Deep Dive in

Section 3.

The derivative at a binding corner has an economic reading. If the optimum is β and $h'(\beta) > 0$, relaxing the limit to $\beta + \Delta$ raises the objective by about $h'(\beta) \Delta$ when Δ is small. That derivative is the **shadow price** of the constraint: the rate at which relaxing it raises the objective, a marginal value rather than a total. In LM7 it becomes the Lagrange multiplier.

PROPOSITION Clipping

Let h be differentiable and concave on an open interval containing $[\alpha, \beta]$, with an unconstrained maximizer x^* . Then the clipped point $\min(\max(x^*, \alpha), \beta)$ maximizes h on $[\alpha, \beta]$; if h is strictly concave, it is the only maximizer.

Proof. If $x^* \in [\alpha, \beta]$ there is nothing to show. If $x^* > \beta$, then h' is non-increasing with $h'(x^*) = 0$, so $h' \geq 0$ on $[\alpha, \beta]$ and h is non-decreasing there. Hence β is a maximizer, and the only one if h is strictly concave, since then $h' > 0$ on $[\alpha, \beta]$. The case $x^* < \alpha$ is symmetric. ■

Two corners in the 0050 data

A long-only investor without a margin account can still reach any leverage between 0 and 2. Mixing cash, 0050 and the daily 2× fund 00631L (LM3) does it, if rebalanced daily and ignoring costs and replication basis. The feasible set is $[0, 2]$. Exhibit 4 shows what in-sample optimization recommends year by year, with and without that interval.

Exhibit 4: Empirical Kelly leverage of 0050 in the full calendar years 2015–2025, and the optimum on $[0, 2]$

Year	0050 total return	$\hat{\mu}$ (bp a day)	s (% a day)	Unconstrained f^*	Optimum on $[0, 2]$
2015	−6.31%	−2.13	1.044	−1.93	0
2016	+19.64%	+7.81	0.956	+8.37	2
2017	+18.13%	+6.95	0.592	+19.41	2
2018	−4.95%	−1.50	1.046	−1.40	0
2019	+33.52%	+12.26	0.782	+19.73	2
2020	+31.08%	+12.12	1.462	+5.44	2
2021	+21.97%	+8.77	1.126	+6.97	2
2022	−21.34%	−8.79	1.389	−4.46	0
2023	+27.40%	+10.53	0.894	+13.58	2
2024	+48.67%	+17.66	1.580	+5.95	2
2025	+36.85%	+14.43	1.575	+5.11	2
Nov 2014 – Oct 2026	+902.25%	+8.71	1.239	+5.42	2

Note: the rows are the full calendar years 2015–2025. The partial years 2014 (from 3 November: +2.37%, $f^* = 7.97$) and 2026 (to 8 October: +78.71%, $f^* = 8.80$) enter only the full-sample row. The rows compound to +447.8%, not +902.25%. f^* maximizes the period's mean of $\ln(1 + fR_t)$ over its own domain $(-1/R_{\max}, 1/|R_{\min}|)$. The optimum on $[0, 2]$ is f^* clipped to the interval. Source: as Exhibit 1.

Two facts stand out. Year by year the unconstrained optimum ranges from a $4.46\times$ short position to $19.73\times$ long. And every constrained optimum is a corner: 0 in the three years with $\hat{\mu} < 0$, 2 in the eight with $f^* > 2$.

EXAMPLE 3

Two corners: the $2\times$ ceiling and the long-only floor

(1) For the full sample, find the optimal leverage on $[0, 2]$, the growth it delivers and the shadow price of the ceiling. Estimate the gain from raising the ceiling to 2.5. (2) For 2022 alone, show that the long-only floor binds.

SOLUTION

(1) $\bar{g}$ is concave and $f^* = 5.42 > 2$, so by clipping the optimum is the corner $f = 2$: everything in the 2× fund. Growth is 34.87% a year (×64.2 over the sample), against 58.84% at f^* . The corner condition holds: $A \bar{g}'(2) = 13.68\%$ a year per unit of leverage, a positive number. That is the shadow price.

Raising the ceiling to 2.5 should add about $0.5 \times 13.68\% = 6.84\%$ a year; the exact gain is 6.37%, smaller because $\bar{g}'$ falls as f rises. In practice 00631L grew ×44.5, not ×64.2, because of the costs and replication basis analysed in LM3.

(2) In 2022, $\hat{\mu} = -8.79$ bp a day and the unconstrained optimum is $f^* = -4.46$, a 4.46× short position. On $[0, 2]$ the optimum is the corner $f = 0$, and the corner condition $\bar{g}'(0) = \hat{\mu} < 0$ holds. In hindsight, cash beat every long position that year.

PITFALL Forcing the derivative to zero at a boundary

At a binding constraint the derivative is not zero, and solving $h'(x) = 0$ returns an infeasible point: 5.42 when the ceiling is 2. Solve the unconstrained problem, then clip (for a concave objective) or compare endpoints (in general). Exhibit 4 also shows a benefit of corners: once a ceiling binds, the decision is the same for every estimate of f^* above it. Corners make decisions robust to estimation error, a point Section 7 returns to.

導讀 角解與影子價格

現實中的槓桿有上下限：現金帳戶只用 0050 加 00631L，最多約 2 倍；不能放空，下限就是 0。閉區間上的最大值可能落在端點（角解），這時導數不一定為零。在上限 2 倍處 $\bar{g}'(2) > 0$ ，代表「如果能再多借一點，成長率還會增加」；這個導數就是限制的影子價格，是放寬一點點時的邊際增加速度，不是總增加量，也是 LM7 拉格朗日乘數的一維版本。目標函數是凹函數時，把無限制的最適解截到區間內就是答案。2015–2025 每個完整年度的答案都是角解：平均日報酬為負的三年落在 0，其餘八年的 f^* 都超過 2，落在上限 2。

KNOWLEDGE CHECK 3

Find the global maximum and minimum of $h(x) = x^3 - 3x$ on $[-2, 3]$.

ANSWER

$h'(x) = 3x^2 - 3 = 0$ at $x = \pm 1$, and the candidates give $h(-2) = -2$, $h(-1) = 2$, $h(1) = -2$ and $h(3) = 18$. The global maximum is the corner $x = 3$, even though $x = -1$ is a strict local maximum. The global minimum, -2 , is attained twice: at the corner $x = -2$ and at the interior point $x = 1$. The function is not concave, so clipping does not apply and every candidate must be compared.

5 THE BINARY KELLY BET

- ☐ derive the growth-optimal position of a binary bet, $f^* = p/a - q/b$, and distinguish position size from risk size

The simplest sizing problem has one source of risk, two outcomes and an exact solution. It exposes the confusion between position size and risk size, and shows how sharply the optimum depends on the win probability. [LM41](#) extends it to general discrete bets.

Why maximize expected log growth

A trade gains a fraction b of the position with probability p and loses a fraction a with probability $q = 1 - p$, independently of other trades. With a position of f times equity, one trade multiplies wealth by $1 + fb$ or $1 - fa$. After N trades with j wins,

$$\frac{W_N}{W_0} = (1 + fb)^j (1 - fa)^{N-j}, \quad \frac{1}{N} \ln \frac{W_N}{W_0} = \frac{j}{N} \ln(1 + fb) + \left(1 - \frac{j}{N}\right) \ln(1 - fa).$$

By the law of large numbers ([LM17](#)), $j/N \rightarrow p$, so the per-trade log growth converges to

$$g(f) = p \ln(1 + fb) + q \ln(1 - fa), \quad -\frac{1}{b} < f < \frac{1}{a}. \quad (4.7)$$

If $g(f) > g(\tilde{f})$, the ratio of the two wealth paths behaves like $e^{N[g(f) - g(\tilde{f})]}$ and grows without bound. The fraction with the highest g therefore ends up with more wealth than any other fixed fraction, with probability one (Breiman 1961; Thorp 2006). That is Kelly's (1956) criterion: maximize the expected log growth. MacLean, Thorp and Ziemba (2011) collect the classic papers on it.

Deriving the Kelly position

From (4.7),

$$g'(f) = \frac{pb}{1+fb} - \frac{qa}{1-fa}, \quad g''(f) = -\frac{pb^2}{(1+fb)^2} - \frac{qa^2}{(1-fa)^2} < 0.$$

The growth function is strictly concave (the Knowledge Check of Section 3), so the root of g' is the unique global maximum. Setting $g'(f) = 0$ and cross-multiplying gives $pb(1-fa) = qa(1+fb)$, so $pb - qa = fab(p+q) = fab$, and

$$f^* = \frac{pb - qa}{ab} = \frac{p}{a} - \frac{q}{b}. \quad (4.8)$$

No approximation was made: (4.8) is exact. Substituting it back gives $1 + f^*b = p(a+b)/a$ and $1 - f^*a = q(a+b)/b$, so the maximal growth rate is

$$g^* = p \ln \frac{p(a+b)}{a} + q \ln \frac{q(a+b)}{b}. \quad (4.9)$$

A rewriting makes (4.8) easier to read. Let $\pi := a/(a+b)$ be the **break-even probability**, at which the expected return $pb - qa$ is zero. Then

$$f^* = \left(\frac{1}{a} + \frac{1}{b} \right) (p - \pi). \quad (4.10)$$

The Kelly position is proportional to the probability edge $p - \pi$, with slope $1/a + 1/b$. It is positive exactly when the expected return is positive. With no edge, $f^* = 0$ is an interior optimum with $g'(0) = 0$; with a negative edge, a long-only rule gives the corner solution $f = 0$.

RESULT The binary Kelly bet

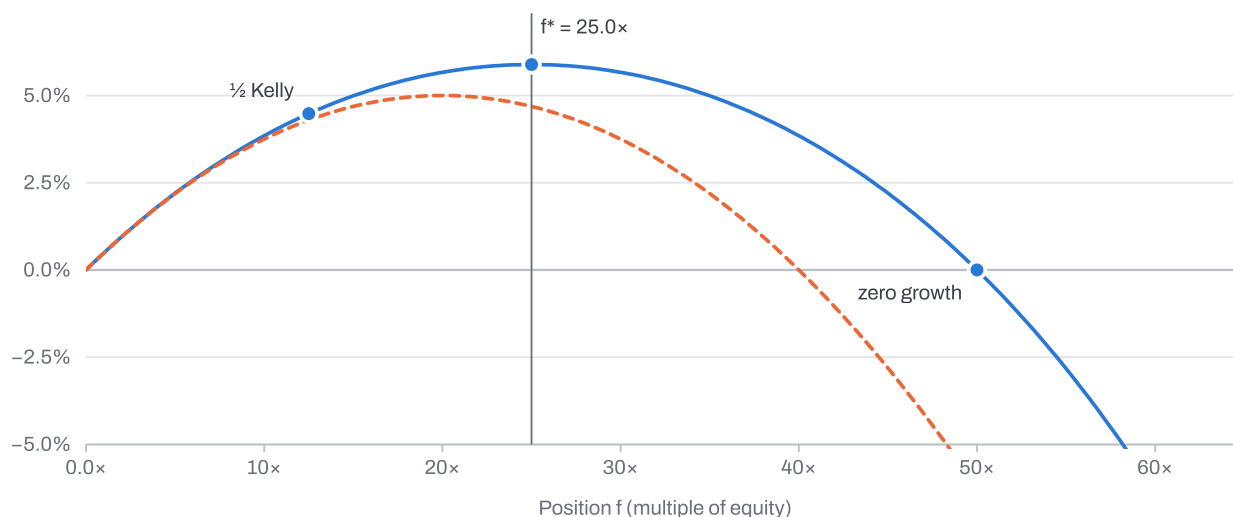
For a bet that gains b with probability p and loses a with probability $q = 1 - p$, the growth-maximizing position is exactly $f^* = p/a - q/b = (1/a + 1/b)(p - \pi)$ times equity, where $\pi = a/(a+b)$. The growth function is strictly concave, so f^* is unique, and the maximal growth (4.9) is positive whenever $p > \pi$.

Binary Kelly bet: growth per trade against position size

INTERACTIVE

— Exact $g(f)$ - - - Quadratic model

Log growth per trade



Bet: p, b, a

0.500, +2.0%, -1.0%

Kelly position f^*

25.0x equity

Risk per losing trade f^*a

25.0% of equity

Growth per trade at f^*

5.89%

Half Kelly keeps

76.1% of it

Zero growth at

50.0x (2.00 f^*)

Break-even probability

0.333

One loss is ruin at $f = 1/a$

100.0x

Solid: exact growth per trade $g(f) = p \ln(1 + fb) + q \ln(1 - fa)$ for a position of f times equity. Dashed: the quadratic model $f\mu - f^2E[R^2]/2$. Dots: half Kelly, full Kelly $f^* = p/a - q/b$ and the zero-growth point. One loss at $f = 1/a$ wipes out the account.

Position size versus risk size

On a losing trade the account loses fa of equity, the **risk fraction**. Multiplying (4.8) by a ,

$$f^*a = p - q \frac{a}{b}. \quad (4.11)$$

When the whole stake can be lost, $a = 1$ and position equals risk. Then (4.11) is the classic Kelly formula $p - q/b$ for a bet paying b to 1, which Thorp (2006) writes as $(bp - q)/b$. A stop-loss trade loses only a of the position, so the two numbers differ by the factor $1/a$: 100 for a 1% stop.

Equation (4.11) carries a second message. Scaling a and b by the same factor leaves the payoff ratio b/a , f^*a , $1 + f^*b$ and $1 - f^*a$ unchanged, so by (4.9) it leaves g^* unchanged. Only the payoff ratio and the win probability determine the attainable growth. The stop distance sets the position size that delivers it.

EXAMPLE 4

The +2% / -1% bet

A trade gains 2% or loses 1% of the position with equal probability. Find the Kelly position, the risk per trade and the maximal growth. Then evaluate a trader who reads "Kelly = 25%" as "put 25% of capital into the trade".

SOLUTION

By (4.8), $f^* = 0.5/0.01 - 0.5/0.02 = 50 - 25 = 25$: a position of 25 times equity. A loss costs $25 \times 1\% = 25\%$ of equity, the risk fraction of (4.11): $f^*a = 0.5 - 0.5 \times 0.01/0.02 = 0.25$. Growth at the optimum is $g^* = 0.5 \ln 1.5 + 0.5 \ln 0.75 = 5.889\%$ per trade, from a levered arithmetic mean of $25 \times 0.5\% = 12.5\%$.

The trader who holds 0.25 times equity risks only 0.25% per trade. Growth is $0.5 \ln 1.005 + 0.5 \ln 0.9975 = 0.1242\%$ per trade, 2.1% of the maximum. "25× equity" and "25% at risk" are the same decision in two units; mixing the units costs a factor of 100.

The same curve has two more landmarks. Its domain is $(-50, 100)$, so one loss at 100× ends the account. And growth returns to zero at exactly $50 = 2f^*$, because $(1 + 50 \times 0.02)(1 - 50 \times 0.01) = 2 \times 0.5 = 1$.

Sensitivity to the win probability

Equation (4.10) makes the sensitivity explicit: $df^*/dp = 1/a + 1/b$, which is 150 for the +2%/−1% bet. One percentage point of win probability moves the Kelly position by 1.5 times equity. Exhibit 5 follows the bet as its true win probability falls, and asks what happens to a trader who keeps the 25× position.

Exhibit 5: The +2% / −1% bet as the win probability falls

Win probability p	Kelly position f^*	Risk f^*a	Maximal growth g^* per trade	Growth at a fixed 25×
0.60	40.0	40.0%	14.834%	+12.821%
0.55	32.5	32.5%	9.856%	+9.355%
0.50	25.0	25.0%	5.889%	+5.889%
0.45	17.5	17.5%	2.924%	+2.423%
0.41504	12.3	12.3%	1.451%	0.000%
0.40	10.0	10.0%	0.971%	−1.042%
1/3 (break-even)	0	0	0	−5.663%

Note: $f^* = 150p - 50$ by (4.10). A fixed 25× position has zero growth at $p = \ln(1/0.75) / \ln(1.5/0.75) = 0.41504$, where $f^* = 12.26$: the fixed position has become 2.04 times Kelly.

Growth at a fixed 25× stays positive until p falls to 0.415. There the true Kelly position is 12.26, so the trader holds about twice Kelly, the zero-growth point of Section 7. Below it the trader loses money in the long run even though every trade still has a positive expected return. At $p = 0.40$ the expected

return is $0.40 \times 2\% - 0.60 \times 1\% = +0.2\%$ per unit of position, yet the 25× position shrinks wealth by 1.04% per trade.

DEEP DIVE Kelly's information interpretation: maximal growth is a relative entropy

optional · click to expand

Write (4.9) with $\pi = a/(a+b)$ and $1-\pi = b/(a+b)$:

$$g^* = p \ln \frac{p}{\pi} + q \ln \frac{q}{1-\pi} =: D_{\text{KL}}(p \parallel \pi).$$

D_{KL} is the **relative entropy**, or Kullback–Leibler divergence, between the true probability p and the break-even probability π . Using $\ln x \leq x - 1$ (Practice Problem 5 of LM3),

$$-D_{\text{KL}}(p \parallel \pi) = p \ln \frac{\pi}{p} + q \ln \frac{1-\pi}{q} \leq p \left(\frac{\pi}{p} - 1 \right) + q \left(\frac{1-\pi}{q} - 1 \right) = 0,$$

with equality only if $p = \pi$.

Allowing $f < 0$ (betting against the trade), maximal growth is positive whenever the true probability differs from the break-even one. It measures that difference in information units. This is the "new interpretation of information rate" of Kelly (1956); Cover and Thomas (2006, chapter 6) develop it for horse races. ■

PITFALL "Kelly says 25%" without a unit

The same decision can be stated as a 25× position or as 25% of equity at risk. Read as "25% of capital in the trade", it becomes a position one hundred times too small. Read as "25× leverage" on a trade with a 10% stop, it puts 250% of equity at risk: one loss is ruin.

State a Kelly number with its unit: a multiple of equity, or the fraction of equity lost if the trade fails. Recompute the position whenever the stop distance changes.

導讀 部位大小 vs 風險大小

凱利公式 $f^* = p/a - q/b$ 算的是「部位是權益的幾倍」，風險公式 $f^*a = p - qa/b$ 算的是「一筆交易賠掉時損失權益的比例」（整筆可能賠光、 $a = 1$ 時就是經典的 $p - q/b$ ），兩者差一個 $1/a$ 。+2%/-1%、勝率五成：部位 25 倍、每筆風險 25%，說的是同一件事；把「25%」誤讀成「拿 25% 資金下單」，部位只剩凱利的百分之一，成長率只剩最大值的約 2%。另一個重點是對勝率極度敏感： $f^* = (1/a + 1/b)(p - \pi)$ ，勝率每掉 1 個百分點，部位就少 1.5 倍權益。固定開 25 倍而勝率掉到 0.415 時，每筆交易的期望報酬仍是正的，長期複利成長率卻已歸零。

KNOWLEDGE CHECK 4

A trade gains 3% with probability 0.45 and otherwise loses 1.5%. Find the Kelly position and the risk per losing trade, and verify the risk with (4.11).

ANSWER

$f^* = 0.45/0.015 - 0.55/0.03 = 30 - 18.33 = 11.67$ times equity. A loss costs $11.67 \times 1.5\% = 17.5\%$ of equity. By (4.11), $p - q a/b = 0.45 - 0.55 \times 0.5 = 0.175$: the same 17.5%.

6

ROOT FINDING: BISECTION AND NEWTON'S METHOD

- ☐ apply bisection and Newton's method to equations without closed-form solutions, such as a yield to maturity and the empirical Kelly condition, and compare their rates of convergence

The binary bet is the exception: most first-order conditions have no closed form. The empirical Kelly condition $\bar{g}'(f) = 0$ is the sample version of $\mathbb{E}[R/(1 + fR)] = 0$. Once denominators are cleared, for 0050 it becomes a polynomial equation of degree 2,709, one less than the number of distinct non-zero daily returns.

A bond's yield and a project's internal rate of return are roots of polynomials too. Numerical root finding solves $u(x) = 0$ to any accuracy, and for a smooth concave objective h , maximizing h is root finding on $u = h'$.

Bisection

DEFINITION Bisection method

Let u be continuous on $[x_{lo}, x_{hi}]$, with $u(x_{lo})$ and $u(x_{hi})$ of opposite signs. Evaluate u at the midpoint $x_{mid} = (x_{lo} + x_{hi})/2$ and keep the half-interval whose endpoints still have opposite signs. Repeat.

By the intermediate value theorem (Section 3) every interval in the sequence contains a root. After n steps the bracket has width $(x_{hi} - x_{lo})/2^n$, so its midpoint is within $(x_{hi} - x_{lo})/2^{n+1}$ of a root. A bracket of width ε takes

$$n \geq \log_2 \frac{x_{hi} - x_{lo}}{\varepsilon}$$

steps: one binary digit per step, or about 3.3 steps per decimal digit. The error bound shrinks by a constant factor each step, which is called **linear convergence**. Bisection needs only continuity and a sign change, and it cannot fail.

Newton's method

Newton's method replaces u by its tangent at the current point, the first-order Taylor polynomial of LM3, and jumps to the tangent's root:

$$x_{n+1} = x_n - \frac{u(x_n)}{u'(x_n)}. \quad (4.12)$$

Applied to $u = h'$ for a maximization, the step is $x_{n+1} = x_n - h'(x_n)/h''(x_n)$: each iteration maximizes the second-order Taylor model of h at x_n .

The speed comes from the remainder. Let x^* be a root with $u'(x^*) \neq 0$ and write $e_n = x_n - x^*$. Taylor's theorem gives $0 = u(x^*) = u(x_n) + u'(x_n)(x^* - x_n) + \frac{1}{2}u''(\xi_n)(x^* - x_n)^2$ for some ξ_n between x_n and x^* . Dividing by $u'(x_n)$ and using (4.12),

$$e_{n+1} = \frac{u''(\xi_n)}{2u'(x_n)} e_n^2. \quad (4.13)$$

The error is squared at each step: **quadratic convergence**. Once $|e_n|$ is small, the number of correct digits roughly doubles per iteration. The constant $C = |u''(x^*)/(2u'(x^*))|$ says how small is small: squaring helps once $C|e_n| < 1$.

THEOREM Local quadratic convergence of Newton's method

Let u be twice continuously differentiable near a root x^* with $u'(x^*) \neq 0$. Then there is an interval around x^* such that Newton's method started anywhere in it converges to x^* . Moreover, $|e_{n+1}| \leq K|e_n|^2$ for some constant K .

The theorem is stated without its full proof; (4.13) is the core, and Burden, Faires and Burden (2016, chapter 2) supply the details. It is a local result. Far from the root the tangent can point anywhere. Where $u'(x_n) \approx 0$ the step is huge, and a step can leave the domain or land near a different root.

Safeguarded methods combine the two ideas. They keep a bracket, accept a fast step (Newton or secant) when it lands inside it, and bisection otherwise. Brent's (1973) method, available as `scipy.optimize.brentq`, is the standard version: it never loses the bracket and converges superlinearly near the root.

The empirical Kelly condition

Newton's method on $u = \bar{g}'$ uses (4.5) and (4.6). Start from $f_0 = 0$, where $\bar{g}'(0) = \hat{\mu}$ and $\bar{g}''(0) = -m_2$:

$$f_1 = 0 - \frac{\hat{\mu}}{-m_2} = \frac{\hat{\mu}}{m_2} = \frac{\hat{\mu}}{s^2 + \hat{\mu}^2}.$$

The quadratic rule (4.3) is exactly one Newton step from zero. Each later step refits the quadratic model at the current point. Exhibit 6 runs both methods on 0050. Bisection starts from the bracket $[0, 9.9]$, since $\bar{g}'(0) = \hat{\mu} > 0$ and $\bar{g}'(9.9) < 0$: $\bar{g}'$ falls towards $-\infty$ as f approaches 10, where it is undefined.

Exhibit 6: Newton's method and bisection on 0050's empirical Kelly condition

Step n	Newton f_n	Newton error $ f_n - f^* $	Bisection bracket width	Bisection midpoint error
0	0	5.4	9.9	0.47
1	5.651642810	2.3×10^{-1}	4.95	2.0
2	5.423914969	2.4×10^{-3}	2.475	0.77
3	5.421488036	2.5×10^{-7}	1.2375	0.15
4	5.421487783	2.8×10^{-15}	0.61875	0.16
10			9.7×10^{-3}	2.6×10^{-3}
20			9.4×10^{-6}	3.6×10^{-7}
30			9.2×10^{-9}	4.2×10^{-9}
44			5.6×10^{-13}	4.8×10^{-14}

Note: $f^* = 5.421487782858$. Errors use 50-digit arithmetic on the exact returns, because the last ones are at the level of double-precision rounding. Newton's unrounded errors give $e_2/e_1^2 = 0.046$ and $e_3/e_2^2 = 0.043$, close to $C = |\bar{g}'''(f^*)/(2\bar{g}''(f^*))| = 0.043$ from (4.13). Bisection's bracket halves every step, while its midpoint error need not fall every step; 44 steps bring the width below 10^{-12} .

EXAMPLE 5

Yield to maturity by Newton's method and by bisection

A five-year bond pays an annual coupon of 1.5 per 100 face value and trades at 97.80 (illustrative numbers). Its yield y solves $P(y) = 97.80$, where

$$P(y) = \sum_{t=1}^5 \frac{1.5}{(1+y)^t} + \frac{100}{(1+y)^5};$$

LM6 derives the pricing formula. Find y by Newton's method from $y_0 = 1.5\%$, and count the bisection steps needed to shrink the bracket $[0, 0.10]$ to a width of 10^{-10} .

SOLUTION

Let $u(y) = P(y) - 97.80$, with $u'(y) = -\sum_t t \times 1.5/(1+y)^{t+1} - 500/(1+y)^6$. At $y_0 = 1.5\%$ the bond would trade at par, so $u(y_0) = 2.200$, and $u'(y_0) = -478.26$; hence $y_1 = 1.5\% + 2.200/478.26 = 1.96000\%$. Next, $u(y_1) = 0.0293$ and $u'(y_1) = -465.59$ give $y_2 = 1.96629\%$, and $y_3 = 1.9662919\%$ prices the bond to within 2×10^{-13} . The yield is 1.9663%.

Bisection starts from $u(0) = 107.50 - 97.80 > 0$ and $u(0.10) = 67.78 - 97.80 < 0$. It needs $\lceil \log_2(0.10/10^{-10}) \rceil = \lceil 29.9 \rceil = 30$ steps against Newton's three. Newton is faster here because P is smooth and gently curved and the coupon rate is a good starting point. Bisection's advantage is that it needs none of these.

PITFALL Newton's method without a bracket

Newton's method trusts the tangent. Cash flows of -100 , $+230$ and -132 (a project with a closing cost) have two internal rates of return, 10% and 20%. Their NPV is flat at 14.8%. Newton started at 14% first jumps to -1.20% and ends at 10%; started at 15%, it jumps to 72.50% and ends at 20%.

On 0050's Kelly condition Newton converged from all 199 starting points from -9.9 to 9.9 (step 0.1), in at most eight steps. That is a property of the function, not of the method: use a bracketing method (`brentq`) whenever a sign change is known.

導讀 二分法與牛頓法

二分法每一步把區間砍半，穩但慢：每多一位小數大約要 3.3 步，只要函數連續、端點異號就一定收斂。牛頓法用切線逼近，誤差每一步大約平方一次，正確位數接近翻倍；但它需要好的起點，切線太平時會跳得很遠，可能跳出定義域或跑到另一個根。實務上兩者混合 (scipy 的 `brentq`)：用區間保證不跑掉，再用快速的步驟收斂。最妙的是：從 $f = 0$ 出發的第一步牛頓法，剛好就是二次近似 $\hat{\mu}/(s^2 + \hat{\mu}^2)$ 。在 0050 上牛頓法四步就到機器精度，二分法則要 44 步才把區間縮到 10^{-12} 以下。

KNOWLEDGE CHECK 5

A yield is known to lie between 0% and 10%. How many bisection steps reduce the bracket to a width of 10^{-6} (0.01 bp)?

ANSWER

The width after n steps is $0.10/2^n$, so we need $2^n \geq 0.10/10^{-6} = 10^5$, i.e. $n \geq \log_2 10^5 = 16.6$. Seventeen steps suffice.

7

SENSITIVITY OF THE OPTIMUM: FLAT TOPS AND THE ENVELOPE THEOREM

- ☐ evaluate how sensitive an optimum and its value are to sizing errors and parameter changes, using the flat-top approximation and the envelope theorem
- ☐ estimate the in-sample growth-optimal leverage of a return series in Python, and explain why it is not a recommendation

Two questions remain once an optimum is found: what does missing it cost, and how do the optimum and its value move with the inputs? The second-order term of a Taylor expansion answers the first, the envelope theorem the second. Together they resolve the case.

Flat tops

At an interior maximum the first-order term of the Taylor expansion vanishes, so

$$g(f) = g^* + \frac{1}{2}g''(\xi)(f - f^*)^2 \approx g^* + \frac{1}{2}g''(f^*)(f - f^*)^2, \quad (4.14)$$

where the approximation drops terms of third order in $f - f^*$. The loss from mis-sizing is of second order in the sizing error. An error twice as large costs four times as much, and a small error costs very little. The top of a smooth maximum is flat.

In the quadratic model the flat top has an exact form. Write (4.2) as $g(f) = f\mu - \frac{1}{2}f^2m_2$, with $f^* = \mu/m_2$ and $g^* = \mu^2/(2m_2)$. Substituting $f = cf^*$,

$$g(cf^*) = c \frac{\mu^2}{m_2} - \frac{c^2}{2} \frac{\mu^2}{m_2} = (2c - c^2) g^* = [1 - (1 - c)^2] g^*. \quad (4.15)$$

RESULT Fractional Kelly in the quadratic model

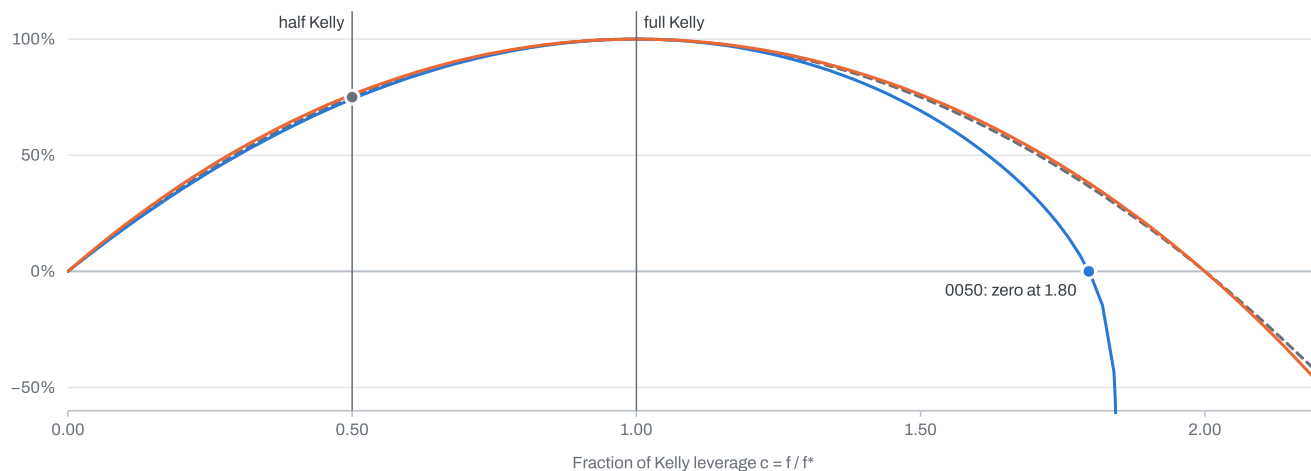
A position of c times the Kelly leverage keeps the fraction $1 - (1 - c)^2$ of the maximal growth rate, with c times the volatility. Half Kelly keeps 75% of the growth at half the volatility; quarter Kelly keeps 43.75% at a quarter. Over-betting mirrors under-betting: $1.5f^*$ also keeps 75%, at three times the volatility of half Kelly. Growth is zero at $c = 2$ and negative beyond.

This is Thorp's (2006) $c(2 - c)$. Exhibit 7 compares (4.15) with the exact curves for 0050 and for the +2%/−1% bet.

Exhibit 7: Growth kept at a fraction c of the Kelly leverage

--- Quadratic model: $1 - (1 - c)^2$ — 0050 in-sample (exact) — +2% / -1% bet, $p = 0.5$ (exact)

Growth kept, $\bar{g}(cf^*) / \bar{g}(f^*)$



Note: 0050 uses the exact in-sample ratio $\bar{g}(cf^*) / \bar{g}(f^*)$ with $f^* = 5.42$; the curve ends at $c = 10 / f^* = 1.845$, where 10× leverage meets the -10% day. The bet uses (4.7) with $f^* = 25$.

Below full Kelly the three curves nearly coincide: half Kelly keeps 75% in the model, 74.2% for 0050 and 76.1% for the bet. Beyond full Kelly they part: the bet's curve is exactly symmetric about $c = 1$ and reaches zero at $c = 2$, as Example 4 found. 0050 falls away: 1.5 times Kelly keeps 69.1% instead of 75%, 1.75 times keeps 17.9%, and growth is zero at 1.80 times. Over-betting a fat-tailed asset is far worse than the parabola suggests; under-betting is not.

EXAMPLE 6

Half Kelly on 0050

Compare full and half Kelly on 0050 in-sample, and check the flat-top approximation (4.14) for the growth lost.

SOLUTION

From Exhibit 1, half Kelly ($f = 2.71$) earned 43.63% a year against 58.84%, keeping 74.2% of the maximal growth rate; the quadratic model says 75%. In exchange it halved the volatility (52.3% against 104.7%), halved the worst day (-27.1% against -54.2%) and cut the maximum drawdown from -93.3% to -70.1%. Terminal wealth was 183 against 1,122: a smaller multiple from a far more survivable path.

For (4.14), $\bar{g}''(f^*) = -1.804 \times 10^{-4}$ a day, or -0.0439 a year. The predicted loss at $f^*/2$ is $\frac{1}{2} \times 0.0439 \times 2.711^2 = 16.1\%$ a year, against an actual loss of $58.84\% - 43.63\% = 15.2\%$. The quadratic term gets the size of the loss right to within one percentage point, 2.7 units of leverage away from the top.

The envelope theorem

THEOREM Envelope theorem (one variable)

Let $F(x, \theta)$ be differentiable, and for each θ let $x^*(\theta)$ be an interior maximizer of $F(\cdot, \theta)$ that depends differentiably on θ . Then the value function $V(\theta) := F(x^*(\theta), \theta)$ satisfies

$$V'(\theta) = \frac{\partial F}{\partial \theta}(x^*(\theta), \theta). \quad (4.16)$$

Proof. By the chain rule, $V'(\theta) = F_x(x^*(\theta), \theta) x^{*'}(\theta) + F_\theta(x^*(\theta), \theta)$. The first-order condition sets $F_x = 0$ at the maximizer, which leaves F_θ . ■

The theorem is the flat top seen from the side. When a parameter moves the optimizer moves too, but to first order this does not change the value. To price a parameter change, differentiate the objective with respect to the parameter at the old optimum; no re-optimization is needed. Simon and Blume (1994, chapter 19) give the multivariable version; two applications follow.

- **The value of edge.** In the quadratic model without the μ^2 term, $g(f) = f\mu - \frac{1}{2}f^2\sigma^2$ and $V(\mu) = \mu^2/(2\sigma^2)$, so $V'(\mu) = \mu/\sigma^2 = f^*$. The envelope theorem gets there in one line: $\partial g/\partial \mu = f$, evaluated at f^* . The marginal value of edge is the position you hold.
- **The cost of financing**, worked out in Example 7.

EXAMPLE 7

What a higher financing rate costs a levered 0050 position

Suppose the investor borrows (and lends idle cash) at a simple rate r_f per trading day, so that one day multiplies wealth by $1 + r_f + f(R_t - r_f)$. Use the envelope theorem to find how maximal in-sample growth responds to r_f near zero. Check the answer for an illustrative financing rate of 2% a year, charged as 2%/A per trading day.

SOLUTION

With $F(f, r_f) = \frac{1}{T} \sum_t \ln(1 + r_f + f(R_t - r_f))$, the partial derivative is $\partial F / \partial r_f = (1 - f) \frac{1}{T} \sum_t 1 / (1 + r_f + f(R_t - r_f))$. At $r_f = 0$ and $f = f^*$ the first-order condition does the work:

$$\frac{1}{T} \sum_t \frac{1}{1 + f^* R_t} = \frac{1}{T} \sum_t \frac{(1 + f^* R_t) - f^* R_t}{1 + f^* R_t} = 1 - f^* \bar{g}'(f^*) = 1.$$

So $V'(0) = 1 - f^*$ exactly, that is -4.42 : each unit of financing rate costs 4.42 units of maximal growth. In annual terms, a rate one percentage point higher costs about 4.42 percentage points a year of growth. The investor is short the financing rate on the $f^* - 1 = 4.42$ borrowed units.

At 2% a year the first-order estimate is $-4.42 \times 2\% = -8.84\%$. Re-optimizing exactly gives $f^* = 4.96$ and growth of 50.46% a year, a fall of 8.38%. Keeping the old leverage of 5.42 still earns 49.99%: re-optimizing is worth only 0.47% a year, which is precisely the envelope theorem's point.

How well the data locate the optimum

The flat top has a second face. If a wide range of leverages earns nearly the same growth, the data can barely tell those leverages apart. The value of the optimum is robust; its location is fragile.

For independent returns, the standard error of a sample mean is $s/\sqrt{T}$ (stated without proof; LM20 derives it and develops confidence intervals). A 95% confidence interval for the true mean is $\hat{\mu} \pm 1.96$ standard errors, where 1.96 is the value a standard normal variable exceeds with probability 2.5%.

EXAMPLE 8

How precisely the sample locates 0050's Kelly leverage

Turn the sampling error of $\hat{\mu}$ into a 95% band for 0050's Kelly leverage. Shift all daily returns equally, putting the sample mean at each end of the 95% interval, and re-solve the exact Kelly condition. What would holding 5.42 achieve at either end of the band?

SOLUTION

The standard error is $1.2385\% / \sqrt{2,902} = 0.0230\%$ a day against $\hat{\mu} = 0.0871\%$, a relative error of 26%. The 95% interval for the mean runs from 0.042% to 0.132% a day. On the shifted samples the exact Kelly leverage runs from 2.7 to 7.6. The quadratic rule $\hat{\mu}/s^2$ would give 2.7 to 8.6, missing the fat tails at the high end.

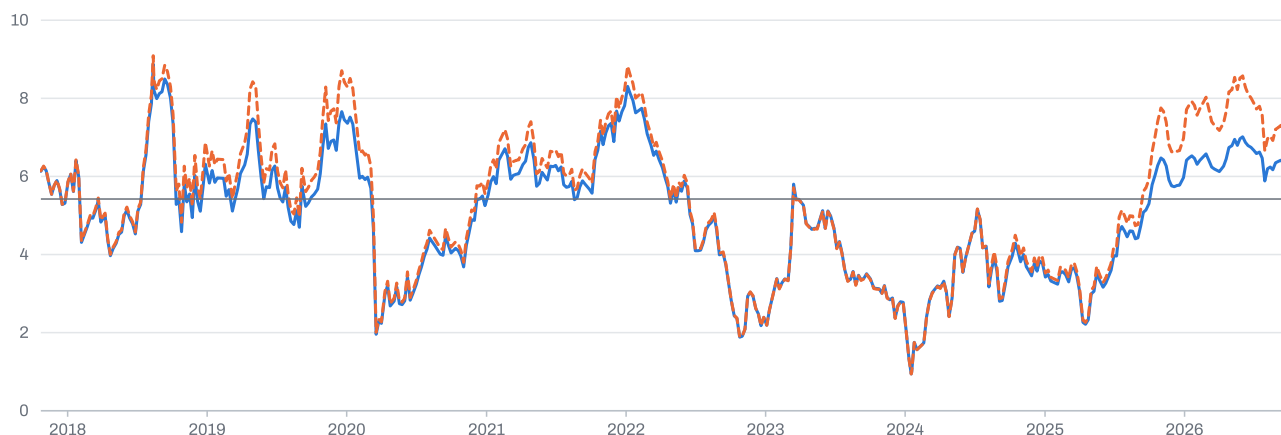
The cost of the error is asymmetric. If the true Kelly leverage were 2.7, holding 5.42 would be $c \approx 2.0$, where (4.15) leaves almost nothing. On the shifted sample's exact curve, 5.42 even loses 0.63% a year. If the true value were 7.6, holding 5.42 would be $c = 0.71$, which keeps 92% by (4.15) and 90% on the shifted sample.

The data's own history agrees. Over trailing three-year windows (Exhibit 8) the in-sample optimum wandered from 0.94 to 8.87, a range of more than nine to one. The low came in the window ending 17 January 2024, the high in the window ending 13 August 18. The latest window says 6.63.

Exhibit 8: Empirical Kelly leverage of 0050 over trailing three-year windows

— Exact empirical Kelly - - - Quadratic rule: sample mean $\div$ s^2 (same window)

Empirical Kelly leverage f^* ($\times$ equity), trailing 729 trading days



Note: each point maximizes the mean of $\ln(1 + fR_t)$ over the preceding 729 trading days (about three years); the first window ends on 23 October 2017. Every fifth window is plotted, plus the latest and the two extremes. The dashed line is the quadratic rule $\hat{\mu}/s^2$ on the same window, and the horizontal line marks the full-sample optimum, 5.42. Source: as Exhibit 1.

Under-betting costs a little; over-betting can cost everything. That asymmetry, not timidity, is the case for fractional Kelly, a quarter to a half of f^* . [LM24](#) and [LM44](#) formalize it, and [LM45](#) adds parameter uncertainty. MacLean, Thorp and Ziemba (2010) summarize the trade-off: maximal long-run growth, but large positions and a real chance of losing most of the wealth.

PITFALL "The backtest Sharpe ratio is 3, so use full Kelly"

In the quadratic model $f^* \approx \mu/\sigma^2 = \text{SR}/\sigma$ and $g^* \approx \text{SR}^2/2$, with SR and σ annualized ([LM42](#)). A backtest Sharpe ratio of 3 at 15% volatility suggests $f = 3/0.15 = 20$ and log growth of 4.5 a year. That is a 90-fold gain every year, which should itself raise suspicion.

If the true Sharpe ratio is half the backtest's, 20 $\times$ is twice Kelly and growth is zero. If it is 1, 20 $\times$ is three times Kelly, and by (4.15) growth is $(2 \times 3 - 3^2) \times 1^2/2 = -1.5$ a year. Wealth then falls 78% a year despite a genuine edge. Backtests overstate Sharpe ratios ([LM38](#), [LM39](#)): size from a haircut estimate, then take a fraction of Kelly.

Resolving the case

- 5.42 is the unique global maximizer of in-sample growth, because $\bar{g}$ is strictly concave (Section 3); Newton's method finds it in four steps from zero (Section 6).

- **5.68** is the quadratic rule $\hat{\mu}/s^2$, and 5.65, one Newton step from zero, its version with $\hat{\mu}^2$ kept. Both overshoot because 0050's daily returns are fat-tailed (Section 2).
- **74.2%** is the flat top: half Kelly kept almost three-quarters of the growth rate at half the volatility.
- **9.74**, the zero-growth leverage, is 1.80 times f^* rather than 2, pulled in by the largest losses (Section 3).
- **2.7 to 7.6** is the range of Kelly leverages consistent with the sampling error in $\hat{\mu}$; the rolling windows wandered from 0.94 to 8.87.

The in-sample 5.42× therefore describes 2014–2026; it is not a recommendation. It treats the sample mean as truth, assumes free daily rebalancing and ignores financing, which costs 8.38% a year of growth at 2%. Even with perfect hindsight, it came with a 93.3% drawdown. [LM20](#) and [LM45](#) supply the statistics that must come before any leverage decision.

導讀 最適點的位置脆弱，價值卻很穩

最大值附近曲線很平（一階項為零，只剩二階項）：槓桿抓錯一點，成長率幾乎不變。在二次模型裡， c 倍凱利保留 $1 - (1 - c)^2$ 的成長率：半凱利保留 75%、波動減半，0050 實際是 74.2%。但「平」的另一面是資料很難分辨最適點在哪：0050 平均報酬的估計誤差約 26%，凱利槓桿的 95% 區間約 2.7 到 7.6 倍，滾動三年的樣本內最適值在 0.94 到 8.87 倍之間跳動。而且押太多比押太少危險得多，0050 超過 1.8 倍凱利，長期成長率就轉負。包絡定理則說：參數變動對「最大成長率」的一階影響，只要對參數偏微分、在原最適點取值即可，不必重新最佳化；例如融資利率每高 1 個百分點，最大成長率約少 $f^* - 1 = 4.42$ 個百分點。

KNOWLEDGE CHECK 6

In the quadratic model, what share of maximal growth does 1.5 times Kelly keep, and how does its volatility compare with half Kelly's? What does 0050's exact in-sample curve say?

ANSWER

$1 - (1 - 1.5)^2 = 75\%$, the same as half Kelly, at three times half Kelly's volatility: identical growth for triple the risk. On 0050's exact curve 1.5 times Kelly keeps only 69.1%, because fat tails punish over-betting more than the parabola does.

8 OPTIMIZATION IN PYTHON

- ☐ estimate the in-sample growth-optimal leverage of a return series in Python, and explain why it is not a recommendation

SciPy's `optimize` module contains every method of this module. Three habits keep results honest: pass the domain explicitly, give a bracket whenever one is known, and compare at least two methods.

PYTHON

```
import numpy as np
import polars as pl
from scipy.optimize import brentq, minimize_scalar, newton

# 0050 total returns on days common with 00631L, from the lab's data/ folder
common = pl.read_csv("data/00631L_daily_raw.csv").select("date")
div = pl.read_csv("data/0050_dividends.csv")
px = (pl.read_csv("data/0050_daily_raw.csv").join(common, on="date")
      .join(div, on="date", how="left").sort("date"))
split = pl.when(pl.col("date") == pl.lit("2025-06-18")).then(4.0).otherwise(1.0)
P, D = pl.col("close"), pl.col("cash_dividend").fill_null(0.0)
R = (px.select((P + D) * split / P.shift(1) - 1)
     .to_series().drop_nulls().to_numpy())

def gbar(f):
    return np.log1p(f * R).mean()

def dg(f):
    return (R / (1 + f * R)).mean()

def d2g(f):
    return -((R / (1 + f * R)) ** 2).mean()

lo, hi = -1 / R.max(), 1 / -R.min()
f_brent = brentq(dg, 0.0, 0.999 * hi, xtol=1e-14)
f_newton = newton(dg, x0=0.0, fprime=d2g, tol=1e-14)
f_bound = minimize_scalar(lambda f: -gbar(f), bounds=(0.0, 0.999 * hi),
                          method="bounded").x
print(f"domain ({lo:.2f}, {hi:.2f})  brentq {f_brent:.6f}  "
      f"newton {f_newton:.6f}  bounded {f_bound:.4f}")
# domain (-10.00, 10.00)  brentq 5.421488  newton 5.421488  bounded 5.4215
```

`minimize_scalar` with `method="bounded"` maximizes $\bar{g}$ directly by Brent's derivative-free search. It stops at its default tolerance of about 10^{-5} , which is why it is printed with fewer digits. The lab applies `gbar` to the fractional-Kelly curves of Exhibit 7 and traces Newton's method and bisection for the yield of Example 5.

KNOWLEDGE CHECK 7

Called without bounds, `minimize_scalar(lambda f: -gbar(f))` returns `nan` on the 0050 data, and SciPy reports that it found no valid bracket. Explain why, and give two fixes.

ANSWER

The unbounded Brent search expands its trial points beyond the domain $(-10, 10)$. There $1 + fR_t \leq 0$ on some day, `np.log1p` returns `nan` or $-\infty$, and the comparisons that drive the search break down. Either restrict the search, `minimize_scalar(..., bounds=(0.0, 0.999 * hi), method="bounded")`, or solve the first-order condition with `brentq` on a bracket inside the domain.

IN PYTHON Companion lab

The Polars-first notebook `LM04_lab.ipynb` rebuilds every computational exhibit and example 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.

SUMMARY

- At an interior maximum of a differentiable function the derivative is zero (Fermat's theorem). A negative second derivative confirms a strict local maximum; a zero second derivative decides nothing; use the tests of Exhibit 2.
- For a differentiable concave function a zero of the derivative is a global maximum, and strict concavity makes it unique. The empirical growth function $\bar{g}(f)$ is strictly concave, so the empirical Kelly leverage exists and is unique whenever the sample contains gains and losses.
- On a closed interval, compare critical points with endpoints. At a binding corner the derivative is the constraint's shadow price, a marginal rate; for a concave objective the constrained optimum is the unconstrained one clipped to the interval.
- The binary Kelly bet has the exact solution $f^* = p/a - q/b = (1/a + 1/b)(p - \pi)$. It is a position size; the risk size is $f^*a = p - qa/b$. The position moves by $1/a + 1/b$ per unit of win probability, and attainable growth depends only on p and the payoff ratio b/a .
- Bisection gains a binary digit per step; Newton's method squares the error near a simple root but needs a good start. On 0050's Kelly condition, Newton's first step from zero is the quadratic rule, and four steps reach machine precision. Bisection needs 44 steps to shrink the bracket below 10^{-12} .
- Near an optimum, growth is flat: cf^* keeps $1 - (1 - c)^2$ of the maximum in the quadratic model. The envelope theorem prices parameter changes without re-optimizing: at the empirical Kelly point,

each unit of financing rate costs $f^* - 1$ units of growth.

- In-sample, 0050's growth-optimal leverage over 2014–2026 is 5.42, with 58.84% a year of log growth and a 93.3% drawdown. Sampling error spans 2.7 to 7.6, and over-betting costs far more than under-betting: the number describes the past and is not a recommendation.

PRACTICE PROBLEMS

1. The function $h(x) = xe^{-x}$ on $x \geq 0$ attains its maximum at:
A. $x = 0$
B. $x = 1/e$
C. $x = 1$
2. Which statement about $h(x) = x^4$ at $x = 0$ is correct?
A. The second-derivative test is inconclusive, and $x = 0$ is a strict local minimum.
B. The second-derivative test shows a local maximum.
C. $x = 0$ is an inflection point.
3. A strategy's annual log growth as a function of leverage is approximately $g(f) = 0.12f - 0.08f^2$. Find the Kelly leverage, the maximal growth, the growth at half Kelly and the leverage at which growth returns to zero.
4. A trade gains 5% with probability 0.55 and otherwise loses 5%. The Kelly position is *closest* to:
A. 0.10 times equity
B. 2 times equity
C. 11 times equity
5. Show that the Kelly position of a binary bet is positive if and only if the expected return per unit of position, $pb - qa$, is positive. Show also that the maximal growth g^* is then positive.
6. A trading rule wins +8% with probability 0.30 and stops out at -2%. (a) Find the Kelly position, the risk per trade and the maximal growth per trade. (b) If the true win rate is 0.25, by what factor is the position from (a) too large, and what growth per trade does it earn?
7. The number of bisection steps needed to shrink a bracket $[0, 1]$ around a root to a width of 10^{-8} is *closest* to:
A. 8
B. 19
C. 27

8. An investor's annual growth is approximately $g(f) = 0.10f - 0.05f^2$, but the account limits leverage to $0 \leq f \leq 0.8$. Find the optimal leverage, the growth it delivers and the shadow price of the limit. What would raising the limit to 0.9 add?
9. In the quadratic model $g(f) = f\mu - \frac{1}{2}f^2\sigma^2$, use the envelope theorem to find $dg^*/d\sigma$, and verify the result by differentiating $g^* = \mu^2/(2\sigma^2)$ directly.
10. In the quadratic model, half Kelly keeps:
 - A. 75% of the maximal growth rate at 50% of the full-Kelly volatility
 - B. 50% of the maximal growth rate at 50% of the full-Kelly volatility
 - C. 75% of the maximal growth rate at 25% of the full-Kelly volatility
11. A backtest reports an annualized Sharpe ratio of 2.4 at 12% annualized volatility. (a) What leverage does full Kelly suggest in the quadratic model? (b) For which true Sharpe ratio does that leverage produce zero long-run growth? (c) What are the annual log growth and the annual wealth factor if the true Sharpe ratio is 0.8?
12. (Python) Using the lab data, compute the empirical Kelly leverage of 0050 from the returns of 2 January 2020 to 8 October 2026. Compare it with $\hat{\mu}/s^2$ for the same window.
13. Expand the binary growth function (4.7) to second order around $f = 0$, and show that the resulting approximate Kelly position is $(pb - qa)/(pb^2 + qa^2)$. Evaluate it for the +2%/-1% bet at $p = 0.5$ and explain why it falls short of the exact 25.

SOLUTIONS

1. C is correct. $h'(x) = (1 - x)e^{-x} = 0$ at $x = 1$, and $h''(x) = (x - 2)e^{-x}$ gives $h''(1) = -e^{-1} < 0$. The maximum value is $e^{-1} = 0.3679$, which B mistakes for the maximizer. The endpoint $x = 0$, where $h'(0) = 1 > 0$, is the minimum on $x \geq 0$.
2. A is correct. $h'(0) = h''(0) = h'''(0) = 0$ and $h^{(4)}(0) = 24 > 0$: the first non-zero derivative has even order and is positive, so $x = 0$ is a strict local (indeed global) minimum (Exhibit 2). C would require the first non-zero derivative to have odd order, as for x^3 .
3. $g'(f) = 0.12 - 0.16f = 0$ gives $f^* = 0.75$, and $g'' = -0.16 < 0$ confirms the maximum. Maximal growth is $g^* = 0.12 \times 0.75 - 0.08 \times 0.5625 = 0.045$, or 4.5% a year. Half Kelly gives $g(0.375) = 0.045 - 0.01125 = 0.03375$, which is 75% of g^* , as (4.15) predicts. Growth returns to zero at $f = 0.12/0.08 = 1.5 = 2f^*$.
4. B is correct. $f^* = 0.55/0.05 - 0.45/0.05 = 11 - 9 = 2$. The risk per trade is $2 \times 5\% = 10\%$, which is also $p - q$ a/b with $a = b$; A mistakes that risk fraction for the position. C keeps only the first term, p/a , and ignores the losses.
5. By (4.10), $f^* = (1/a + 1/b)(p - \pi)$ with $1/a + 1/b > 0$, so $f^* > 0$ if and only if $p > \pi = a/(a + b)$, that is, $p(a + b) > a$, that is, $pb - qa > 0$. In that case f^* is the unique maximizer of the strictly concave g ,

and $g(0) = 0$, so $g^* = g(f^*) > g(0) = 0$. (Alternatively, the Deep Dive shows $g^* = D_{\text{KL}}(p||\pi) > 0$ for $p \neq \pi$.)

6. (a) $f^* = 0.30/0.02 - 0.70/0.08 = 15 - 8.75 = 6.25$ times equity, risking $6.25 \times 2\% = 12.5\%$ per trade; check: (4.11) gives $0.30 - 0.70 \times 0.02/0.08 = 0.125$. With $1 + f^*b = 1.5$ and $1 - f^*a = 0.875$, $g^* = 0.30 \ln 1.5 + 0.70 \ln 0.875 = 2.817\%$ per trade.

(b) At $p = 0.25$, $f^* = 12.5 - 9.375 = 3.125$, so $6.25\times$ is exactly twice Kelly: by (4.10), f^* halves with the edge over break-even ($\pi = 0.20$), from 0.10 to 0.05 . Growth at $6.25\times$ is $0.25 \ln 1.5 + 0.75 \ln 0.875 = 0.122\%$ per trade, 16.5% of the attainable 0.738% . Equation (4.15) predicts zero, but the exact zero is at $6.56\times$, 2.10 times Kelly. A rare large gain ($p < 0.5$) makes growth fall more slowly beyond f^* than it rises below it; at $p = 0.5$ the curve is symmetric.

7. C is correct. $n \geq \log_2(1/10^{-8}) = 26.6$, so 27 steps. A counts decimal digits instead of binary ones; B uses the natural logarithm, $\ln 10^8 = 18.4$, instead of $\log_2$.

8. The unconstrained optimum is $f^* = 0.10/0.10 = 1.0 > 0.8$, so by clipping the optimum is the corner $f = 0.8$, with $g(0.8) = 0.08 - 0.032 = 0.048$ against an unconstrained 0.050 . The shadow price is $g'(0.8) = 0.10 - 0.10 \times 0.8 = 0.02$ a year per unit of leverage: a marginal rate, not a total. Raising the limit to 0.9 adds about $0.02 \times 0.1 = 0.2$ percentage points to first order and exactly $g(0.9) - g(0.8) = 0.15$ points, because g' falls as f rises. A limit above 1 adds nothing more: the whole gain on offer is 0.2 points.

9. $\partial(f\mu - \frac{1}{2}f^2\sigma^2)/\partial\sigma = -f^2\sigma$; at $f^* = \mu/\sigma^2$ this is $-\mu^2/\sigma^3$. Directly, $d(\mu^2/(2\sigma^2))/d\sigma = -\mu^2/\sigma^3$. The two agree, as (4.16) requires; the change in f^* never enters.

10. A is correct. By (4.15) with $c = \frac{1}{2}$, the growth kept is $1 - (1/2)^2 = 75\%$, and volatility scales with c . B scales growth with the position, as if growth were linear in f ; C confuses volatility with variance, which falls to 25% .

11. (a) $f \approx \text{SR}/\sigma = 2.4/0.12 = 20$. (b) Growth is zero at $c = 2$, so when the true Kelly leverage is 10 , which means a true Sharpe ratio of $10 \times 0.12 = 1.2$: half the backtest's. (c) A true Sharpe ratio of 0.8 gives $f^* = 6.67$ and $c = 3$, so growth is $(2 \times 3 - 3^2) \times 0.8^2/2 = -0.96$ a year, a wealth factor of $e^{-0.96} = 0.383$ a year.

12. Over the 1,635 returns from 2 January 2020 to 8 October 2026, the empirical Kelly leverage is 5.38 and $\hat{\mu}/s^2 = 5.66$. As over the full sample (5.68 against 5.42), $\hat{\mu}/s^2$ is about 5% too high. The window's optimum differs from the full-sample 5.42 by far less than the sampling error of Section 7.

13. Using $\ln(1+x) \approx x - x^2/2$ with $x = fb$ and $x = -fa$: $g(f) \approx f(pb - qa) - \frac{1}{2}f^2(pb^2 + qa^2)$, a parabola with its maximum at $(pb - qa)/(pb^2 + qa^2)$. This is the quadratic rule (4.3) with $\mu = pb - qa$ and $\mathbb{E}[R^2] = pb^2 + qa^2$. For $+2\%/-1\%$ at $p = 0.5$ it gives $0.005/0.00025 = 20$, against the exact 25 .

The neglected third-order term, $f^3(pb^3 - qa^3)/3$, is positive here, so it raises growth at high leverage and pushes the true optimum up. The bet is symmetric about its mean $\mu = 0.5\%$ ($\pm 1.5\%$), so $pb^3 - qa^3 = \mu^3 + 3\mu\sigma^2 = 3.5 \times 10^{-6}$ with $\sigma = 1.5\%$: positive because $\mu > 0$. A positive mean alone does not settle the sign. For $+1\%$ with probability 0.9 against -5% , $pb - qa = +0.4\%$ but $pb^3 - qa^3 < 0$, and the quadratic rule overshoots: 11.8 against an exact 8 .

GLOSSARY

Term	中文	Meaning
Critical point	臨界點	Interior point where $h' = 0$ or h' does not exist.
First-order condition	一階條件	$h'(x^*) = 0$ at an interior extremum (Fermat's theorem).
Second-order condition	二階條件	Sign of h'' at a critical point: negative for a strict local maximum.
Local / global maximum	局部／全域極大值	Best value in a neighbourhood / on the whole domain.
Inflection point	反曲點	Point where curvature changes sign; not an extremum.
Concave function	凹函數	Chords lie on or below the graph ($h'' \leq 0$); if h is differentiable, a point with $h' = 0$ is a global maximum.
Convex function	凸函數	Chords lie on or above the graph ($h'' \geq 0$).
Intermediate value theorem	中間值定理	A continuous function that changes sign on an interval is zero somewhere in it.
Extreme value theorem	極值定理	A continuous function on $[\alpha, \beta]$ attains its maximum and minimum.
Corner solution	角解	Optimum at a boundary of the feasible interval.
Shadow price	影子價格	Rate at which relaxing a binding constraint raises the objective.
KKT conditions	KKT 條件	First-order conditions with inequality constraints; at a corner, a sign condition on h' .
Kelly criterion	凱利準則	Choose the position that maximizes expected log growth.
Kelly position	凱利部位	Growth-optimal position f^* , as a multiple of equity.
Risk fraction	風險比例	Fraction of equity lost if a trade fails, fa .
Break-even probability	損益兩平勝率	$\pi = a/(a + b)$: the win probability with zero expected return.
Fractional Kelly	分數凱利	A position cf^* with $c < 1$, such as half or quarter Kelly.
Bisection method	二分法	Halve a sign-change bracket at each step; linear convergence.

Term	中文	Meaning
Newton's method	牛頓法	$x_{n+1} = x_n - u(x_n)/u'(x_n)$; quadratic convergence near a simple root.
Quadratic convergence	二次收斂	$e_{n+1} \approx Ce_n^2$; correct digits roughly double per step.
Envelope theorem	包絡定理	$V'(\theta) = \partial F / \partial \theta$ evaluated at the optimum.
Yield to maturity	到期殖利率	Discount rate that equates a bond's present value to its price.
Internal rate of return	內部報酬率	Discount rate at which net present value is zero.
Relative entropy	相對熵	$D_{\text{KL}}(p \pi) = p \ln(p/\pi) + q \ln(q/(1-\pi))$; equals the binary bet's g^* .

REFERENCES

- Kelly, J. L., Jr. (1956). "A New Interpretation of Information Rate." *Bell System Technical Journal* 35 (4): 917–926.
- Breiman, L. (1961). "Optimal Gambling Systems for Favorable Games." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 65–78. University of California Press.
- Thorp, E. O. (2006). "The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market." In S. A. Zenios and W. T. Ziemba (eds.), *Handbook of Asset and Liability Management*, Vol. 1, 385–428. Amsterdam: Elsevier.
- MacLean, L. C., E. O. Thorp, and W. T. Ziemba (eds.) (2011). *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific Handbook in Financial Economics Series, Vol. 3. Singapore: World Scientific.
- MacLean, L. C., E. O. Thorp, and W. T. Ziemba (2010). "Long-Term Capital Growth: The Good and Bad Properties of the Kelly and Fractional Kelly Capital Growth Criteria." *Quantitative Finance* 10 (7): 681–687.
- Burden, R. L., J. D. Faires, and A. M. Burden (2016). *Numerical Analysis*, 10th ed., chapter 2, "Solutions of Equations in One Variable". Boston: Cengage Learning.
- Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall (Dover reprint, 2002).
- Stewart, J., D. K. Clegg, and S. Watson (2021). *Calculus: Early Transcendentals*, 9th ed., chapter 2 (continuity and the intermediate value theorem) and chapter 4, "Applications of Differentiation" (maximum and minimum values, optimization problems, Newton's method). Cengage (ISBN 9781337613927).

- Simon, C. P., and L. Blume (1994). *Mathematics for Economists*. New York: W. W. Norton (chapter 19 on envelope theorems; chapter 21 on concave functions).
- Cover, T. M., and J. A. Thomas (2006). *Elements of Information Theory*, 2nd ed., chapter 6, "Gambling and Data Compression". Hoboken, NJ: Wiley.
- 臺灣證券交易所. 《臺灣證券交易所60週年特刊》, 05 證券交易市場, p. 181: the daily price limit was widened from 7% to 10% on 1 June 2015. Online edition, page 181.
- FinMind open data: `TaiwanStockPrice` and `TaiwanStockDividendResult` (0050, 00631L), retrieved 10 October 2026.