Vanguard Total World Stock ETF (VT)

Price analysis since inception, 26 Jun 2008 – 30 Oct 2026

Cédric Masset · 4,616 trading days · Yahoo Finance data, validated against Vanguard

Summary

Metric	Value
Price at inception → latest price	USD 49.58 → 155.33
Cumulative return (price / total return)	213% / 360%
Annualised return (price / total return)	6.4% / 8.7%
Annualised total return in CHF	7.4%
Annualised volatility (monthly returns; daily $\times\sqrt{252}$ in brackets)	16.5% (20.4%)
Sharpe ratio, 95% confidence interval (average risk-free rate 1.4%)	0.50 [0.02, 0.99]
Sortino ratio	0.73
Maximum drawdown (total return)	−50.3% (trough on 09 Mar 2009)
Calmar ratio	0.17
Kelly exposure with borrowing spread, in-sample (block-bootstrap 95% interval)	1.81x [0.24x, 4.28x]
Probability of loss over 1 / 5 / 10 years: historical windows	22% / 0% / 0%
Probability of loss over 1 / 5 / 10 years: block bootstrap	24% / 10% / 4%

Key findings

- Over 18.3 years, VT returned 8.7% a year with dividends reinvested, of which roughly 2.2% a year came from dividends. These figures match Vanguard's published market-price returns to within 0.01 percentage point.
- Measured on monthly returns, annualised volatility is 16.5%; scaling daily volatility by $\sqrt{252}$ gives 20.4% because daily returns are negatively autocorrelated. The maximum drawdown was −50% during the global financial crisis.
- The Sharpe ratio of 0.50 has a 95% confidence interval of [0.02, 0.99]: eighteen years of data are barely enough to distinguish it from zero.
- Daily returns have fat tails (kurtosis 16.1) and strongly persistent volatility (GARCH persistence 0.993, half-life of about 96 trading days). Out of sample, a GARCH(1,1)-t model gives the best variance forecasts by the QLIKE criterion, but its edge over a simple EWMA is not statistically significant. It passes the 5% VaR backtest, unlike rolling historical and normal VaR, whose breaches cluster.
- After correcting for multiple testing, the only rejection of the random walk is a small negative autocorrelation of daily returns (−0.105), not detectable at weekly and monthly frequencies. No calendar effect is detected, and moving-average timing rules underperformed buy and hold whatever the window and execution lag.
- In hindsight, a Kelly exposure of about 1.81x would have maximised growth after borrowing costs, but its 95% bootstrap interval runs from 0.24x to 4.28x and a real-time estimate did not

beat buy and hold. Leverage raised the drawdown of 2008–2009 from 50% to 80% at 2x, and the compounding gap cost 4.2% a year at that exposure.

- For a Swiss franc investor, the weaker dollar cut the annualised return by 1.3% (7.4% in CHF), before taxes.

1 Return and risk

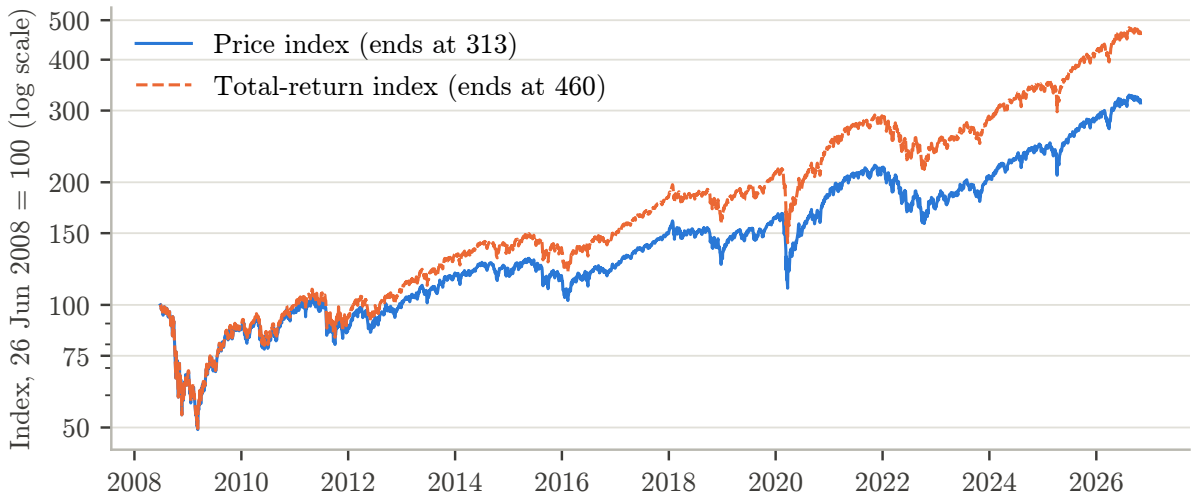


Figure 1. Price index and total-return index (dividends reinvested), both rebased to 100 at inception, log scale. The gap between the two lines is the cumulative contribution of reinvested dividends; equal slopes represent equal percentage returns.

1.1 Calendar-year returns

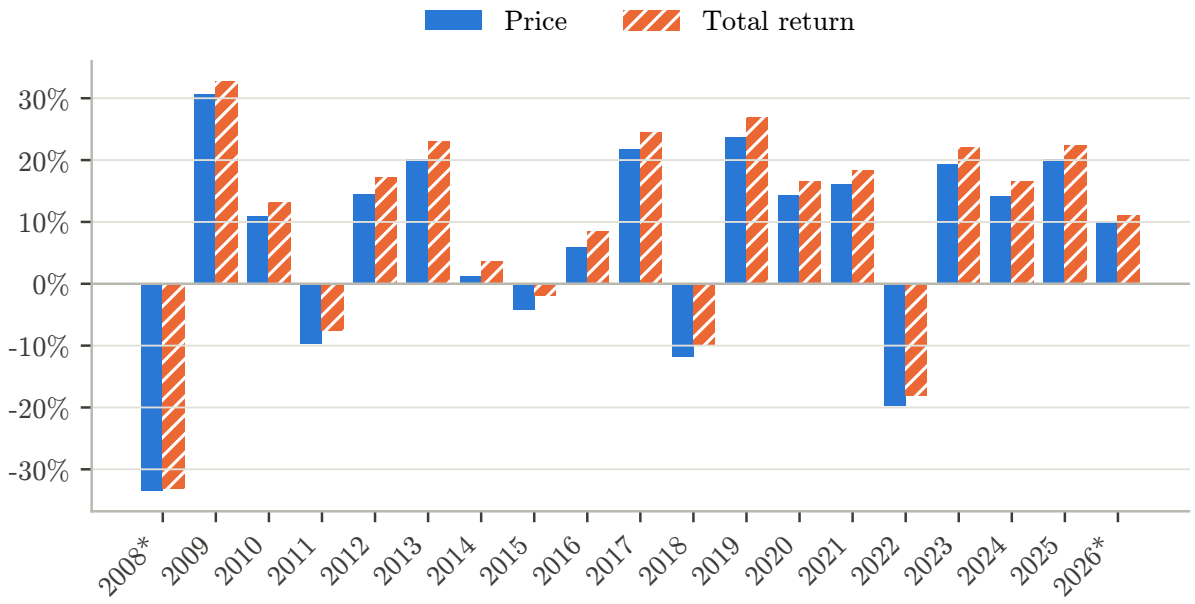


Figure 2. Calendar-year returns, price and total return. * Partial years.

* Partial years: 2008 (from inception) and 2026 (up to 30 Oct 2026). Of the 17 full years, the best was 2009 (+32.7%) and the worst 2022 (−18.0%); 4 years were negative.

1.2 Volatility

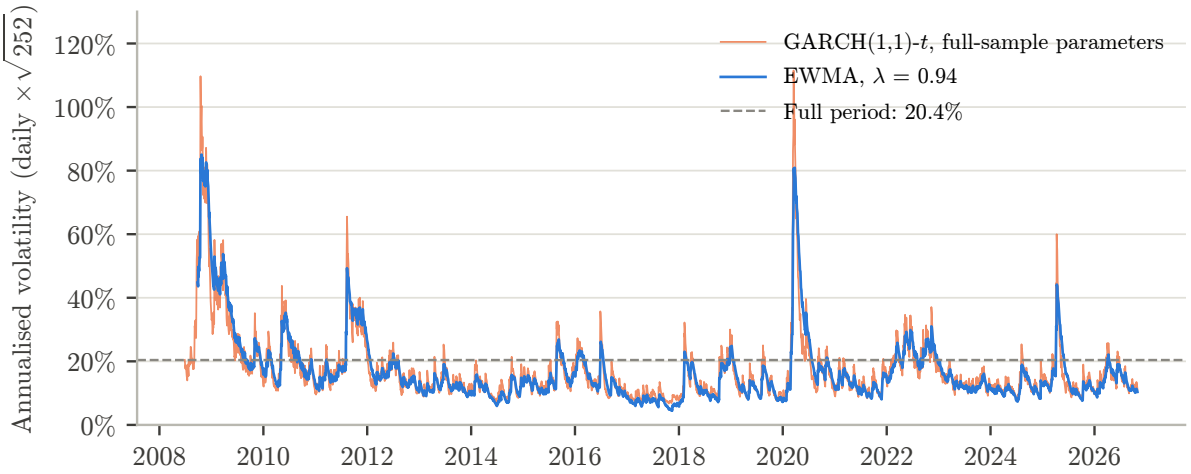


Figure 3. Annualised daily volatility: exponentially weighted moving average (RiskMetrics, $\lambda = 0.94$) and GARCH(1,1)-t conditional volatility with parameters estimated on the full sample. Dashed line: full-period volatility on the same daily basis.

Table 1. Annualised volatility by sampling frequency (non-overlapping total returns).

Frequency	Observations	Annualised volatility
Daily	4,615	20.4%
Weekly	957	18.3%
Monthly	220	16.6%
Quarterly	74	17.8%

Volatility varies sharply across regimes. The exponentially weighted estimate ranges from 4% (Nov 2017) to 85% (23 Oct 2008), and the GARCH conditional volatility peaks at 111% on 17 Mar 2020. Both measures react within days to a shock and decay as markets calm down. An equally weighted one-year window would instead keep an extreme day at full weight for exactly 252 sessions and then drop it abruptly, creating plateaus that do not reflect the actual level of risk. Both curves use daily returns scaled by $\sqrt{252}$, so the reference line is the full-period figure on the same basis (20.4%). The GARCH curve uses parameters estimated on the whole sample and is therefore descriptive; the out-of-sample forecasts are evaluated in Section 2.

The annualised level also depends on the sampling frequency. Scaling a k-period standard deviation by $\sqrt{k}$ is valid only if returns are serially uncorrelated. Daily returns are negatively autocorrelated (Section 3), so daily data annualised with $\sqrt{252}$ give 20.4%, whereas monthly data give 16.6%.

The variance of a sum of q daily returns decomposes into variances and autocovariances:

$$\operatorname{Var}\Big(\sum_{k=1}^{q} r_k\Big) = q\,\gamma_0\Big[1 + 2\sum_{k=1}^{q-1}\Big(1 - \frac{k}{q}\Big)\rho_k\Big].$$

With the first 20 daily autocorrelations, the bracket equals 0.72 for $q = 21$: serial dependence alone would reduce the monthly variance to 72% of 21 times the daily variance. The observed ratio is 0.66, with a 95% interval of about [0.48, 0.83] from the sampling error of 220 monthly variances (kurtosis of monthly returns 5.1). Autocorrelation thus explains most of the gap, and the remainder is within estimation error. This report uses monthly returns for annualised volatility and for the Sharpe and Sortino ratios.

1.3 Drawdowns



Figure 4. Decline of the total-return index from its previous peak.

Table 2. Five largest drawdowns (total return).

Peak	Trough	Recovery	Drawdown	Decline (days)	Recovery (days)
26 Jun 2008	09 Mar 2009	29 Dec 2010	−50.3%	175	457
12 Feb 2020	23 Mar 2020	24 Aug 2020	−34.2%	27	107
08 Nov 2021	12 Oct 2022	29 Jan 2024	−26.4%	233	324
29 Apr 2011	03 Oct 2011	02 Jan 2013	−23.8%	108	313
26 Jan 2018	24 Dec 2018	28 Oct 2019	−20.0%	229	212

Lengths in trading days. Because VT was launched during the global financial crisis, the first peak is the launch date itself.

1.4 Risk-adjusted performance by sub-period

Table 3. Return and risk by sub-period (total return; volatility and ratios from monthly returns).

Period	Ann. return	Volatility	Sharpe	95% CI	Sortino	Max. DD
2008–2012	1.9%	22.8%	0.19	[−0.74, 1.12]	0.26	−50.3%
2013–2017	11.1%	10.0%	1.07	[0.15, 1.99]	1.88	−19.8%
2018–2022	5.3%	17.9%	0.31	[−0.59, 1.20]	0.44	−34.2%
2023–2026	18.8%	12.2%	1.11	[0.10, 2.13]	2.10	−16.5%
Full period	8.7%	16.5%	0.50	[0.02, 0.99]	0.73	−50.3%

Excess returns over the 13-week US T-bill (`^IRX`). Sharpe: mean excess return divided by its volatility. Sortino: mean excess return divided by the downside deviation of excess returns. Confidence intervals for the Sharpe ratio use the standard error of Mertens (2002) and Opdyke (2007), which accounts for skewness and kurtosis.

The confidence intervals are wide: with about five years of monthly data per sub-period, all four sub-period intervals contain the full-period Sharpe ratio, and two contain zero. Differences between sub-periods should be read as descriptive rather than as evidence of a change in the fund's risk-return profile.

2 Distribution, volatility dynamics and Value at Risk

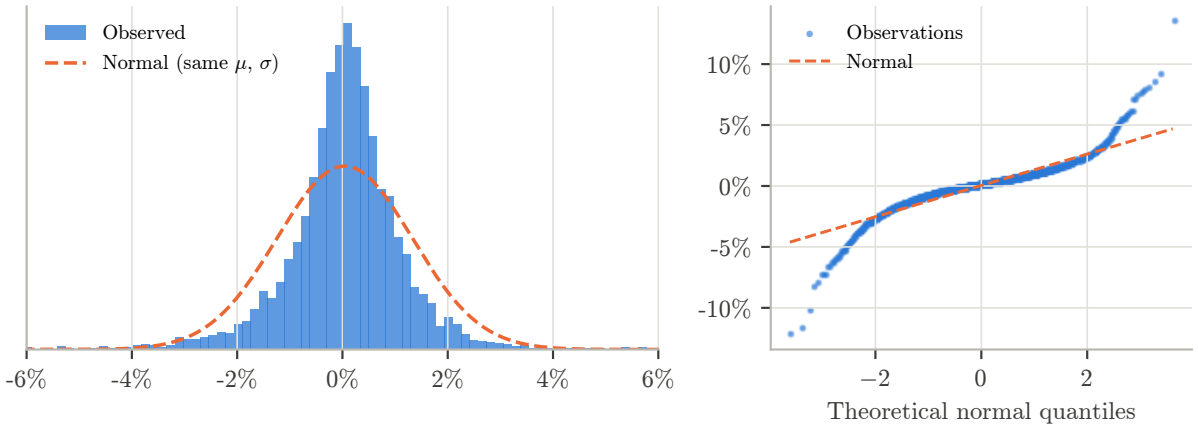


Figure 5. Daily total returns: histogram against a normal density with the same mean and variance (left), and normal Q–Q plot (right).

Table 4. Descriptive statistics of daily returns.

Statistic	Value	Normal benchmark
Mean / standard deviation	0.041% / 1.29%	–
Skewness	−0.30	0
Kurtosis	16.08	3
Jarque–Bera statistic	32,989 ($p < 0.001$)	–
Days beyond ±4 standard deviations	47	0.29 expected
Worst day	−12.1% (15 Oct 2008)	–
Best day	+13.5% (13 Oct 2008)	–

The tails are far fatter than those of a normal distribution: 47 days moved by more than four standard deviations, whereas a normal distribution would predict 0.3 over the period. The negative skewness means that large declines are more pronounced than large gains. Part of this excess kurtosis reflects time-varying volatility: returns are closer to normal once scaled by their conditional volatility, as the GARCH estimates below show.

2.1 Volatility clustering and GARCH estimates

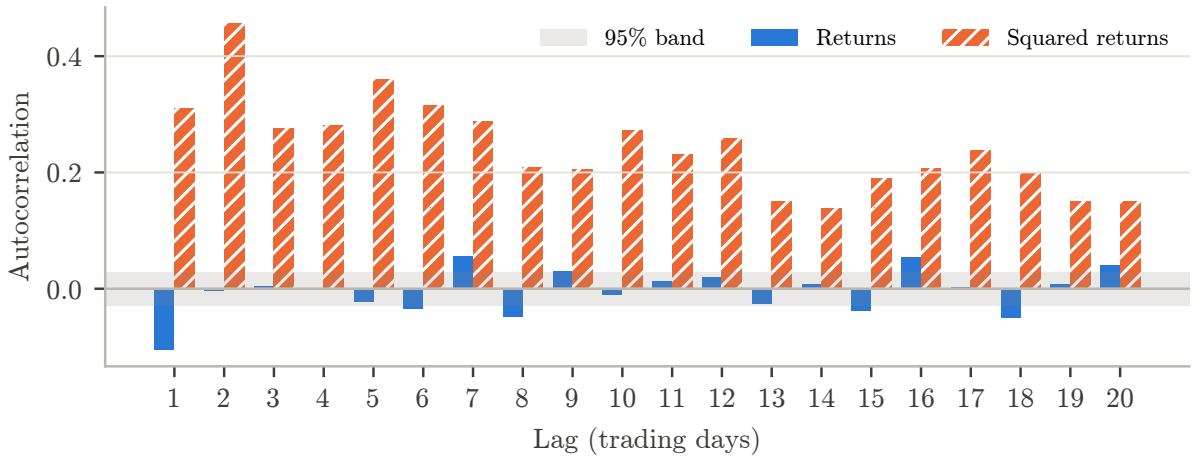


Figure 6. Autocorrelation of daily log returns and squared log returns, lags 1 to 20, with the 95% band under independence.

Returns show little autocorrelation beyond the first lag (Ljung–Box with 10 lags: $Q = 89.7$). Squared returns, by contrast, are strongly and persistently autocorrelated ($Q = 4,305$, $p < 0.001$): a large move tends to be followed by further large moves of either sign.

Table 5. GARCH(1,1) with standardised Student-t innovations, maximum-likelihood estimates (full sample).

Parameter	Estimate	Standard error
μ (% per day)	0.0846	0.0110
ω	0.0199	0.0041
α	0.1345	0.0134
β	0.8583	0.0125
ν (degrees of freedom)	6.1515	0.5502

Standard errors from the inverse of the numerical Hessian of the log-likelihood.

The model $\sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2$ with standardised Student-t innovations is estimated by maximum likelihood on daily log returns in percent (Bollerslev, 1986). Persistence $\alpha + \beta = 0.993$ implies that a volatility shock loses half its effect after about 96 trading days. The estimated $\nu = 6.2$ confirms that, even after accounting for volatility clustering, daily shocks have fatter tails than the normal distribution. The implied long-run volatility, 26.4%, uses daily scaling and is therefore comparable with the daily-scaled figure in Section 1.

2.2 Value at Risk: out-of-sample backtest

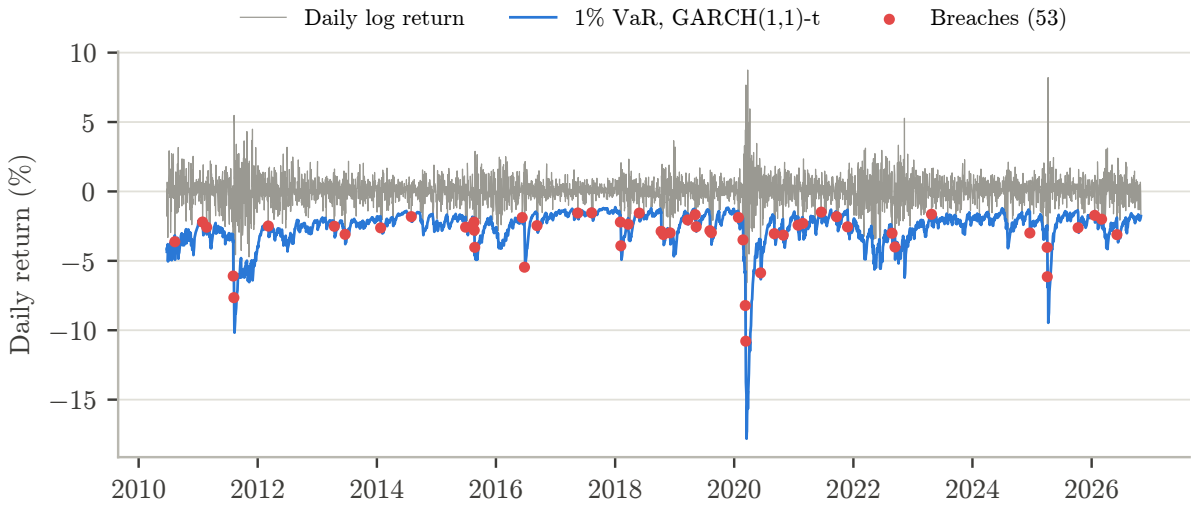


Figure 7. Daily returns and one-day 1% VaR forecast by the GARCH(1,1)-t model; red dots mark breaches.

Table 6. Backtest of one-day VaR forecasts (4,115 days from 23 Jun 2010); p-values of the coverage tests.

Level	Method	Breach rate	Breaches / expected	Kupiec	Christoff.	CC
5%	Historical (250 days)	5.20%	214 / 206	0.558	< 0.001	< 0.001
5%	Normal (250 days)	5.22%	215 / 206	0.511	< 0.001	< 0.001
5%	GARCH(1,1)-t	5.30%	218 / 206	0.385	0.877	0.678
1%	Historical (250 days)	1.36%	56 / 41	0.027	< 0.001	< 0.001
1%	Normal (250 days)	2.38%	98 / 41	< 0.001	< 0.001	< 0.001
1%	GARCH(1,1)-t	1.29%	53 / 41	0.075	0.005	0.004

One-day VaR on daily log returns, forecasts from 23 Jun 2010 to 30 Oct 2026 (4,115 days), each using only information available the day before. Historical and normal VaR use a 250-day rolling window. GARCH parameters are re-estimated at the start of each calendar year on all prior data, and the conditional variance is updated daily. Kupiec: unconditional coverage; Christoffersen: independence; CC: conditional coverage (joint).

Rolling historical and normal VaR react slowly to changes in volatility. Their breach rates at 5% are close to the target, but the breaches cluster in turbulent periods, so the independence test of Christoffersen (1998) rejects them decisively. At 1%, both also fail the unconditional coverage test of Kupiec (1995): 1.36% and 2.38% of breaches instead of 1%. The GARCH(1,1)-t forecasts, from a model estimated with a zero mean, pass both tests at 5% (breach rate 5.30%; p-values 0.385 and 0.877). At 1%, their breach rate of 1.29% is acceptable ($p = 0.075$), but extreme losses still arrive in clusters ($p = 0.005$ for independence). A symmetric t distribution cannot reproduce the negative skewness of large losses; a skewed distribution or an extreme-value model would be the natural next step.

Table 7. GARCH(1,1)-t VaR backtest under three treatments of the conditional mean.

Level	Mean specification	Breach rate	Breaches / exp.	Kupiec	Christoff.	CC
5%	Zero mean, estimated as such (main)	5.30%	218 / 206	0.385	0.877	0.678
5%	ML mean kept	6.08%	250 / 206	0.002	0.814	0.009
5%	ML parameters, mean set to zero afterwards	5.32%	219 / 206	0.348	0.850	0.632
1%	Zero mean, estimated as such (main)	1.29%	53 / 41	0.075	0.005	0.004
1%	ML mean kept	1.48%	61 / 41	0.004	0.014	< 0.001
1%	ML parameters, mean set to zero afterwards	1.26%	52 / 41	0.102	0.004	0.004

The specification of the conditional mean matters. With a symmetric Student-t distribution, the maximum-likelihood location estimate (0.085% a day) lies well above the sample mean (0.033%), because negative skewness pulls the mean below the centre of the distribution. Keeping that estimate shifts the VaR upwards and raises the breach rates to 6.08% and 1.48%, which the coverage tests reject. The main specification therefore estimates the model with the mean fixed at zero, the usual convention at a one-day horizon. Setting the mean to zero only after estimation gives almost the same breaches, so the conclusion rests on the treatment of the mean, not on how it is imposed.

2.3 Variance forecasts: out-of-sample comparison

Table 8. Out-of-sample variance forecasts over the backtest period. Lower losses are better; Diebold–Mariano test of QLIKE against EWMA.

Horizon	Model	QLIKE	MSE	Rank	DM stat.	p-value
1 day	Historical variance (250 days)	1.087	20.6	3	3.87	< 0.001
1 day	EWMA ($\lambda = 0.94$)	0.819	17.4	2	–	–
1 day	GARCH(1,1)-t	0.795	16.2	1	-2.20	0.028
21 days	Historical variance (250 days)	4.312	2,843.1	3	1.92	0.055
21 days	EWMA ($\lambda = 0.94$)	4.123	2,936.3	2	–	–
21 days	GARCH(1,1)-t	4.050	3,318.8	1	-1.47	0.141

Variances in squared percent. Proxy: squared daily return (1 day) and realised variance, the sum of the next 21 squared daily returns (21 days). GARCH: zero-mean model re-estimated each year, multi-step forecasts from its persistence and long-run variance; EWMA and historical variance are held flat over the horizon. Rank by QLIKE. Diebold–Mariano statistics use a Newey–West variance (5 and 25 lags); a negative value favours the model over EWMA.

A good VaR backtest does not show that a model forecasts variance well, so the three variance forecasts are also compared directly, out of sample, over the same period. QLIKE, $\log h_t + \hat{\sigma}_t^2 / h_t$, and the mean squared error are both robust to noise in the variance proxy (Patton, 2011); QLIKE penalises under-prediction of risk more heavily and is less dominated by a few extreme days. By QLIKE, GARCH ranks first at both horizons, but the Diebold–Mariano test (Diebold and Mariano, 1995) against EWMA gives $p = 0.028$ at one day and $p = 0.141$ at 21 days: the improvement over a model with a single fixed parameter is not statistically significant. At one day, EWMA clearly beats the equally weighted 250-day variance ($p < 0.001$), and GARCH has a lower loss still. By MSE, the ranking changes at 21 days (the 250-day historical variance first), because squared errors are dominated by the few crisis months. The richer model helps describe volatility, but its

forecasting advantage is modest.

3 Market efficiency

3.1 First-order autocorrelation

Table 9. First-order autocorrelation of log returns.

Frequency	**Obs.**	$\hat{\rho}_1$	**Naive** t	**Robust** z	p**-value**	**Holm** p
Daily	4,615	−0.105	−7.16	−2.98	0.003	0.029
Weekly	957	−0.023	−0.72	−0.34	0.734	1.000
Monthly	220	0.033	0.49	0.33	0.742	1.000

Naive $t = \hat{\rho}_1\sqrt{n}$ assumes independent, identically distributed returns. The robust z uses the heteroskedasticity-consistent standard error $\sqrt{\sum e_t^2 e_{t-1}^2} \,/\, \sum e_t^2$; p-values refer to the robust statistic. Holm: adjusted for the 10 tests of this section.

Daily autocorrelation is negative (−0.105): a gain tends to be slightly reversed the following day. The naive t-statistic overstates its significance; the heteroskedasticity-robust statistic is −2.98, still significant after the Holm correction. The effect is concentrated in crises: excluding 2008–2009 and February to June 2020, the autocorrelation falls to −0.044. Plausible explanations, which these data cannot separate, are overreaction on turbulent days and temporary gaps between the ETF's closing price and the value of its foreign holdings, which trade at different hours. Testing the second would require daily net asset values. At weekly and monthly frequencies, no autocorrelation is detectable.

3.2 Variance ratio test

Under the random-walk hypothesis, the variance of q-period returns is q times the one-period variance, so

$$VR(q) = \frac{\operatorname{Var}\left(r_t + r_{t-1} + \cdots + r_{t-q+1}\right)}{q \operatorname{Var}(r_t)} = 1.$$

A ratio below 1 signals mean reversion, a ratio above 1 persistence. Following Lo and MacKinlay (1988), the statistic $z^*(q)$ is robust to heteroskedasticity, which is essential given the volatility clustering documented above.

Table 10. Lo–MacKinlay variance ratio test on daily log total returns.

Horizon q **(days)**	$VR(q)$	$z^*(q)$	p**-value**	**Holm** p	**Conclusion (Holm)**
2	0.895	−2.97	0.003	0.029	mean reversion
5	0.830	−2.09	0.037	0.296	not rejected
10	0.778	−1.80	0.072	0.504	not rejected
20	0.760	−1.38	0.167	1.000	not rejected

3.3 Calendar effects

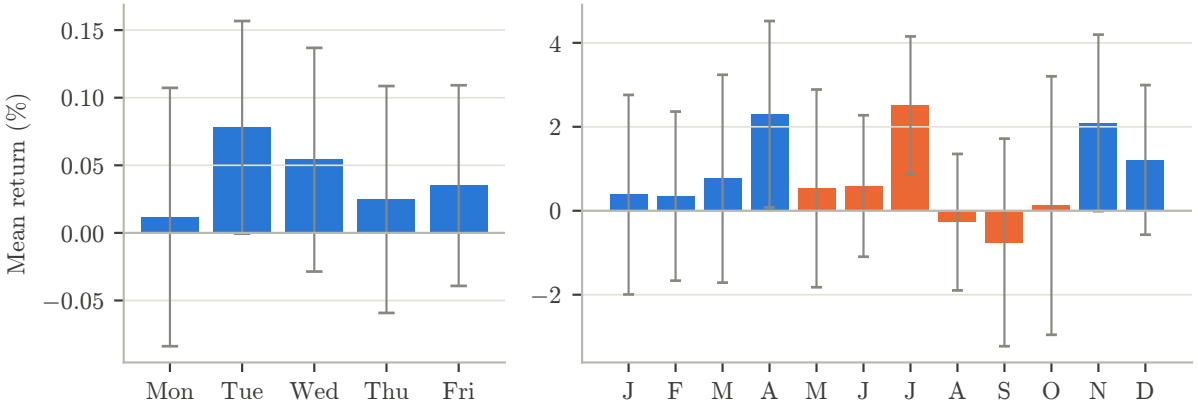


Figure 8. Mean daily return by weekday (left) and mean monthly return by calendar month (right), with 95% confidence intervals. May to October in orange.

Table 11. Tests of calendar effects.

Test	Result	p-value	Holm p
Equal mean returns across weekdays (ANOVA)	$F = 0.38$	0.826	1.000
January effect: January vs other months (Welch)	0.38% vs 0.85% per month	0.714	1.000
"Sell in May": Nov–Apr vs May–Oct, per half-year (18 and 18 half-years)	7.1% vs 4.7%	0.483	1.000

With fewer than twenty observations per calendar month, these tests have little power: the absence of significance does not prove the absence of an effect, only that any effect is too small to detect in eighteen years of data.

3.4 Multiple testing

Section 3 reports 10 hypothesis tests. If all null hypotheses were true and the tests independent, 0.5 false rejections would be expected at the 5% level, and the probability of at least one would be about 40%. Each p-value is also adjusted with the step-down procedure of Holm (1979), which controls the probability of any false rejection. 3 tests are significant before adjustment and 2 after. Both remaining rejections concern the same first-order daily autocorrelation, since $VR(2) \approx 1 + \hat{\rho}_1$.

3.5 Moving-average timing rules

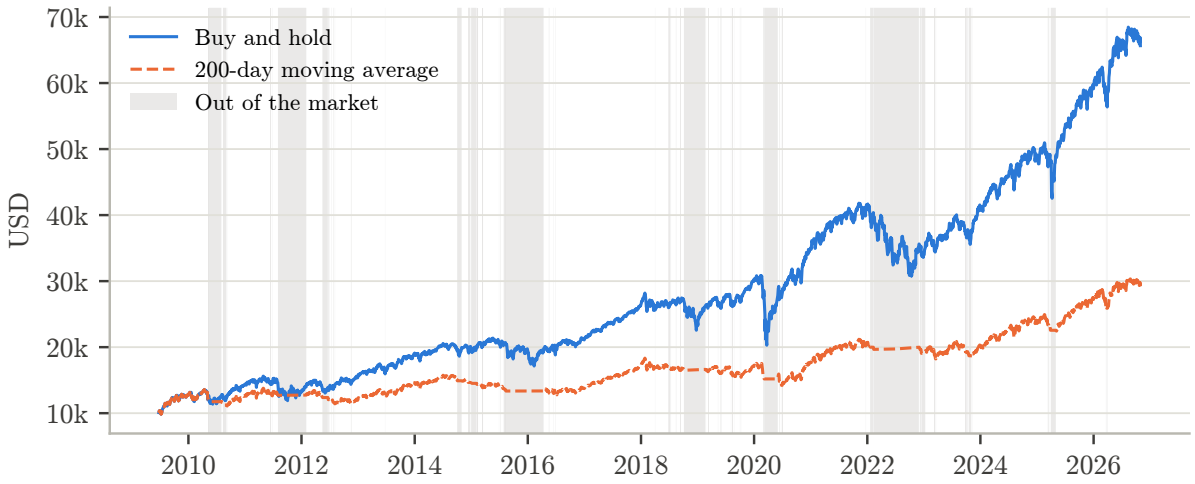


Figure 9. Growth of USD 10,000 from 24 Jun 2009: buy and hold vs the 200-day moving-average rule. Shaded areas: rule out of the market.

Table 12. Buy and hold vs 200-day moving average (Sharpe ratio with 95% confidence interval).

	Buy and hold	**200-day moving average**
Annualised return	11.5%	6.4%
Volatility	14.7%	11.2%
Sharpe ratio	0.71 [0.22, 1.21]	0.48 [0.00, 0.95]
Maximum drawdown	−34.2%	−22.5%
Final value of USD 10,000	USD 65,660	USD 29,231

Table 13. Sensitivity to the moving-average window and to execution (common period). Volatility, Sharpe and drawdown refer to the main execution.

Window	**Ann. return**	**Delayed**	**Volatility**	**Sharpe**	**Max. DD**	**Switches**	**Invested**
50 days	3.6%	4.2%	9.6%	0.26	−25.9%	319	68%
100 days	5.9%	5.3%	10.3%	0.47	−25.2%	203	72%
150 days	5.6%	7.5%	10.7%	0.42	−26.9%	176	74%
200 days	6.4%	7.0%	11.2%	0.48	−22.5%	140	77%
250 days	6.3%	6.3%	11.0%	0.48	−23.1%	141	77%
Buy and hold	11.5%	11.5%	14.7%	0.71	−34.2%	0	100%

Rule: the position earning the return of day t is decided with prices up to the close of day $t-1$, so no future information is used. The main case assumes the trade is executed at that same close (a market-on-close order placed just before the close); the delayed case executes one session later. Out of the market, the rule earns the T-bill rate; each switch costs 0.10%. All windows are compared over the same period, starting 24 Jun 2009, the first day on which a 250-day average is available.

Every window underperformed buy and hold, by 5.1% to 7.9% a year, while reducing the maximum drawdown. For the 200-day rule, a circular block bootstrap (2,000 replications, 63-day blocks) gives a 95% confidence interval of [0.2%, 10.3%] for the annual return gap ($p = 0.051$): the underperformance is borderline significant. The sample starts in mid-2009, after the financial-crisis crash that such rules are designed to avoid, which works against them; a longer index history

would be needed for a fair assessment. Late and repeated signals ("whipsaws") are costly in a market that recovers quickly. Delaying execution by one session changes the results by up to 1.9% a year, in both directions: the ranking of windows is fragile, but the underperformance against buy and hold is not.

4 Holding period and investor behaviour

4.1 Rolling returns and the probability of loss

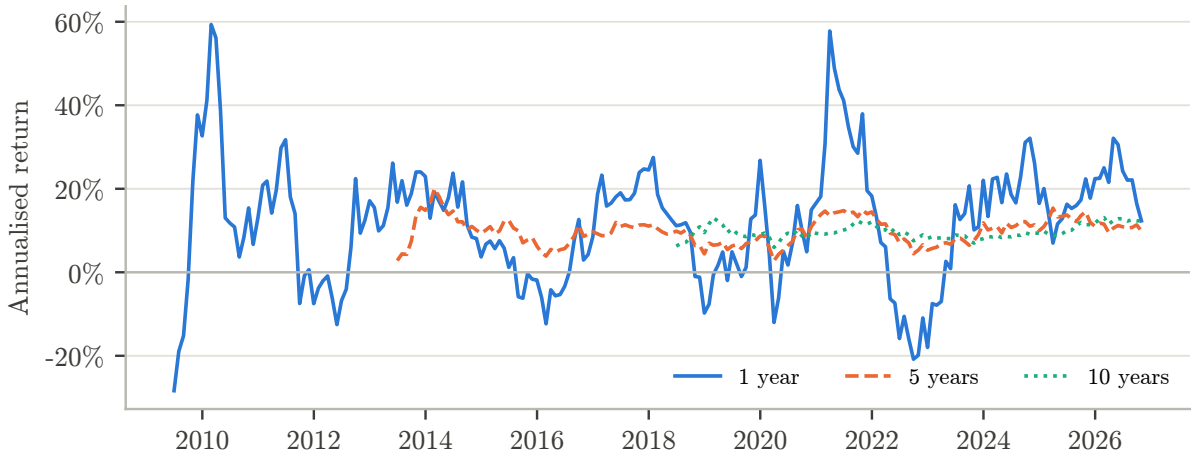


Figure 10. Annualised total return over rolling 1-, 5- and 10-year windows, by end of window.

Table 14. Rolling returns (overlapping monthly windows, annualised) and bootstrap probability of loss.

Horizon	Windows	P(loss)	Bootstrap	Worst (window end)	Median	Best
1 year	209	22%	24%	−28.4% (Jun 2009)	13.1%	59.3%
5 years	161	0%	10%	2.8% (Mar 2020)	10.1%	20.0%
10 years	101	0%	4%	6.0% (Mar 2020)	9.3%	13.1%

Overlapping windows understate the probability of loss at long horizons: the 10-year windows contain only about one and a half independent periods, none of which happened to lose money. A stationary bootstrap (Politis and Romano, 1994) that resamples blocks of monthly returns (mean length 12 months, 20,000 paths) gives a more defensible estimate: 10% over five years and 4% over ten.

These figures are fragile. The Monte Carlo error is negligible (about 0.1 percentage point), but the estimate depends on the assumed block length, which governs how much of the slow-moving dependence in returns is preserved: from 1% to 6% over ten years across the block lengths tested. Above all, no resampling scheme can capture the uncertainty about the true expected return. Treating monthly log returns as independent and normal, the probability of a ten-year loss is 6%; moving the mean return to either end of its 95% confidence interval gives 0.1% to 45%. Eighteen years of data simply do not pin down the long-run probability of loss.

Table 15. Bootstrap probability of loss by mean block length (months).

Block	1 year	5 years	10 years
1 (i.i.d.)	29%	13%	5.9%
6	25%	13%	6.3%
12 (main)	25%	10%	4.0%
24	24%	7%	1.9%
36	25%	5%	1.0%

Table 16. Normal i.i.d. model: probability of loss with the mean at its estimate and at the ends of its 95% interval.

Horizon	Estimate	High μ	Low μ
1 year	31%	17.0%	48%
5 years	13%	1.6%	47%
10 years	6%	0.1%	45%

4.2 Sensitivity to the starting date

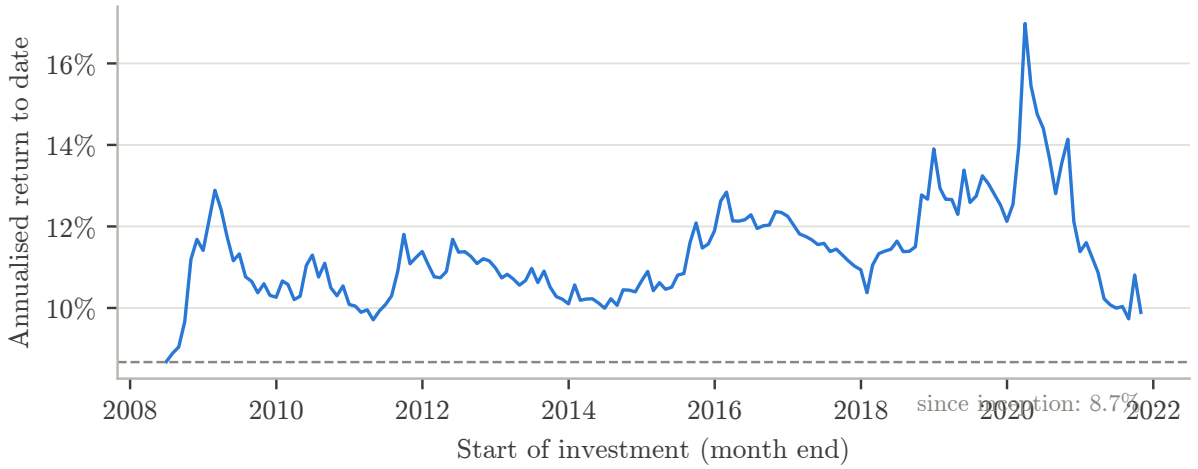


Figure 11. Annualised total return from each month end to the latest close (starting points with at least five years of history).

Because VT was launched just before the 2008 crash, "since inception" figures are among the least favourable available. Starting at any year end between 2008 and 2021 would have delivered between 9.9% (2021) and 13.9% (2018) a year to date, against 8.7% since inception. Performance statistics should always be read together with their start and end dates.

4.3 The cost of missing the best days

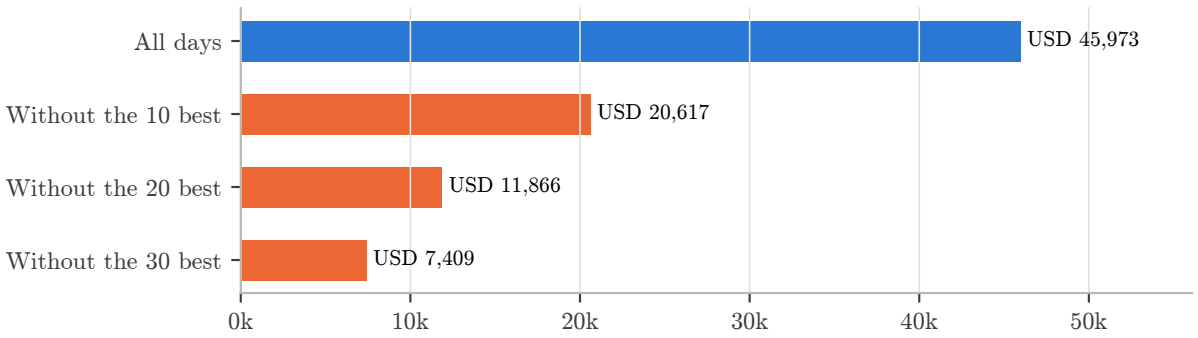


Figure 12. Final value of USD 10,000 invested at inception, with and without the best trading days.

Missing the 10 best days out of 4,615 cuts the annualised return from 8.7% to 4.0%. The reverse exercise is just as striking: avoiding the 10 worst days would have lifted the final value to USD 112,534. 7 of the 10 best days fall within two weeks of one of the 10 worst. They occur in the same crisis periods: avoiding the worst days without missing the best would require perfect timing.

4.4 Monthly savings plan and lump-sum investing

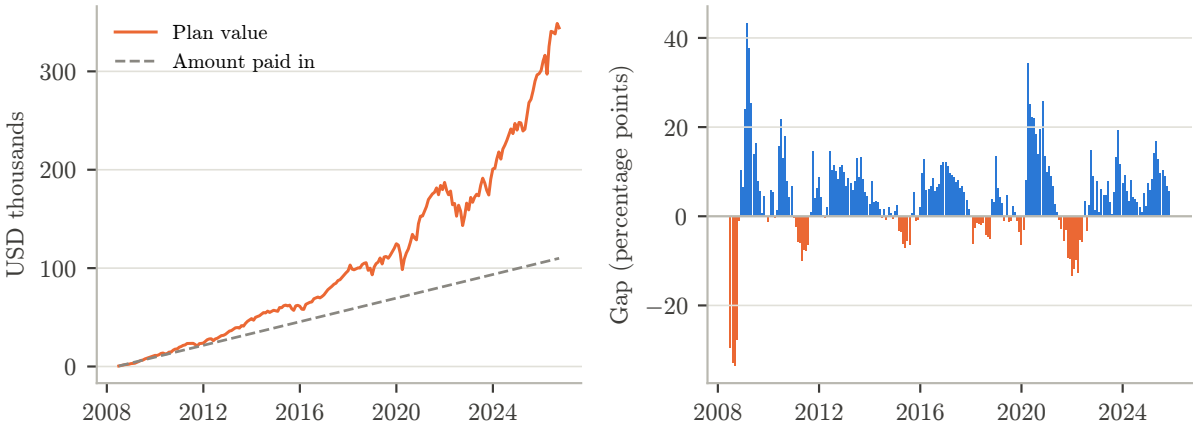


Figure 13. Left: value of a USD 500 monthly plan vs amount paid in. Right: lump sum minus 12-month averaging, value gap after 12 months, by starting month.

Investing USD 500 at the end of every month since inception (220 contributions, USD 110,000 in total) would be worth USD 338,740 today, an internal rate of return of 11.1% a year, above the 8.7% of a single investment at inception because most of the money was invested after the 2008–2009 crash.

Across the 209 possible monthly starting points, investing a sum all at once beat spreading it over 12 months (with the uninvested cash earning the T-bill rate) in 72% of cases, with an average advantage of 4.5% after 12 months. Because the market rises most of the time, waiting costs money on average; averaging in mainly limits the regret of investing just before a decline. The starting points overlap, so this frequency is descriptive rather than a precise probability.

4.5 Optimal leverage, Kelly criterion and volatility drag

Can leverage raise the long-run compounded growth of an investment in VT, and by how much? This subsection derives the growth-optimal exposure, estimates it, tests it out of sample and separates the effects that make leveraged compounding differ from a simple multiple of the fund's return. The aim is not to show that leverage pays, but to identify the assumptions under which it maximises expected growth and to show how fragile that optimum is.

Framework. Suppose the fund's value follows a geometric Brownian motion with arithmetic expected return μ and volatility σ, and an investor holds a constant exposure L to it, rebalanced continuously, with the rest in cash at the rate r_f ($L < 1$) or financed by borrowing ($L > 1$). $L = 1$ is the unleveraged fund, $L = 0.5$ a 50% exposure and $L = 2$ a 200% exposure. If borrowing costs r_f, wealth W grows at the expected logarithmic rate

$$g(L) = \frac{1}{\mathrm{d}t}\,\mathbb{E}\left[\mathrm{d}\log W_t\right] = r_f + L(\mu - r_f) - \tfrac{1}{2}L^2\sigma^2.$$

The arithmetic expected return of the portfolio, $r_f + L(\mu - r_f)$, rises linearly with L, but the expected log growth carries the quadratic penalty $\frac{1}{2}L^2\sigma^2$. Since $g'(L) = (\mu - r_f) - L\sigma^2$ and $g''(L) = -\sigma^2 < 0$, the function is strictly concave and maximised at the Kelly exposure (Kelly, 1956; Breiman, 1961; Merton, 1969)

$$L^* = \frac{\mu - r_f}{\sigma^2}, \qquad g(L^*) = r_f + \frac{(\mu - r_f)^2}{2\sigma^2}.$$

μ is an arithmetic mean: substituting the historical CAGR, which is already a log-growth rate net of the drag $\frac{1}{2}\sigma^2$, would count the drag twice.

Borrowing costs. A retail investor borrows above the rate earned on cash. With a lending rate r_l and a borrowing rate $r_b > r_l$, the growth rate is piecewise:

$$g(L) = \begin{cases} r_l + L(\mu - r_l) - \frac{1}{2}L^2\sigma^2, & 0 \le L \le 1, \\ r_b + L(\mu - r_b) - \frac{1}{2}L^2\sigma^2, & L > 1. \end{cases}$$

It remains concave, with a kink at $L = 1$. The optimum is $(\mu - r_l)/\sigma^2$ if that is below one, $(\mu - r_b)/\sigma^2$ if that is above one, and exactly one otherwise: leverage is worth taking only if the marginal expected return exceeds the marginal borrowing cost by more than the marginal variance penalty. Margin rates are not available in the data, so the borrowing rate is set to the T-bill yield plus a spread of 1.5% a year, comparable to the lower end of retail margin rates; Table 21 reports spreads from 0 to 3%. The T-bill yield itself is a lending rate, not an observable margin rate.

Simulation. The strategies are simulated on daily simple total returns r_t from the adjusted prices, so dividends are included and fund expenses, already in the price, are not charged again. For an exposure L decided at the close of day $t - 1$,

$$W_t = W_{t-1}\left(1 + R_{p,t} - c_t\right), \qquad R_{p,t} = \begin{cases} Lr_t + (1 - L)\, r_{l,t}, & 0 \le L \le 1, \\ Lr_t - (L - 1)\, r_{b,t}, & L > 1, \end{cases}$$

where $r_{l,t}$ is the T-bill yield known at the previous close divided by 252, $r_{b,t} = r_{l,t} + \text{spread}/252$, and c_t is a cost of 2 bp per unit of exposure traded when the position is rebalanced to its target at the close. With $L = 1$ nothing is traded and the simulation reproduces buy-and-hold exactly. If $1 + R_{p,t} - c_t \le 0$, the strategy is insolvent and its wealth stays at zero. Five numerical checks are

run on every update: $L = 1$ reproduces the fund's total-return path, $L = 0$ reproduces the cash account, the closed-form L^* matches the numerical maximum of g, the exact decomposition below reconciles with simulated wealth to machine precision, and wealth cannot recover after insolvency.

Table 17. Kelly inputs and leverage estimates (full sample).

Input	**Estimate**
Arithmetic mean return μ (daily mean $\times 252$)	10.4%
Cash rate r_f (13-week T-bill, average)	1.4%
Expected excess return $\mu - r_f$ (standard error)	9.0% (4.8%)
Volatility σ, daily returns $\times\sqrt{252}$	20.4%
Volatility σ, monthly returns $\times\sqrt{12}$	16.6%
Borrowing rate (assumption)	T-bill + 1.5%
Exposure	
Theoretical Kelly, borrowing at the T-bill rate, daily variance (In-sample, hindsight)	2.17x
Theoretical Kelly, borrowing at the T-bill rate, monthly variance (In-sample, hindsight)	3.05x
Theoretical Kelly, borrowing spread 1.5% (In-sample, hindsight)	1.81x
Ex-post log-growth maximiser, frictionless (Exact daily recursion)	2.13x
Ex-post log-growth maximiser, spread and 2 bp costs (Exact daily recursion)	1.76x
Half Kelly / quarter Kelly (from the spread case) (Fractional)	0.90x / 0.45x
Feasible under Regulation T initial margin (Constraint)	$\leq$ 2.00x

With daily rebalancing, the variance that governs compounding is the variance of daily returns. Using it, the frictionless Kelly exposure is 2.17x; with the assumed borrowing spread it falls to 1.81x. The exact daily recursion, which needs neither the diffusion approximation nor normality, is maximised at 2.13x without frictions and at 1.76x with the spread and rebalancing costs, close to the theoretical values. The monthly variance gives 3.05x: because daily returns are negatively autocorrelated (Section 3), the Kelly exposure depends on the rebalancing frequency it is meant for. All these figures are computed on the full sample and are therefore in-sample and hindsight-calibrated: they describe what would have maximised growth over this particular history, not an achievable ex-ante optimum.

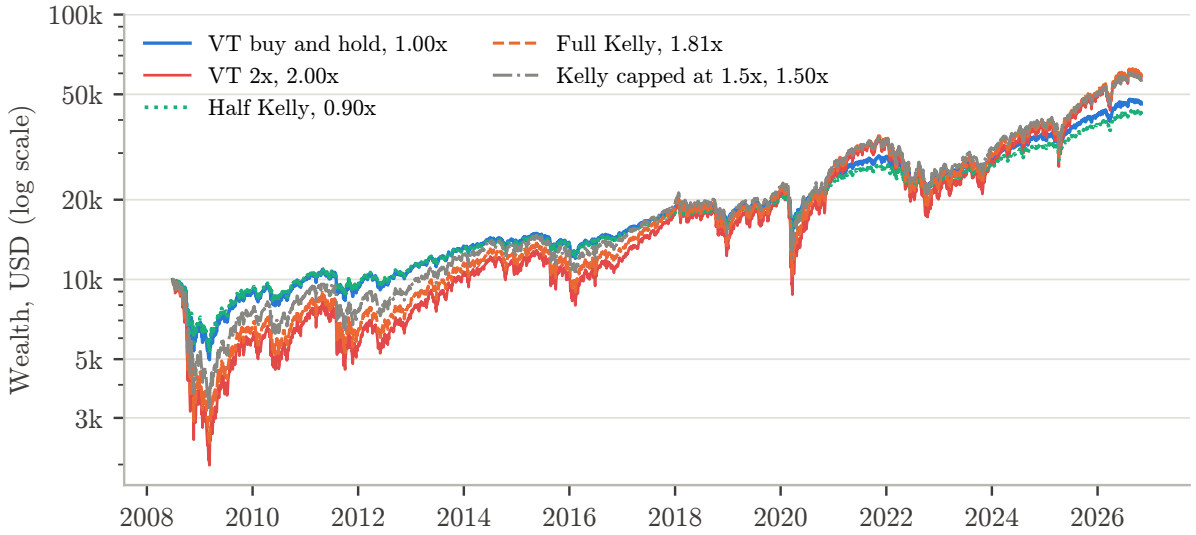


Figure 14. Historical simulation, not a forecast: growth of USD 10,000 under constant exposures rebalanced daily, borrowing at the T-bill rate plus 1.5%, 2 bp rebalancing costs. Kelly exposures are estimated on the full sample (hindsight).

Table 18. Constant-exposure strategies, full sample, daily rebalancing, borrowing spread 1.5%. Volatility and Sharpe ratio from monthly returns, as in Section 1.

Strategy	L	CAGR	Vol.	Sharpe	Max. DD	Final (USD)	Margin days
VT buy and hold	1.00x	8.7%	16.5%	0.50	−50.3%	45,973	0
VT 2x	2.00x	9.9%	33.4%	0.42	−80.1%	56,439	0
Quarter Kelly	0.45x	5.2%	7.4%	0.53	−25.1%	25,142	0
Half Kelly	0.90x	8.1%	14.9%	0.51	−46.3%	42,013	0
Full Kelly	1.81x	10.0%	30.1%	0.43	−75.9%	57,749	0
Kelly capped at 1.5x	1.50x	9.9%	24.9%	0.45	−67.7%	56,295	0
VT 3x	3.00x	6.3%	50.4%	0.37	−93.6%	30,659	2
VT 4x	4.00x	−1.9%	67.0%	0.34	−98.4%	7,056	2139

Leverage raised the historical CAGR only up to a point. Full Kelly (1.81x) and 2x ended with about 58k and 56k against 46k for buy and hold, a gain of roughly 1.4% a year, bought with maximum drawdowns of −76% and −80% instead of −50%. Their Sharpe ratios are lower than buy and hold, not higher: leverage financed above the cash rate scales risk more than excess return. At 3x, the CAGR drops to 6.3%, and at 4x to −1.9% with a −98% drawdown: beyond the optimum, more leverage lowers long-run growth. Quarter and half Kelly give up most of the growth gain but cut drawdowns sharply, the usual argument for fractional Kelly. The four objectives that appear in this table, log growth, Sharpe ratio, maximum drawdown and terminal wealth in one realised path, are different and point to different exposures.

The simulation is frictionless in one important respect: it assumes the target exposure can always be restored at the close. In practice, a US margin account is limited by Regulation T to an initial loan of 50% of the purchase price (2x), and FINRA Rule 4210 requires equity of at least 25% of the position's value, with many brokers demanding more. The last column counts the days on which equity fell below 25% of assets before rebalancing, a proxy for a margin call. A 3x position, which starts with 33% equity, breaches the threshold on 2 days (15 Oct 2008, 16 Mar 2020), when VT fell by more than 11% in a single session; a 4x position starts at the 25% limit and breaches it on almost every down day. A forced sale at such moments would lock in the loss and break the constant-leverage path assumed here.

Volatility drag and the compounding gap. Three different effects separate a leveraged position from "L times the fund", in addition to trading costs. To isolate the effect of leverage on compounding, compare the frictionless daily-reset position financed at a rate b_t with the geometrically scaled reference $\log W_T^{\mathrm{ref}} = L\sum_t \log(1+r_t) - (L-1)\sum_t \log(1+b_t)$. The difference is

$$\begin{aligned} D_T(L) &= \sum_{t=1}^{T} \Big[\log\big(1 + Lr_t - (L-1)b_t\big) - L\log(1+r_t) + (L-1)\log(1+b_t) \Big] \\ &\approx -\tfrac{1}{2} L(L-1) \sum_{t=1}^{T} (r_t - b_t)^2, \end{aligned}$$

where the approximation uses second-order expansions of the logarithms and requires every wealth factor to be positive (Cheng and Madhavan, 2009; Avellaneda and Zhang, 2010). The gap is zero for $L = 0$ and $L = 1$, negative for $L > 1$, and grows with realised variance and with $L(L-1)$.

Table 19. Exact decomposition of annualised log growth, full sample, spread 1.5%, 2 bp costs. Components add up to the total.

L	$L\times$ log growth of VT	T-bill financing	Borrowing spread	Compounding gap	Trading costs	Total log growth
1.00x	8.3%	0.0%	0.0%	0.0%	0.00%	8.3%
1.50x	12.5%	−0.7%	−0.7%	−1.6%	−0.03%	9.4%
2.00x	16.6%	−1.4%	−1.5%	−4.2%	−0.08%	9.4%
3.00x	24.9%	−2.8%	−3.0%	−12.8%	−0.25%	6.1%

Over the full sample, VT's arithmetic mean return was 10.4% a year, but its log growth only 8.3%: the difference, 2.1%, is the ordinary volatility drag that affects any risky investment, close to $\frac{1}{2}\sigma^2 = 2.1\%$. Leverage adds a second, distinct effect. Table 19 decomposes the realised log growth of each strategy exactly into five parts: L times the log growth of VT; financing at the T-bill rate; the borrowing spread; the compounding gap $D_T(L)$; and trading costs. At 2x, the compounding gap costs 4.2% a year, more than the T-bill financing (1.4%) or the spread (1.5%), and costs are negligible. At 3x the gap grows to 12.8%, roughly in proportion to $L(L-1)$ as the approximation predicts.

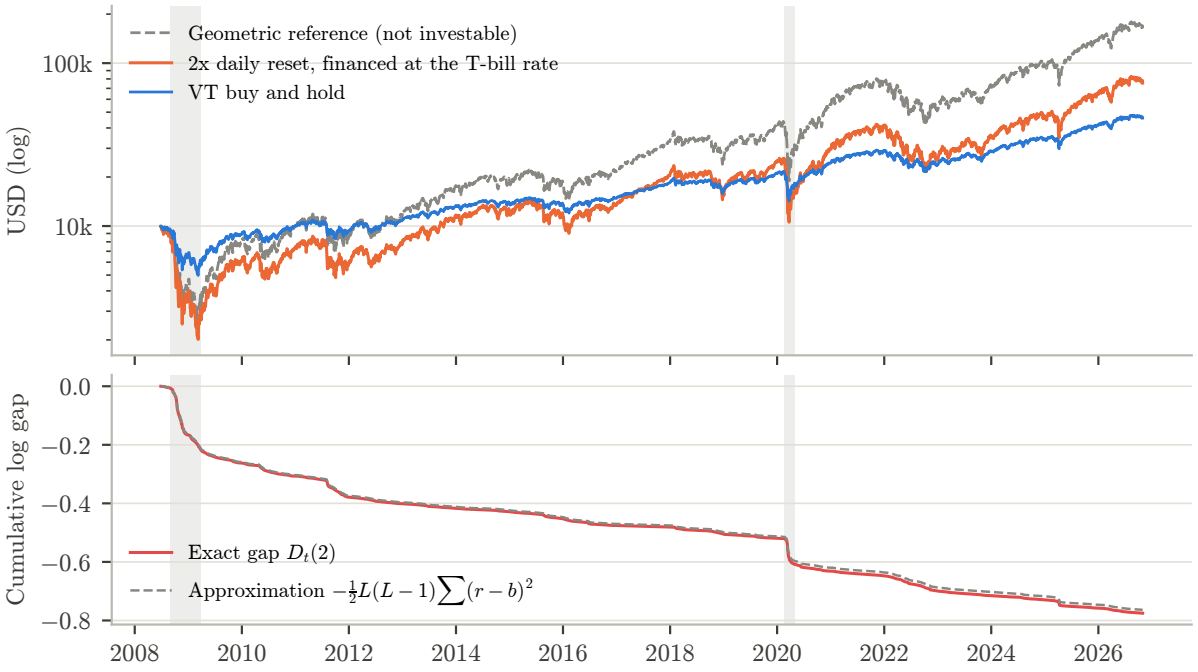


Figure 15. Historical simulation: a frictionless 2x daily-reset position financed at the T-bill rate, its geometrically scaled reference and VT (top); cumulative exact compounding gap and its approximation (bottom). Shaded: September 2008 to March 2009 and February to April 2020.

The gap is path-dependent. If VT gains 10% and then loses 9.09%, it is back where it started; a daily-reset 2x position gains 20% and then loses 18.18%, ending 1.8% lower ($1.2 \times 0.8182 = 0.982$). Over the sample, the exact gap at 2x reached -0.775 in log terms (approximation -0.764), so the frictionless 2x position ended at USD 75,426 against USD 163,737 for the geometric reference. Most of the gap accumulated in the 2008–2009 and 2020 crises, when daily variance was highest. The reference is a decomposition device, not an investable strategy. Compared with another common benchmark, twice the cumulative simple return of VT (USD 81,945), the daily-reset position also ended lower in this sample, but that comparison has no fixed sign: in a steady trend

with low volatility, daily resetting compounds gains and can beat the simple multiple.

Growth as a function of leverage. Figure 16 plots the model growth curve and the growth actually realised by each constant exposure. The curves are concave: below the optimum, more leverage adds growth; beyond it, the quadratic penalty dominates and growth falls, turning negative around 4x. The realised curve follows the model closely because both use the same sample, which is precisely why its maximum is not a guide to the future. The shaded band shows how far the model curve moves when μ is set to either end of its 95% confidence interval: the uncertainty grows with L, and at 2x the band is 37 percentage points of annual growth wide.

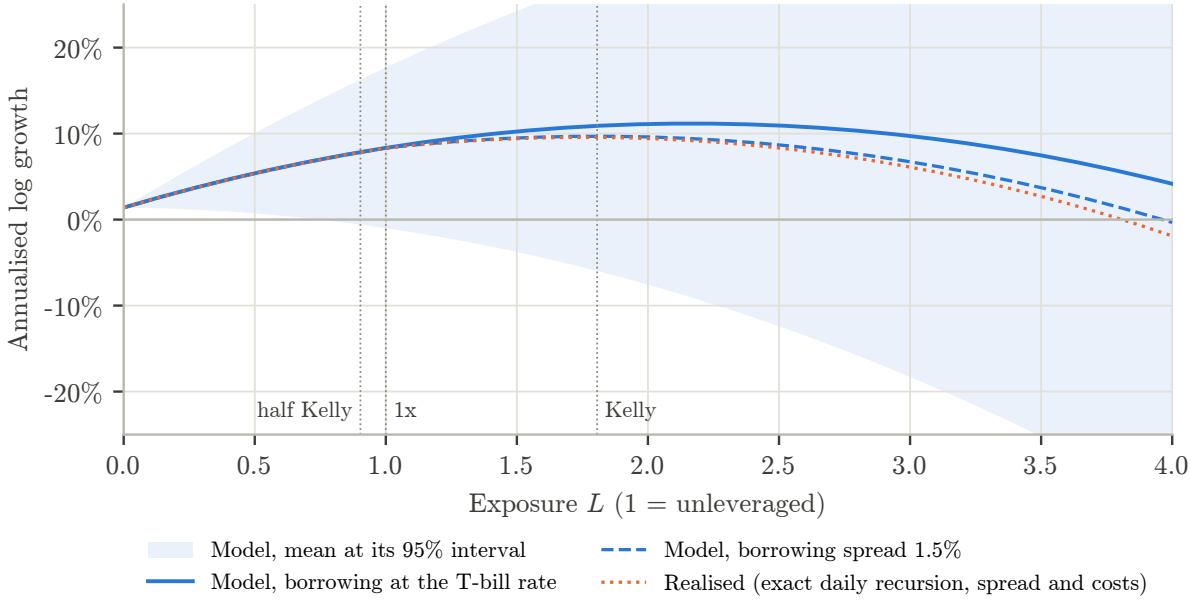


Figure 16. Model-implied expected log growth $g(L)$ (blue, with and without the borrowing spread; band: mean return at the ends of its 95% interval) and realised log growth of the daily simulations (orange, in-sample). Vertical lines: 1x, half Kelly and Kelly with the spread.

Estimation uncertainty. Kelly leverage inherits the imprecision of the expected return. With $\hat{\mu} - r_f = 9.0\%$ estimated with a standard error of 4.8%, the approximate standard error of L^* is $1/(\sigma\sqrt{T}) \approx 1.14$. The block bootstrap confirms it: with 21-day blocks, the 95% interval runs from 0.24x to 4.64x, the probability that the true optimum is below 1x is 13%, and the results hardly depend on the block length. Table 21 shows that plausible changes to the expected return, the variance or the period move the estimate by about one unit of leverage or more: an excess return of 3% instead of the historical 9.0% gives 0.72x; the GARCH long-run variance gives 1.29x; and the five-year sub-periods range from 0.69x to 6.99x.

Table 20. Circular block bootstrap of the frictionless Kelly exposure (2,000 resamples of daily excess returns).

Block length	Median	95% interval	$\mathbf{P}(L^* < 1)$	$\mathbf{P}(L^* > 2)$
1 (i.i.d.)	2.15x	[−0.02x, 4.33x]	15%	56%
21 days	2.17x	[0.24x, 4.64x]	13%	56%
63 days	2.23x	[0.15x, 4.92x]	14%	57%
252 days	2.20x	[0.26x, 5.34x]	14%	56%

Table 21. Kelly exposure under alternative inputs, one input changed at a time (borrowing at the T-bill rate unless stated).

Input	Value	L^*
Expected excess return	Historical estimate	2.17x
Expected excess return	Excess return 5%	1.20x
Expected excess return	Excess return 3%	0.72x
Variance estimate	Daily returns, $\sigma = 20.4\%$	2.17x
Variance estimate	Monthly returns, $\sigma = 16.6\%$	3.29x
Variance estimate	GARCH long run, $\sigma = 26.4\%$	1.29x
Borrowing spread	0.0%	2.17x
Borrowing spread	0.5%	2.05x
Borrowing spread	1.5%	1.81x
Borrowing spread	3.0%	1.45x
Sub-period	2008–2012	0.69x
Sub-period	2013–2017	6.99x
Sub-period	2018–2022	1.42x
Sub-period	2023–2026	6.92x

Out-of-sample test. To remove hindsight, Kelly exposure is re-estimated at every month end from 30 Jun 2011, after a 3-year training period, using only data available at that date: the expanding-window mean excess return, the expanding-window daily variance and the current T-bill yield plus the spread. The estimate is held for the next month and rebalanced daily. It ranged from 0.03x to 1.86x, starting at 0.64x because the training period was dominated by the 2008 crash; caps at 1.5x or 2x would never have bound. Out of sample, estimated full Kelly earned 9.6% a year against 10.1% for buy and hold, with higher volatility and a lower Sharpe ratio; the leverage that looks optimal in hindsight was not identifiable in real time. Fixed 2x did better over this particular period, which began after the crash, but choosing it also relies on hindsight.

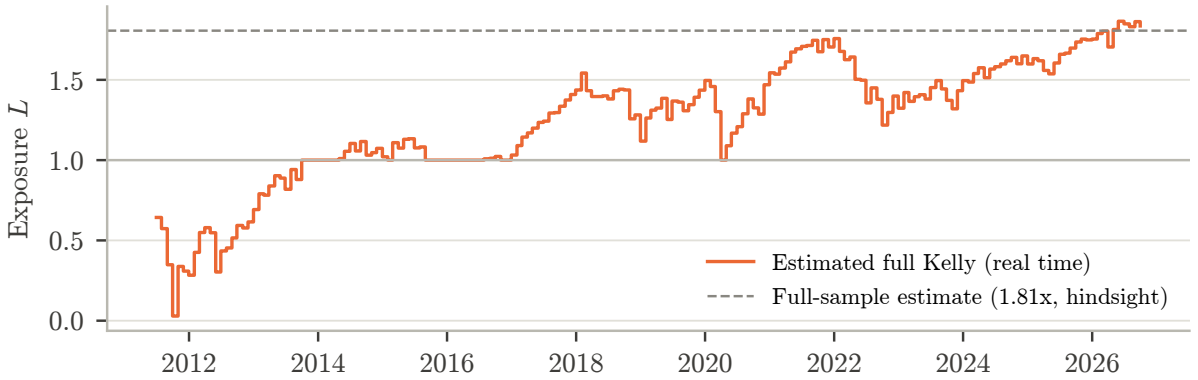


Figure 17. Kelly exposure estimated in real time at each month end (expanding window, borrowing spread 1.5%), compared with the full-sample estimate available only in hindsight.

Table 22. Out-of-sample comparison from 01 Jul 2011 (common period, daily rebalancing, spread 1.5%, 2 bp costs).

Strategy	L: mean (range)	CAGR	Vol.	Sharpe	Max. DD	Final (USD)
VT buy and hold	1.00x	10.1%	14.4%	0.63	−34.2%	43,610
VT 2x	2.00x	13.9%	29.1%	0.54	−59.6%	73,973
Full-sample Kelly, 1.81x (hindsight)	1.81x	13.5%	26.2%	0.56	−55.3%	69,295
Estimated full Kelly (real time)	1.25x (0.03x–1.86x)	9.6%	18.1%	0.51	−44.0%	40,683
Estimated half Kelly (real time)	0.63x (0.01x–0.93x)	6.4%	9.0%	0.56	−24.0%	25,832

Interpretation. Historical VT returns imply a growth-optimal exposure of about 1.81x with realistic borrowing costs, but this number is an in-sample artefact of one favourable eighteen-year path rather than a parameter that can be relied on. Its 95% interval spans almost the whole range from no leverage to 4x, it moves by about one unit of leverage or more with plausible changes to the expected return, the variance or the period, and a real-time estimate would not have found it. The features documented earlier in this report work against aggressive leverage: fat tails and volatility clustering (Section 2) mean that losses arrive in bursts, exactly when the compounding gap widens and margin calls occur; the 50% drawdown of 2008–2009 becomes 76% at full Kelly and 80% at 2x; and the confidence interval of the Sharpe ratio (Section 1) is the same uncertainty that makes L^* unknowable. Because the growth curve is flat near its maximum and falls steeply beyond it, overestimating L^* is far more costly than underestimating it. This asymmetry is the case for fractional Kelly (Thorp, 2006; MacLean, Thorp and Ziemba, 2011): under the model, half Kelly keeps three quarters of the growth gain while halving the volatility, and at the point estimate it does not even require borrowing.

5 A Swiss investor's perspective

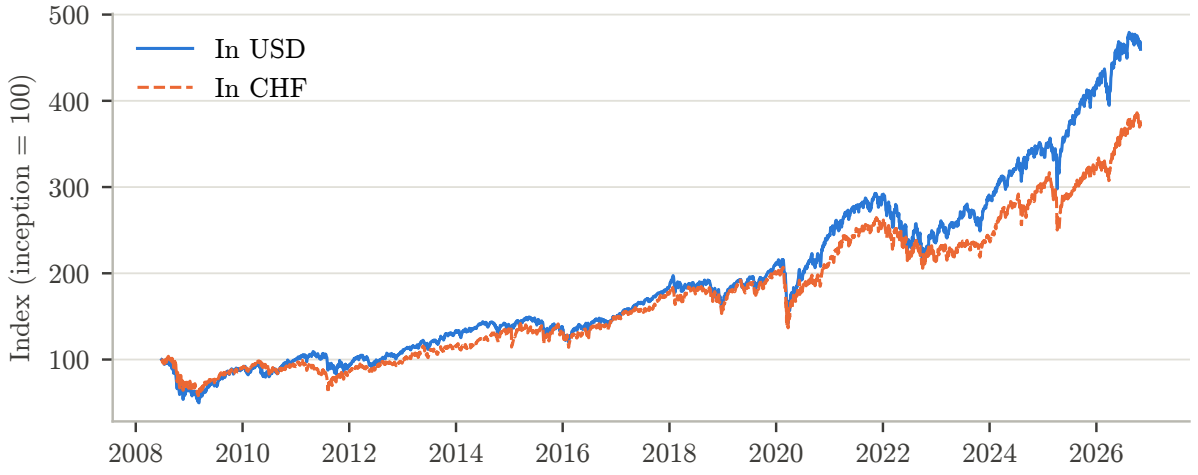


Figure 18. Total return in USD and in CHF, rebased to 100 at inception.

Table 23. VT in US dollars and in Swiss francs.

	In USD	In CHF
Annualised total return	8.7%	7.4%
Annualised volatility (monthly returns)	16.5%	14.9%
USD/CHF over the period	–	1.0237 → 0.8249 (−1.2% a year)
Correlation of VT with USD/CHF (weekly / daily)	–	−0.24 / −0.05

The dollar depreciated against the franc, costing an unhedged Swiss investor about 1.3% of return a year. The weekly correlation between VT and USD/CHF is −0.24; the daily correlation, −0.05, is biased towards zero because Yahoo's daily currency and ETF closes are taken at different times of day. The dollar tends to strengthen against the franc when equities fall, which cushions part of the losses in CHF. Because VT is a global fund, its true exposure is to a basket of currencies: USD/CHF captures only part of the currency effect. All figures are before Swiss income tax and the 15% US withholding tax on dividends, which a Swiss resident can usually credit against Swiss tax.

6 Data validation

Table 24. Total returns from Yahoo adjusted prices vs Vanguard's published returns, to 30 September 2026.

Period	Vanguard, market price	This report	Gap (pp)	Vanguard, NAV
Third quarter 2026	0.83%	0.83%	+0.00	0.76%
Year to date	12.86%	12.86%	−0.00	12.83%
1 year	16.45%	16.45%	+0.00	16.32%
3 years	21.46%	21.45%	−0.01	21.42%
5 years	11.36%	11.35%	−0.01	11.34%
10 years	12.32%	12.32%	−0.00	12.32%
Since inception	8.78%	8.81%	+0.03	8.78%

Periods of one year or more are annualised. Source: Vanguard fact sheet F3141 (data as of 30 September 2026), stored in `data/reference/`.

Yahoo's adjusted prices are not an official source, so the total returns used in this report were checked against the market-price returns published by Vanguard. For all standard periods the two agree to within 0.01 percentage point. The small gap since inception arises because Vanguard measures from the launch on 24 June 2008, two trading days before the first Yahoo closing price. The ETF's market price and its net asset value (NAV) also track each other closely.

7 Limitations

- **One asset, one period.** All results come from a single fund over eighteen years that include one global financial crisis, one pandemic crash and a long bull market. Estimates of mean returns and Sharpe ratios are imprecise over such horizons, as the confidence intervals show.
- **Start date.** VT was launched just before the 2008 crash, which depresses every "since inception" figure and works against timing rules that would have protected investors during that crash.
- **Data source.** Yahoo Finance is not an official source, and its adjusted prices can be revised retroactively. The validation against Vanguard covers a single reference date.
- **Market price vs NAV.** The analysis uses closing market prices. The daily autocorrelation may partly reflect gaps between the market price and the net asset value, which cannot be tested without daily NAV data.
- **Taxes and costs.** Returns are gross of investor-level taxes and trading costs (fund expenses are already reflected in the price). Dividends are assumed to be reinvested at the closing price on the ex-dividend date.
- **Leverage simulations.** Section 4.5 assumes the target exposure is restored at each close at a small cost, ignores intraday moves, forced liquidations and taxes, and uses an assumed borrowing spread rather than observed margin rates. Its results are historical simulations, not achievable track records.
- **Model choices.** GARCH(1,1) with symmetric Student-t errors does not capture the asymmetric response of volatility to losses (leverage effect) nor the skewness of large losses. Bootstrap results assume that the dependence structure of the past repeats.

8 Conclusion

The evidence describes a broadly diversified equity fund that has delivered the equity risk premium with the usual features of equity returns: fat tails, volatility that clusters and persists, and deep but temporary drawdowns.

Three statements about efficiency need to be kept apart. Statistically, the random walk is rejected only for a small negative daily autocorrelation; failing to reject it at longer horizons does not prove that prices are efficient, because the tests have limited power with eighteen years of data. In terms of predictability, the effect explains about 1.1% of the variance of daily returns and is concentrated in crises. Economically, none of the simple rules examined, moving averages or calendar strategies, created value after costs. The data are consistent with an efficient market at the horizons that matter to investors, without proving it.

Leverage illustrates the same point from another angle. Over this sample, a Kelly exposure of about 1.81x would have maximised growth after borrowing costs, but the estimate is too imprecise to act on, a real-time version did not beat buy and hold, and the compounding gap, deep drawdowns and margin constraints all bite hardest in the crises that dominate long-run outcomes.

On risk, the richer models help to describe volatility, but their forecasting advantage over a simple exponentially weighted average is modest, and every model still underestimates how extreme losses cluster. For long-term investors, the lessons are about horizon and discipline rather than prediction: the probability of loss falls with the holding period but remains uncertain and not negligible, and missing a handful of days in the middle of a crisis is as costly as suffering the worst of it.

Data, methods and reproducibility

Source: Yahoo Finance, downloaded with the `yfinance` library and updated automatically (GitHub repository `https://github.com/cedricmasset98/donnees-marche`). VT: daily closing prices and prices adjusted for dividends and splits ("Adj Close"). Risk-free rate: 13-week US T-bill (`^IRX`). Currency: USD/CHF (`CHF=X`). Annualised returns use calendar time; annualised volatility, Sharpe and Sortino ratios use monthly returns; Sharpe confidence intervals follow Mertens (2002) and Opdyke (2007), allowing for skewness and kurtosis. Bootstrap results use the fixed seed 20080626. Data file fingerprint (SHA-256 of `data/VT.csv`): `bfdcdac71dc770d2`. Software: Python 3.13.17, pandas 3.0.5, NumPy 2.5.3, SciPy 1.18.1, Matplotlib 3.11.2. The five numerical checks of Section 4.5 passed on this run (largest error 1.6e-14); they are also run as `tests/test_kelly.py` before each update. These analyses are descriptive and historical; they do not constitute investment advice.

References

Avellaneda, M. and Zhang, S. (2010). Path-dependence of leveraged ETF returns. *SIAM Journal on Financial Mathematics*, 1(1), 586–603.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

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.

Cheng, M. and Madhavan, A. (2009). The dynamics of leveraged and inverse exchange-traded funds. *Journal of Investment Management*, 7(4), 43–62.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4), 917–926.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1), 41–66.

MacLean, L. C., Thorp, E. O. and Ziemba, W. T. (Eds.) (2011). *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3), 247–257.

Mertens, E. (2002). Comments on variance of the IID estimator in Lo (2002). Working paper, University of Basel.

Opdyke, J. D. (2007). Comparing Sharpe ratios: So where are the p-values? *Journal of Asset Management*, 8(5), 308–336.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.

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. Elsevier.