# In-Sample and Out-of-Sample Sharpe Ratios in Linear Models


Joseph Mulligan
*Imperial College London*
*Qube Research & Technologies*


July 2024

IMPERIAL QRT

# Overview

Question: What is the expected in-sample and out-of-sample Sharpe ratio for a trading strategy?

## Overview

**Question:** What is the expected in-sample and out-of-sample Sharpe ratio for a trading strategy?

**Proposition:** Study a linear predictive model fit and tested on one period (in-sample), and used to trade in another period (out-of-sample)

## Overview

Question: What is the expected in-sample and out-of-sample Sharpe ratio for a trading strategy?

Proposition: Study a linear predictive model fit and tested on one period (in-sample), and used to trade in another period (out-of-sample)

Findings: We give analytic expressions for the expected in-sample and out-of-sample Sharpe ratios, and find:

**Question:** What is the expected in-sample and out-of-sample Sharpe ratio for a trading strategy?

**Proposition:** Study a linear predictive model fit and tested on one period (in-sample), and used to trade in another period (out-of-sample)

**Findings:** We give analytic expressions for the expected in-sample and out-of-sample Sharpe ratios, and find:

1. Higher model complexity inflates in-sample Sharpe ratios
2. Low true Sharpe ratios are vulnerable to overoptimism
3. Short backtests are insufficient to avoid overfitting

Mclean and Pontiff [MP16]:

> *Portfolio returns are* **26% lower out-of-sample** *and* **58% lower post-publication**.

Mclean and Pontiff [MP16]:

> *Portfolio returns are* **26% lower out-of-sample** *and* **58% lower post-publication**.

Falck, Rej, and Thesmar [FRT22]:

> *Published anomalies evaluated outside the data sample deliver about* **50% of in-sample performance**.

Mclean and Pontiff [MP16]:

> *Portfolio returns are* **26% lower out-of-sample** *and* **58% lower post-publication**.

Falck, Rej, and Thesmar [FRT22]:

> *Published anomalies evaluated outside the data sample deliver about* **50% of in-sample performance**.

Suhonen, Lennkh, and Perez [SLP17]:

> *We find a median Sharpe ratio (SR) of 1.20 across the 215 alternative beta strategies during their respective backtest period, compared to 0.31 during live performance ... (a)* **73% median haircut**.

Mclean and Pontiff [MP16]:

> *Portfolio returns are* **26% lower out-of-sample** *and* **58% lower post-publication**.

Falck, Rej, and Thesmar [FRT22]:

> *Published anomalies evaluated outside the data sample deliver about* **50% of in-sample performance**.

Suhonen, Lennkh, and Perez [SLP17]:

> *We find a median Sharpe ratio (SR) of 1.20 across the 215 alternative beta strategies during their respective backtest period, compared to 0.31 during live performance ... (a)* **73% median haircut**.

i.e. **Past performance does not guarantee future results.**

Excluding transaction costs and operational issues, two main problems:

1. Alpha decay - your model was right, but now it's wrong
2. Overfitting - your model was never right

Excluding transaction costs and operational issues, two main problems:

1. Alpha decay - your model was right, but now it's wrong
2. **Overfitting - your model was never right**
    2.1 Multiple testing
    2.2 Parameter misestimation

Excluding transaction costs and operational issues, two main problems:

1. Alpha decay - your model was right, but now it's wrong
2. **Overfitting - your model was never right**
    2.1 Multiple testing
    2.2 **Parameter misestimation**

- Assume the Researcher *fits* **and** *tests* a linear model on a historical period of length $T_1$ yielding the in-sample Sharpe ratio $\mathsf{SR}_{\mathsf{IS}}$.
- What will the expected out-of-sample Sharpe ratio $\mathsf{SR}_{\mathsf{OOS}}$ be on a future period of length $T_2$?

Assume a linear prediction model

$$\underbrace{\mathbf{r}_{t+1}}_{\text{Returns} \in \mathbb{R}^m} = \underbrace{\boldsymbol{\beta}}_{\text{OLS Param} \in \mathbb{R}^{m\times p}} \underbrace{\mathbf{s}_t}_{\text{Signals} \in \mathbb{R}^p} + \underbrace{\boldsymbol{\epsilon}_{t+1}}_{\text{Residuals} \in \mathbb{R}^m} .$$

If $\boldsymbol{\beta}$ was known, then one could form the portfolio $\mathbf{w}_t = \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}\mathbf{s}_t$, which would have Sharpe ratio

$$\mathsf{SR} = \frac{\mathbb{E}[\mathbf{w}_t^\mathsf{T}\mathbf{r}_{t+1}]}{\sqrt{\mathbb{V}[\mathbf{w}_t^\mathsf{T}\mathbf{r}_{t+1}]}} = \frac{\mathbb{E}\left[(\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}(\boldsymbol{\beta}\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})\right]}{\sqrt{\mathbb{V}\left[(\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}(\boldsymbol{\beta}\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})\right]}}.$$

Let $\mathbf{s}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and $\boldsymbol{\epsilon}_{t+1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$ be mutually independent and IID sequences. Then the Sharpe ratio of the strategy is

$$\mathsf{SR} = \frac{\operatorname{tr}(\boldsymbol{\Gamma})}{\sqrt{2\operatorname{tr}(\boldsymbol{\Gamma}^2) + \operatorname{tr}(\boldsymbol{\Gamma})}},$$

where

$$\boldsymbol{\Gamma} := \boldsymbol{\beta}^\mathsf{T} \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\beta}.$$

Thus the Sharpe ratio is increasing in $\operatorname{tr}(\boldsymbol{\Gamma})$ and the ratio $\operatorname{tr}(\boldsymbol{\Gamma}^2)/\operatorname{tr}(\boldsymbol{\Gamma})$.

In reality, we need to estimate $\boldsymbol{\beta}$. If we observe $T_1$ samples of $\mathbf{r}_{t+1}$ and $\mathbf{s}_t$ then by OLS

Stacked returns $\in \mathbb{R}^{m \times T_1}$

Stacked residuals $\in \mathbb{R}^{m \times T_1}$

$$\widehat{\boldsymbol{\beta}} = \mathbf{R} \ \mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1} = \boldsymbol{\beta} + \mathbf{E}\,\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}$$

Stacked signals $\in \mathbb{R}^{p \times T_1}$

And as $\mathbb{E}\left[\mathbf{E}\right] = \mathbf{0} \implies \mathbb{E}\left[\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$. Great!

So one can then form the portfolio $\widehat{\mathbf{w}}_t = \boldsymbol{\Sigma}_{\epsilon}^{-1}\widehat{\boldsymbol{\beta}}\mathbf{s}_t$ and earn $\widehat{\mathsf{PnL}}_t = \widehat{\mathbf{w}}_t^{\mathsf{T}}\mathbf{r}_{t+1}$ per time step.

## Sample Sharpe Ratios

We can then compute the sample means and variances

Historical Period

$$\widehat{\mathbb{E}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right] := \frac{1}{T_1}\sum_{t=0}^{T_1-1}\widehat{\mathsf{PnL}}_t,$$

$$\widehat{\mathbb{V}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right] := \frac{1}{T_1-1}\sum_{t=0}^{T_1-1}\left(\widehat{\mathsf{PnL}}_t - \widehat{\mathbb{E}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right]\right)^2,$$

and these can be used then to estimate the in-sample Sharpe ratio (and similarly the out-of-sample Sharpe ratio)

$$\mathsf{SR}_{\mathsf{IS}} = \frac{\widehat{\mathbb{E}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right]}{\sqrt{\widehat{\mathbb{V}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right]}}, \quad \mathsf{SR}_{\mathsf{OOS}} = \frac{\widehat{\mathbb{E}}\left[(\widehat{\mathsf{PnL}}_u)_{u\in\mathcal{T}_2}\right]}{\sqrt{\widehat{\mathbb{V}}\left[(\widehat{\mathsf{PnL}}_u)_{u\in\mathcal{T}_2}\right]}}.$$

Future Period

Expanding out the sample P&L,

$$
\begin{aligned}
\widehat{\mathsf{PnL}}_t &= (\boldsymbol{\Sigma}_\epsilon^{-1}\widehat{\boldsymbol{\beta}}\mathbf{s}_t)^\mathsf{T}(\boldsymbol{\beta}\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1}) \\
&= (\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t + (\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1} \quad \longleftarrow \text{Truth} \\
&+ \left(\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{E}\mathbf{S}^\mathsf{T}\left(\mathbf{S}\mathbf{S}^\mathsf{T}\right)^{-1}\mathbf{s}_t\right)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t \quad \longleftarrow \text{Misestimation} \\
&+ \left(\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{E}\mathbf{S}^\mathsf{T}\left(\mathbf{S}\mathbf{S}^\mathsf{T}\right)^{-1}\mathbf{s}_t\right)^\mathsf{T}\boldsymbol{\epsilon}_{t+1} \quad \longleftarrow \text{Overfitting}
\end{aligned}
$$

Because when we look at the in-sample P&L,

$$
\mathbf{E} = \left(\ldots \boldsymbol{\epsilon}_{t+1} \ldots\right) \implies \mathbf{E} \not\perp \boldsymbol{\epsilon}_{t+1}.
$$

But, if we look at the out-of-sample P&L, we will have that $\mathbf{E}$ and $\mathbf{S}$ are independent of the new realisations of the signals and the residuals.

$$\mathbf{S} \perp\!\!\!\perp \mathbf{s}_u \quad \mathbf{E} \perp\!\!\!\perp \boldsymbol{\epsilon}_{u+1} \iff u \notin \mathcal{T}_1$$

$$\implies \textbf{Out-of-Sample Overfitting Disappointment}$$

**Q:** How large is this difference?

## Proposition

*Given the previous setup, the expected in-sample and out-of-sample mean are,*

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\widehat{\mathsf{PnL}}_t\right)_{t\in\mathcal{T}_1}\right]\right] = \operatorname{tr}(\boldsymbol{\Gamma}) + \frac{pm}{T_1}, \quad \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\widehat{\mathsf{PnL}}_u\right)_{u\in\mathcal{T}_2}\right]\right] = \operatorname{tr}(\boldsymbol{\Gamma}),$$

*and the expected in-sample and out-of-sample variance are,*

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left(\widehat{\mathsf{PnL}}_t\right)_{t\in\mathcal{T}_1}\right]\right] \approx 2\operatorname{tr}(\boldsymbol{\Gamma}^2) + (c_1 + \widetilde{c}_1)\operatorname{tr}(\boldsymbol{\Gamma}) + c_2 + \widetilde{c}_2,$$

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left(\widehat{\mathsf{PnL}}_u\right)_{u\in\mathcal{T}_2}\right]\right] = 2\operatorname{tr}(\boldsymbol{\Gamma}^2) + c_1\operatorname{tr}(\boldsymbol{\Gamma}) + c_2,$$

*where $c_i$ are some variables which increase with $m, p$ and decrease with $T_1$.*

We propose approximations to the expected in-sample Sharpe ratio $\mathbb{E}[\mathsf{SR}_{\mathsf{IS}}]$ and the expected out-of-sample Sharpe ratio $\mathbb{E}[\mathsf{SR}_{\mathsf{OOS}}]$

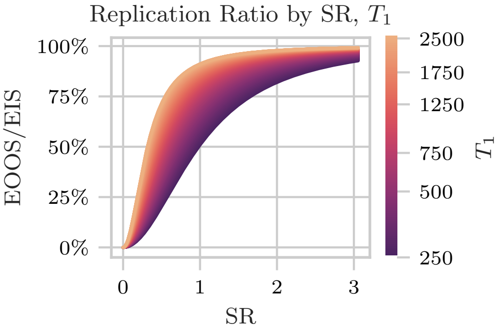
$$\mathbb{E}[\mathsf{SR}_{\mathsf{IS}}] \approx \mathsf{SR}_{\mathsf{EIS}} = \frac{\mathbb{E}\left[\widehat{\mathbb{E}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right]\right]}{\sqrt{\mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\mathsf{PnL}}_t)_{t\in\mathcal{T}_1}\right]\right]}},$$

$$\mathbb{E}[\mathsf{SR}_{\mathsf{OOS}}] \approx \mathsf{SR}_{\mathsf{EOOS}} = \frac{\mathbb{E}\left[\widehat{\mathbb{E}}\left[(\widehat{\mathsf{PnL}}_u)_{u\in\mathcal{T}_2}\right]\right]}{\sqrt{\mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\mathsf{PnL}}_u)_{u\in\mathcal{T}_2}\right]\right]}},$$

and we use these to study the **replication ratio** $\frac{\mathsf{SR}_{\mathsf{EOOS}}}{\mathsf{SR}_{\mathsf{EIS}}}$.

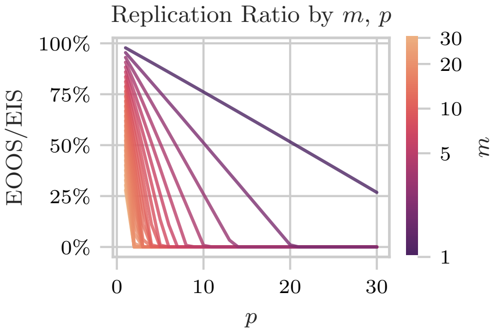
What does this look like in the case where $p = m = 1$?

$$\frac{\mathsf{SR}_{\mathsf{EOOS}}}{\mathsf{SR}_{\mathsf{EIS}}} = \frac{\beta^2 \sqrt{2\beta^4 + \left(1 + \frac{15}{T_1 - 2} - \frac{2}{T_1}\right)\beta^2 + \frac{4}{T_1} - \frac{3}{T_1 + 2} - \frac{1}{T_1^2}}}{\left(\beta^2 + \frac{1}{T_1}\right)\sqrt{2\beta^4 + \left(1 + \frac{2}{T_1 - 2}\right)\beta^2 + \frac{1}{T_1 - 2}}}.$$

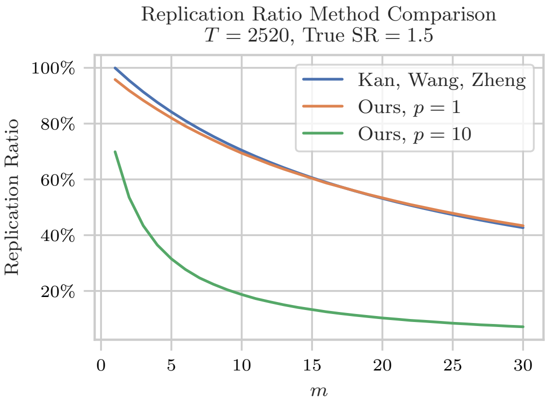


How is the replication ratio affected by increasing the number of assets $m$ and signals $p$?

**Example:** 10 year backtest (2520 days) with in-sample SR 2.



- If we don't have a predictive model and instead just estimate the drift and covariance of our assets, what is the replication ratio? i.e. statically hold the portfolio $\mathbf{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.
- Kan, Wang, and Zheng [KWZ22]: let the true SR be $\theta = \frac{\mathbf{w}^\top \boldsymbol{\mu}}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}} = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}$, the authors compute the expected in-sample SR $\mathbb{E}[\widehat{\theta}]$ and out-of-sample SR $\mathbb{E}[\widetilde{\theta}]$.




Similar haircuts when $p = 1$, but not exactly the same as dynamic vs static.

How does this work when our assumptions are violated?

- Using 12 commodity futures from 1998 to 2023 compute rolling 5-day, 1-year and 5-year t-statistics as signals
- Fit an AR(1) model with $t-$distributed or Normally distributed residuals:

$$\begin{aligned} \mathbf{r}_{t+1} &= \boldsymbol{\beta}\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1}, \\ \mathbf{s}_t &= \boldsymbol{\Phi}\mathbf{s}_{t-1} + \mathbf{u}_t, \end{aligned}$$

- Simulate new samples using this model to check impact of assumption violations

# Commodity Futures

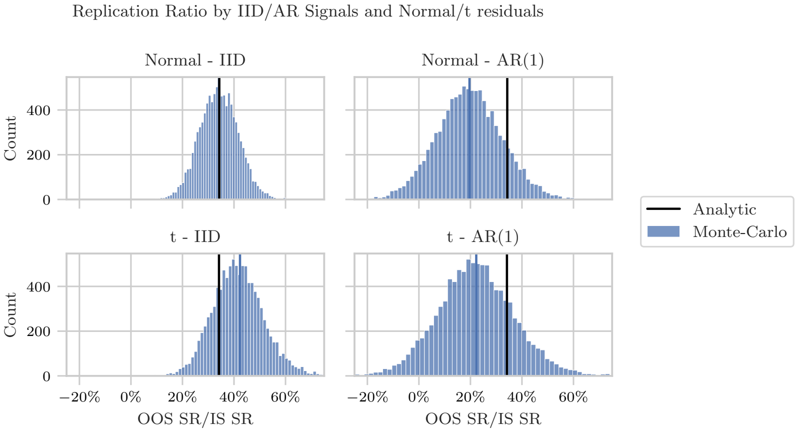



Figure: Distribution of the replication ratio, with iid or AR(1) signals and Normal or $t$ distributed residuals.

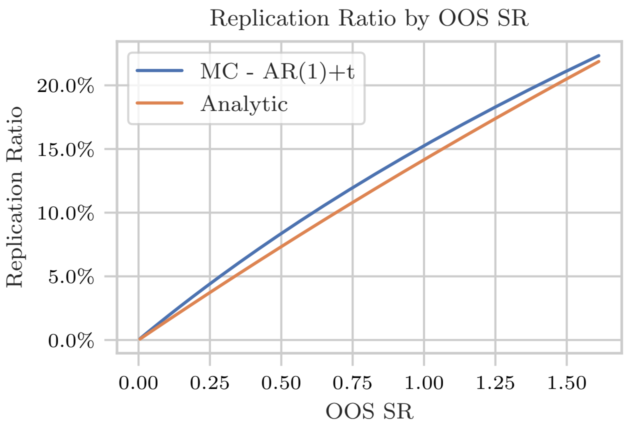



Figure: Distribution of the replication ratio by out-of-sample Sharpe ratio.

Peter Muller (Founder, PDT Partners) [Mul01]:

> *In my opinion it is far better to refine an individual strategy...than to attempt to put together lots of weaker strategies...I would much rather have a single strategy with an expected Sharpe ratio of 2 than a strategy that has an expected Sharpe ratio of 2.5 formed by putting together five supposedly uncorrelated strategies each with an expected Sharpe ratio of 1.*

Nick Patterson (RenTech) [Pat16]:

> *It's funny that I think the most important thing to do in data analysis is to do the simple things right. So, here's a kind of non-secret about what we did at Renaissance: in my opinion, our most important statistical tool was simple regression with one target and one independent variable.*

# Conclusion

Common sense prevails:

1. Use the longest backtest you can
2. Don't use too many signals
3. Don't trust low Sharpe ratios (or too high Sharpe ratios!)

Good luck!

# Bibliography I


[FRT22] Antoine Falck, Adam Rej, and David Thesmar. "When Do Systematic Strategies Decay?" In: *Quantitative Finance* 22.11 (Nov. 2022), pp. 1955–1969. ISSN: 1469-7688. DOI: 10.1080/14697688.2022.2098810.

[KWZ22] Raymond Kan, Xiaolu Wang, and Xinghua Zheng. "In-Sample and Out-of-Sample Sharpe Ratios of Multi-Factor Asset Pricing Models". SSRN Scholarly Paper. Rochester, NY, Mar. 2022. DOI: 10.2139/ssrn.3454628.

[MP16] R. David Mclean and Jeffrey Pontiff. "Does Academic Research Destroy Stock Return Predictability?" In: *The Journal of Finance* 71.1 (2016), pp. 5–32. ISSN: 1540-6261. DOI: 10.1111/jofi.12365.

[Mul01] P. Muller. "Proprietary Trading: Truth and Fiction". In: *Quantitative Finance* 1.1 (Jan. 2001), pp. 6–8. ISSN: 1469-7688. DOI: 10.1080/713665546.

[Pat16] Nick Patterson. "Talking Machines: AI Safety and the Legacy of Bletchley Park, with Nick Patterson - Robohub". https://robohub.org/talking-machines-ai-safety-and-the-legacy-of-bletchley-park-with-nick-patterson/. Feb. 2016.

[SLP17] Antti Suhonen, Matthias Lennkh, and Fabrice Perez. "Quantifying Backtest Overfitting in Alternative Beta Strategies". In: *The Journal of Portfolio Management* 43.2 (Jan. 2017), pp. 90–104. ISSN: 0095-4918, 2168-8656. DOI: 10.3905/jpm.2017.43.2.090.