In-Sample and Out-of-Sample Sharpe Ratios for Linear Predictive Models

Antoine Jacquier¹, Johannes Muhle-Karbe¹, and Joseph Mulligan^{1, 2}

 $^1\mathrm{Department}$ of Mathematics, Imperial College London $^2\mathrm{Qube}$ Research & Technologies

December 2, 2025

Abstract

We study how much the in-sample performance of trading strategies based on linear predictive models is reduced out-of-sample due to overfitting. More specifically, we compute the in- and out-of-sample means and variances of the corresponding PnLs and use these to derive a closed-form approximation for the corresponding Sharpe ratios. We find that the out-of-sample "replication ratio" diminishes for complex strategies with many assets and based on many weak rather than a few strong trading signals, and increases when more training data is used. The substantial quantitative importance of these effects is illustrated with a simulation case study for commodity futures following the methodology of Gârleanu and Pedersen [GP13], and an empirical case study using the dataset compiled by Goyal, Welch, and Zafirov [GWZ24].

Conflict of Interest Statement

The authors have no conflicts of interest to disclose.

Data Availability Statement

The data that support the findings of this study consists of publicly available commodity futures prices, which are accessible from various market data providers, and the dataset compiled by Goyal, Welch and Zafirov, which is available from Amit Goyal's webpage.

Open Access Statement

For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence (where permitted by UKRI, 'Open Government Licence' or 'Creative Commons Attribution No-derivatives (CC BY-ND) licence' may be stated instead) to any Author Accepted Manuscript version 'arising'.

Funding Statement

This research was supported by Qube Research & Technologies and the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1).

Corresponding Author: j.mulligan21@imperial.ac.uk. Please note the disclaimer from QRT at the end of this document. The authors would like to thank Daniel Giamouridis and Paolo Guasoni for their pertinent remarks, and discussants at Imperial College London and Qube Research & Technologies for fruitful discussions throughout the preparation of this work. The pertinent remarks of three anonymous reviewers are also gratefully acknowledged.

1 Introduction

A standard approach to building systematic trading strategies is to first fit some predictive model to historical training data. For example, Gârleanu and Pedersen [GP13] regress price changes of commodity futures against past returns with various lookback periods. Using the forecasting model, one can then form a trading strategy, e.g., by combining forecasts for several assets using mean-variance analysis, and "backtest" its performance on historical data. The strategy's "out-of-sample" performance on new testing data or in live trading will typically be worse than its "in-sample" performance on the training data for a number of reasons. But how much worse?

A number of empirical studies find that the out-of-sample performance of ETFs, anomalies and various other strategies tends to be significantly lower than the corresponding in-sample performance, with estimates ranging widely depending on the dataset [BL15; Wie+16; MP16; SLP17; FRT22]. One commonly used rule of thumb is to haircut the in-sample performance by 50% [HL15]. While such rules of thumb are useful, we seek a more systematic approach, enabling us to estimate the out-of-sample performance of different trading strategies based on their characteristics, and without having to resort to a new large-scale simulation study each time the strategy is modified.

To this end, we focus on one crucial issue in this context: overfitting.² Kan, Wang, and Zheng [KWZ24] obtain analytical results in this spirit for the simplest setting where investment opportunities are constant, in that expected asset returns and covariances do not change over time. When these model parameters are estimated from a finite training dataset, Kan, Wang, and Zheng derive explicit formulas for the "replication ratio", that is, the fraction of the in-sample Sharpe ratio that can be recovered out-of-sample.

In the present paper, we extend this analysis to a setting closer to that of Gârleanu and Pedersen and many related papers, where trading strategies exploit several trading signals per asset and also take into account that these signals are not fully persistent but fluctuate over time. More specifically, we assume that expected returns are linear functions of iid trading signals drawn from a multivariate normal distribution. When the variance of this distribution tends to zero, the constant-parameter model of Kan, Wang, and Zheng is recovered as a special case. Conversely, our iid expected returns correspond to the limiting case of short-lived signals in the setting of Gârleanu and Pedersen.

The relationship between signals and expected returns can be estimated empirically by linear regression. For portfolios corresponding to this plug-in estimate, we compute the (unconditional) in- and out-of-sample means and variances of the P&L and derive closed-form approximations for the replication ratio, i.e., the fraction of the in-sample Sharpe ratio that can be achieved out-of-sample.

We find that as in Kan, Wang, and Zheng [KWZ24] the replication ratio is increasing in the size of the training dataset and the magnitude of the "true" Sharpe ratio of the strategy (i.e., the exact unconditional Sharpe ratio computed in the model rather than based on estimated parameters). Indeed, when the true Sharpe ratio is low the strategies are more prone to overfitting, and small increases in the true Sharpe ratio yield significant improvements in the replication ratio. Once the true Sharpe ratio is sufficiently large, the model is more robust to overfitting and the replication ratio slowly converges towards 100%.

Conversely, the replication ratio is decreasing in the number of model parameters. In Kan, Wang, and Zheng [KWZ24] this quantity only depends on the number of assets; in our setting, it is determined both by the number of assets and by the trading signals per asset. The takeaway message is that complex strategies are more prone to overfitting, which needs to be balanced with sufficiently long backtesting periods.

The practical relevance of these theoretical results is in turn explored in a simulation study that follows the implementation in [GP13]. To wit, we fit a linear predictive model to a set of commodity futures forecast using momentum-style signals, and in turn use this as a simulation

¹ More generally, as in many other quantitative disciplines [Ioa05], there is a controversial ongoing debate whether there is a "replication crisis" in financial economics [HLZ16; JKP23].

² Another key driver of lower out-of-sample performance is *multiple testing* studied by [BL14; HL15]. While multiple testing may lead to overfitting, we focus on overfitting from a single test perspective.

engine. We test our approach on simulations with both Gaussian and fat-tailed returns, and for both iid and autoregressive signals. Our analytical approximations provide very good predictions in the Gaussian iid setting where they were derived, but also perform encouragingly well with autoregressive signals driven by heavy-tailed noise, confirming the practicality of our results in realistic settings.

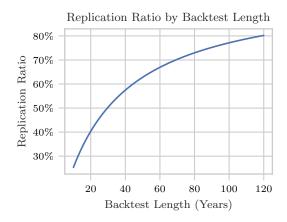


Figure 1: The replication ratio for the case study in Section 5 with varying backtest length.

We observe that a ten-year backtest yields an expected replication ratio of just 30% (Figure 1). This steep haircut arises from the large number of parameters in the model; it can be reduced by imposing constraints on the model, such as regularisation or a panel regression as in the empirical section of [GP13].

To further confirm that our results are also relevant for real data, we test the method with the comprehensive dataset compiled by Goyal, Welch, and Zafirov [GWZ24], which provides over 100 years of monthly signals which have been published in the literature for predicting the equity premium. This allows us to deploy the methodology developed in the present paper to subsets of the predictors and, in doing so, control both complexity of the composite strategy and backtest length. Encouragingly, we recover the key relationships from our model in the data, in that more signals lead to more overfitting, whereas longer backtests and higher strength signals mitigate the issue.

In summary, our results suggest the following broad practical conclusions: be wary of low Sharpe ratio strategies, keep your models as simple as possible, and use the longest sensible backtest period available. While much of this advice aligns with practitioners' intuition, our results allow one to pinpoint how much of a haircut one should expect out-of-sample depending on the characteristics of the backtest, rather than broad rules of thumb. Additionally, estimating the out-of-sample haircut by simulation is often numerically slow. In contrast, our closed-form expressions permit accelerating this time-sensitive process by instantly computing the expected out-of-sample haircuts.

The remainder of this paper is organised as follows. Section 2 sets up the model and differentiates between the true and the sample Sharpe ratios. Section 3 presents the main results and their interpretation. Section 4 compares our results to the existing literature on overfitting. Section 5 applies our results in a simulation case study using commodity futures, Section 6 applies the method to the Goyal, Welch, and Zafirov dataset, and Section 7 concludes. Section A and C provide derivations of results used in the main text, Section B provides the general case of the main result, and Section D provides details of the Monte Carlo analysis we use to check the approximations used.

2 Model Setup

We consider a researcher who wants to fit a linear model to forecast asset returns and to trade based on these predictions. The researcher has access to some historical data observed at discrete times $\mathcal{T}_1 = \{0, \ldots, T_1\}$. We denote the returns per time step by $\mathbf{r}_{t+1} \in \mathbb{R}^m$ and the predictive signals by $\mathbf{s}_t \in \mathbb{R}^p$ for each $t \in \mathcal{T}_1$. The researcher assumes that the relationship between the signals and the returns follows a linear model with parameter $\boldsymbol{\beta} \in \mathbb{R}^{m \times p}$:

$$\mathbf{r}_{t+1} = \beta \mathbf{s}_t + \boldsymbol{\epsilon}_{t+1},$$

where $\epsilon_{t+1} \in \mathbb{R}^m$ denotes the unexplained residual returns. We assume that the sequences $\{\mathbf{s}_t\}$ and $\{\epsilon_{t+1}\}$ are independent of each other, and iid across time with distributions \mathfrak{s} and \mathfrak{e} respectively, which are assumed to be multivariate Gaussian:³

$$\mathfrak{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$$
 and $\mathfrak{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon}).$

We note here that trading signals are typically assumed to be autoregressive as in [GP13]. However, imposing the iid assumption as in [DL18] makes the calculations below substantially more tractable, and we demonstrate using simulation experiments in Section 5 that the resulting formulas also provide very good approximations for autoregressive signals fitted using the methodology from the empirical part of [GP13].

We impose without loss of generality that the signals are centred, except for a possible intercept term, i.e., $\mu_s = (\mu_{s,1}, 0, ...)$ where $\mu_{s,1} = 1$ if the model includes a constant (intercept) term, and $\mu_{s,1} = 0$ otherwise; removing the intercept simplifies many of the expressions, but we leave the expressions general as including a drift term may be desirable for some applications such as long-only equity portfolios.

Assuming $\boldsymbol{\beta}$ is known, the researcher can use the signal \mathbf{s}_t at time t to accurately predict the expected returns $\mathbb{E}_t[\mathbf{r}_{t+1}] = \boldsymbol{\beta}\mathbf{s}_t$ conditional on the information at time t, and in turn trade a portfolio of the form $\mathbf{w}_t = \mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t$ where $\mathbf{Z} \in \mathbb{R}^{m \times m}$ is some symmetric weight matrix. For example, the standard (one-period) Markowitz portfolio is given by $\mathbf{w}_t = \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$ where $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\mu}_t$ denote the covariance matrix and expected returns of the traded assets conditional on the information available at time t. In the present context, these are given by

$$\mathbb{E}_t[\beta \mathbf{s}_t + \boldsymbol{\epsilon}_{t+1}] = \beta \mathbf{s}_t, \quad \text{Cov}_t[\beta \mathbf{s}_t + \boldsymbol{\epsilon}_{t+1}] = \boldsymbol{\Sigma}_{\epsilon},$$

and thus the one-step Markowitz optimal portfolio is of the form $\mathbf{w}_t = \mathbf{Z}\beta\mathbf{s}_t$ with $\mathbf{Z} = \boldsymbol{\Sigma}_{\epsilon}^{-1}$. Somewhat less common but still relevant choices of \mathbf{Z} include $\mathrm{diag}(\boldsymbol{\Sigma}_{\epsilon})^{-1}$, corresponding to an inverse variance allocation by asset, or \mathbf{I}_m , corresponding to a risk indifferent portfolio. The researcher can then test this portfolio against the returns to compute the Profit & Loss per time step,

$$PnL_t = \mathbf{w}_t^\mathsf{T} \mathbf{r}_{t+1}.$$

2.1 The True Sharpe Ratio

The Sharpe ratio typically refers to a summary statistic that compares the sample mean of a time series of returns to their sample standard deviation. In the model from the previous section, we can compare this to its "true" equivalent, defined as the ratio of the expected return and standard deviation of the daily P&L, averaged over the distributions of the signal and noise terms:

$$SR := \frac{\mathbb{E}[PnL_t]}{\sqrt{\mathbb{V}[PnL_t]}} = \frac{\mathbb{E}[\mathbf{w}_t^\mathsf{T} \mathbf{r}_{t+1}]}{\sqrt{\mathbb{V}[\mathbf{w}_t^\mathsf{T} \mathbf{r}_{t+1}]}} = \frac{\mathbb{E}[(\mathbf{Z}\beta\mathbf{s}_t)^\mathsf{T}(\beta\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})]}{\sqrt{\mathbb{V}[(\mathbf{Z}\beta\mathbf{s}_t)^\mathsf{T}(\beta\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})]}}.$$
 (2.1)

Using identities for quadratic expectations of Gaussian vectors, this can be computed as

$$SR = \frac{\operatorname{tr}(\mathbf{G}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{G}\boldsymbol{\mu}_s}{\sqrt{2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_s)^2) + 4\boldsymbol{\mu}_s^\mathsf{T}\mathbf{G}\boldsymbol{\Sigma}_s\mathbf{G}\boldsymbol{\mu}_s + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{F}\boldsymbol{\mu}_s}},$$
(2.2)

³ We focus on the Gaussian case as it simplifies the exposition in many places. One can compute the true Sharpe ratio when the signals/noise are non-Gaussian, but we omit it here as the expression is verbose and broadly similar.

where

$$\mathbf{F} \coloneqq \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta} \quad \text{and} \quad \mathbf{G} \coloneqq \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta}.$$

See Section A for the derivation of this expression. To provide some intuition for this result, suppose for simplicity that the signals include no intercept ($\mu_s = 0$). Then, (2.2) simplifies to

$$SR = \frac{\operatorname{tr}(\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z} \boldsymbol{\beta} \boldsymbol{\Sigma}_{s})}{\sqrt{2 \operatorname{tr}((\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z} \boldsymbol{\beta} \boldsymbol{\Sigma}_{s})^{2}) + \operatorname{tr}(\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z} \boldsymbol{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta} \boldsymbol{\Sigma}_{s})}}.$$

In this form, clearly the expected return (numerator) increases with the Frobenius norm $\|\beta\|_F$ (which measures the "predictability" of future asset returns), however the variance (denominator) also increases. For a single asset or signal, the Sharpe ratio is monotonically increasing in $\|\beta\|_F$, but with multiple assets or signals the relationship is no longer obvious. We also see that the Sharpe ratio increases as the signal-to-noise ratio rises with smaller $\|\Sigma_{\epsilon}\|_F$. Similarly to β , the relationship with the signal variance Σ_s is generally not monotone, but for parameters similar to our simulation case study in Section 5 the Sharpe ratio typically increases when $\|\Sigma_s\|_F$ increases relative to $\|\Sigma_{\epsilon}\|_F$, again indicating an improvement in the signal-to-noise ratio of the model.

2.2 The In-Sample Sharpe Ratio

When a given predictive model is deployed in practice, its parameters β must be estimated. In the linear model we consider here, this can be done using ordinary least-squares (OLS) by stacking the returns across all time steps into the matrix $\mathbf{R} \in \mathbb{R}^{m \times T_1}$ and the signals across all time steps into the matrix $\mathbf{S} \in \mathbb{R}^{p \times T_1}$ and computing the OLS estimator

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

Although this is the best estimator for β in many senses (unbiased, L^2 optimal, consistent, etc.), it is also naturally overfit to the particular realisations of the signal and returns in the in-sample period, which will clearly differ in the out-of-sample period.

Moreover, using this estimate for β for backtesting on the same in-sample dataset naturally induces "look-ahead bias", as data from the full sample period has been used to estimate the model parameters which are then used to test the trading strategy on the same period.⁴

To determine the magnitude of these effects, we can also stack the residuals ϵ_{t+1} into the matrix $\mathbf{E} \in \mathbb{R}^{m \times T_1}$ and with a substitution of the model for the returns into (2.3) we obtain

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

This expression enables us to decompose the estimated regression parameter into the true regression parameter plus some sampling error. If the researcher trades the portfolio $\hat{\mathbf{w}}_t = \mathbf{Z}\hat{\boldsymbol{\beta}}\mathbf{s}_t$ using this estimate, the corresponding P&L per time step is $\widehat{\text{PnL}}_t = \widehat{\mathbf{w}}_t^{\mathsf{T}}\mathbf{r}_{t+1}$. The sample average and variance in the training data set in turn are

$$\widehat{\mathbb{E}}\left[(\widehat{\mathbf{PnL}}_t)_{t\in\mathcal{T}_1}\right] := \frac{1}{T_1} \sum_{t=0}^{T_1-1} \widehat{\mathbf{PnL}}_t, \tag{2.4}$$

$$\widehat{\mathbb{V}}\left[(\widehat{\mathrm{PnL}}_t)_{t\in\mathcal{T}_1}\right] := \frac{1}{T_1 - 1} \sum_{t=0}^{T_1 - 1} \left(\widehat{\mathrm{PnL}}_t - \widehat{\mathbb{E}}\left[(\widehat{\mathrm{PnL}}_t)_{t\in\mathcal{T}_1}\right]\right)^2. \tag{2.5}$$

Together, these lead to the in-sample Sharpe ratio

$$SR_{IS} := \frac{\widehat{\mathbb{E}}\left[(\widehat{PnL}_t)_{t \in \mathcal{T}_1}\right]}{\sqrt{\widehat{\mathbb{V}}\left[(\widehat{PnL}_t)_{t \in \mathcal{T}_1}\right]}}.$$

⁴ The look-ahead bias could be avoided using a walk-forward regression as in [Jou+24], but this does not avoid the issue of overfitting.

The issue here is that $\widehat{\beta}$ has been estimated using the same returns observations \mathbf{r}_{t+1} which the model is then tested against, i.e. the model has been fit to the specific realisation of the noise in the historical period. Expanding out the sample P&L,

$$\widehat{\operatorname{PnL}}_{t} = \widehat{\mathbf{w}}_{t}^{\mathsf{T}} \mathbf{r}_{t+1} = (\boldsymbol{\Sigma}_{\epsilon}^{-1} \widehat{\boldsymbol{\beta}} \mathbf{s}_{t})^{\mathsf{T}} (\boldsymbol{\beta} \mathbf{s}_{t} + \boldsymbol{\epsilon}_{t+1}) \\
= (\boldsymbol{\Sigma}_{\epsilon}^{-1} \boldsymbol{\beta} \mathbf{s}_{t})^{\mathsf{T}} \boldsymbol{\beta} \mathbf{s}_{t} + (\boldsymbol{\Sigma}_{\epsilon}^{-1} \boldsymbol{\beta} \mathbf{s}_{t})^{\mathsf{T}} \boldsymbol{\epsilon}_{t+1} \leftarrow \text{Truth} \\
+ (\boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{E} \mathbf{S}^{\mathsf{T}} (\mathbf{S} \mathbf{S}^{\mathsf{T}})^{-1} \mathbf{s}_{t})^{\mathsf{T}} \boldsymbol{\beta} \mathbf{s}_{t} \leftarrow \text{Misestimation} \\
+ (\boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{E} \mathbf{S}^{\mathsf{T}} (\mathbf{S} \mathbf{S}^{\mathsf{T}})^{-1} \mathbf{s}_{t})^{\mathsf{T}} \boldsymbol{\epsilon}_{t+1} \cdot \leftarrow \text{Overfitting}$$

We observe that by using $\hat{\boldsymbol{\beta}}$ instead of $\boldsymbol{\beta}$ we obtain a combination of the true P&L and some additional P&L as a result of the estimation error in $\boldsymbol{\beta}$. In the in-sample period $\boldsymbol{\epsilon}_{t+1}$ is an element of \mathbf{E} , but in the out-of-sample period it is not (\mathbf{E} is the matrix of stacked noise realisations from the in-sample period). The impact of this is that, in expectation, the misestimation term has expected value zero in both the in- and out-of-sample period due to $\mathbb{E}[\mathbf{E}] = \mathbf{0}$. The overfitting term however only has zero expectation in the out-of-sample period, leading to inflation in the expected return in the in-sample period, this is how the look ahead bias discussed earlier inflates the in-sample performance. However, in both periods the variance of the sample P&L is higher than if we had used the true estimator $\boldsymbol{\beta}$, as we will see in our main result in Proposition 3.1.

2.3 The Out-of-Sample Sharpe Ratio

After estimating the predictive model on the training dataset, the researcher deploys the model out-of-sample at discrete trading times $\mathcal{T}_2 = \{T_1 + 1, \dots, T_1 + T_2 + 1\}$. The corresponding sample mean and variance then lead to the out-of-sample Sharpe ratio

$$SR_{OOS} := \frac{\widehat{\mathbb{E}}\left[(\widehat{PnL}_t)_{t \in \mathcal{T}_2}\right]}{\sqrt{\widehat{\mathbb{V}}\left[(\widehat{PnL}_t)_{t \in \mathcal{T}_2}\right]}}.$$

The difference, of course, is that in the future period $\widehat{\beta}$ is not fit to the particular observations of the returns and signals, and thus the out-of-sample Sharpe ratio will typically be lower than the in-sample Sharpe ratio previously calculated. We also see from the expressions derived below that the out-of-sample Sharpe ratio will be lower than the true Sharpe ratio due to misestimation of the regression parameter causing increased volatility over trading with the true β .

The questions therefore are: How much lower? And what can be done to mitigate this bias?

3 Main Results

The unconditional means of the in-sample and out-of-sample Sharpe ratio, $\mathbb{E}[SR_{IS}]$ and $\mathbb{E}[SR_{OOS}]$, cannot be computed in closed form. We can however compute the expected in-sample and out-of-sample expected return and variance. For clarity of exposition, we present here our main result for a special (but extremely common) case, taking $\mathbf{Z} = \boldsymbol{\Sigma}_{\epsilon}^{-1}$, $\mu_s = \mathbf{0}$ and $\boldsymbol{\Sigma}_s = \mathbf{I}_p$, i.e., the Markowitz portfolio for a set of centred standardised signals. The impact of having to estimate the covariance matrix when constructing the portfolio is discussed in Section 3.3. The general version of this result is reported in Appendix B.

Proposition 3.1. The expected average in-sample PEL and out-of-sample PEL is given by

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\widehat{\mathrm{PnL}}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] = \mathrm{tr}(\mathbf{\Gamma}) + \frac{pm}{T_{1}},$$

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\widehat{\mathrm{PnL}}_{u}\right)_{u\in\mathcal{T}_{2}}\right]\right] = \mathrm{tr}(\mathbf{\Gamma}).$$

The expected variance of the in-sample and out-of-sample P&L is given by

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

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

where

$$\Gamma = \beta^{\mathsf{T}} \Sigma_{\epsilon}^{-1} \beta,
c_{1} = 1 + \frac{p+1}{T_{1} - p - 1}, \qquad \tilde{c}_{1} = \frac{2p+5}{T_{1} - p - 1} + \frac{2m(p^{2} + p + 2T_{1})}{T_{1}(T_{1} - p - 1)},
c_{2} = \frac{mp}{T_{1} - p - 1}, \qquad \tilde{c}_{2} = \frac{mp(2m + p + T_{1} + 4)}{T_{1}(T_{1} + 2)} - \frac{2m^{2}p^{2}}{T_{1}^{2}(T_{1} + 2)} - \frac{mp}{T_{1} - p - 1},$$
(3.1)

and

$$\begin{split} \varepsilon &= \operatorname{tr} \left(\operatorname{Cov} \left[(\mathbf{S} \mathbf{S}^\mathsf{T})^{-1}, \mathbf{s}_t \mathbf{s}_t^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_\epsilon \mathbf{Z} \boldsymbol{\beta} \mathbf{s}_t \mathbf{s}_t^\mathsf{T} \right] \right) + \operatorname{tr} \left(\operatorname{Cov} \left[(\mathbf{S} \mathbf{S}^\mathsf{T})^{-1}, \mathbf{s}_t \mathbf{s}_t^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta} \mathbf{s}_t \mathbf{s}_t^\mathsf{T} \right] \right) \\ &+ \operatorname{tr} \left(\operatorname{Cov} \left[(\mathbf{S} \mathbf{S}^\mathsf{T})^{-1}, \mathbf{s}_t \mathbf{s}_t^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \mathbf{Z} \boldsymbol{\Sigma}_\epsilon \mathbf{Z} \boldsymbol{\beta} \mathbf{s}_t \mathbf{s}_t^\mathsf{T} \right] \right). \end{split}$$

The proof relies on a combination of results on the expectations of Gaussian vectors or matrices, properties of projection matrices and matrix statistics tricks. For better readability, these lengthy computations are deferred to Appendix C. The out-of-sample expected PnL and its variance as well as the in-sample expected PnL can be computed explicitly. The in-sample variance of the PnL is also almost fully explicit, up to the ε term which does not have a closed-form expression. In practice, we approximate the in-sample variance by simply dropping this term as it is typically very small compared to the other terms. The accuracy of this approximation is verified in Section 3.1. Here, let us mention a few implications.

Remark 1 (Expected Mean Return). Examining the difference between the expected in-sample and out-of-sample returns, we see that the in-sample return is inflated by $\frac{pm}{T_1}$. Thus, as the number of in-sample observations T_1 increases, our in-sample P&L will converge to our out-of-sample P&L. Conversely, as the number of signals p increases the overfitting bias becomes arbitrarily large. Increasing the number of assets m will also lead to further inflation of the in-sample P&L (see Figure 6).

Remark 2 (Expected Variance). Inspecting the expected variance, we see that $\tilde{c}_1 > 0$ provided $T_1 > p+1$, which we would always expect to be the case, but whether $\tilde{c}_2 > 0$ depends on the interplay between the length T_1 of the backtest, as well as the number m and p of assets and signals. For a range of backtest lengths, Figure 2 displays the combinations of m, p for which $\tilde{c}_2 > 0$. Generally, when p < m (with some sublinearity for large m) then $\tilde{c}_2 > 0$, and for large enough T_1 then p must be much larger than m for $\tilde{c}_2 < 0$. Overfitting β thus causes the in-sample variance to be marginally larger than the out-of-sample variance for small m, but for large m it can be significantly higher. Further, the out-of-sample variance can be much larger than the true variance, which causes the out-of-sample Sharpe to be lower than the true theoretical Sharpe ratio if β was known.

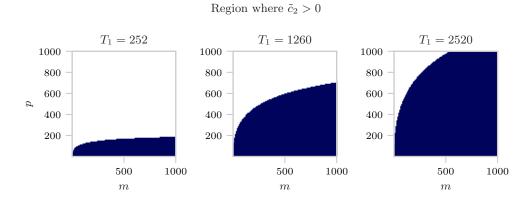


Figure 2: Values of m, p, T_1 where $\tilde{c}_2 > 0$.

3.1 Accuracy of the Approximation of the In-Sample Variance

In our simulation study for commodity futures in Section 5, we find that the error term ε from Proposition 3.1 is negative but small relative to the other terms, with the estimated variance being 3% smaller than the analytical result with $\varepsilon = 0$. To support the robustness of this observation across a wide range of parameter values, we run separate simulations where the model parameters are selected randomly from distributions centred around the empirical point estimates (full details can be found in Appendix D). This is distinct from the case study in Section 5, where all true parameters are calibrated to that particular dataset and held fixed. We find that for T_1 sufficiently larger than p, ε are small, less than 3% for $p/T_1 < 0.1$ as can be seen in Figure 3. This is equivalent to having at least ten data points per signal which in fact is a common rule of thumb [Har06, Sec. 4.4]. Furthermore, m has no strong effect on ε as these terms are driven by the covariance between the inverse Gram matrix (\mathbf{SS}^{T})⁻¹ and the individual signal vectors \mathbf{s}_t , which becomes negligible as the number of signals p increases. Based on these observations, we henceforth disregard ε in our analysis.

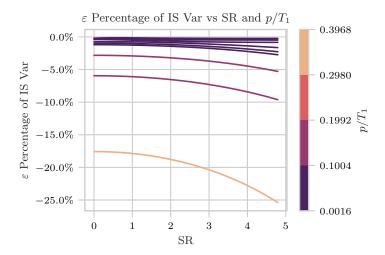


Figure 3: Magnitude of $\sum_{i} \varepsilon_{i}$ as a percentage of the in-sample variance $\mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\mathrm{PnL}}_{t})_{t \in \mathcal{T}_{1}}\right]\right]$.

3.2 The Replication Ratio

In view of Proposition B.1 and the discussion in Section 3.1 we propose the following approximations to the expected in- and out-of-sample Sharpe ratios:

$$\mathbb{E}\left[\operatorname{SR}_{\operatorname{IS}}\right] \approx \operatorname{SR}_{\operatorname{EIS}} := \frac{\mathbb{E}\left[\widehat{\mathbb{P}}\left[(\widehat{\operatorname{PnL}}_{t})_{t \in \mathcal{T}_{1}}\right]\right]}{\sqrt{\mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\operatorname{PnL}}_{t})_{t \in \mathcal{T}_{1}}\right]\right]}},$$

$$\mathbb{E}\left[\operatorname{SR}_{\operatorname{OOS}}\right] \approx \operatorname{SR}_{\operatorname{EOOS}} := \frac{\mathbb{E}\left[\widehat{\mathbb{P}}\left[(\widehat{\operatorname{PnL}}_{t})_{t \in \mathcal{T}_{2}}\right]\right]}{\sqrt{\mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\operatorname{PnL}}_{t})_{t \in \mathcal{T}_{2}}\right]\right]}}.$$
(3.2)

Here we take $\varepsilon = 0$ in $\mathbb{E}[\widehat{\mathbb{V}}[(\widehat{\mathrm{PnL}}_t)_{t \in \mathcal{T}_1}]]$ as discussed in Section 3.1. Moreover, we assume that the difference between taking expectations of the numerator and denominator individually, rather than jointly, is small.

We again check this by the same simulation process as Section 3.1 and find that indeed the convexity adjustment is generally very small, with maximum errors of 0.01 for the in-sample Sharpe ratio and 0.005 for the out-of-sample Sharpe ratio, as can be seen in Figure 4. Compared to the magnitude of the sample Sharpe ratios (shown in the second row), the errors due to convexity are negligible, at most 2.5%. In these simulations we let p = m and $T = T_1, T_2$.

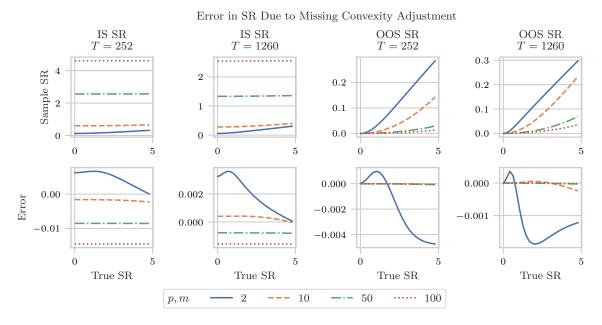


Figure 4: Difference between the expected Sharpe Ratio $\mathbb{E}[SR]$ and the ratio of the expected mean and variance $\mathbb{E}[\widehat{\mathbb{E}}[(\widehat{PnL}_t)_{t\in\mathcal{T}_i}]]/\sqrt{\mathbb{E}[\widehat{\mathbb{V}}[(\widehat{PnL}_t)_{t\in\mathcal{T}_i}]]}$, i.e. the error due to the missing convexity adjustment for both in-sample and out-of-sample Sharpe ratios.

Armed with these analytical approximations, we can study the "replication ratio", i.e., the fraction SR_{EOOS}/SR_{EIS} of the in-sample Sharpe ratio that is recovered out of sample. A replication ratio of 100% means that the strategy completely replicates out of sample. Conversely, a replication

⁵ Note these estimates are purely calculated via simulation, and we do not use our analytical values. This is as the error due to the missing ε terms (as studied in Section 3.1) dominates the small convexity error when $p = 100, T_1 = 252$, yielding no insight into the magnitude of the convexity error.

ratio of 0% means that it returns zero, e.g., because the signals are completely uninformative (i.e., the true parameter β is $\mathbf{0}$) and thus the expected out-of-sample Sharpe ratio is zero.

Univariate Case To gain some intuition we first evaluate the replication ratio in the simplest case with one asset (m=1) and one dynamic signal with no intercept (p=1) and where we scale σ_s to be equal to σ_ϵ , which is always possible as one can always scale the original signal to achieve this. In this case the formulas derived above simplify as

$$\begin{split} \mathrm{SR} &= \frac{\beta^2}{\sqrt{2\beta^4 + \beta^2}}, \\ \mathrm{SR}_{\mathrm{EIS}} &= \frac{\beta^2 + \frac{1}{T_1}}{\sqrt{2\beta^4 + \left(1 + \frac{15}{T_1 - 2} - \frac{2}{T_1}\right)\beta^2 + \frac{4}{T_1} - \frac{3}{T_1 + 2} - \frac{1}{T_1^2}}}, \\ \mathrm{SR}_{\mathrm{EOOS}} &= \frac{\beta^2}{\sqrt{2\beta^4 + \left(1 + \frac{2}{T_1 - 2}\right)\beta^2 + \frac{1}{T_1 - 2}}}. \end{split}$$

Whence, the replication ratio only depends on the size T_1 of the training dataset and the parameter β (or, equivalently, the true Sharpe ratio) in this case:

$$\frac{\text{SR}_{\text{EOOS}}}{\text{SR}_{\text{EIS}}} = \frac{\beta^2 \sqrt{2\beta^4 + \left(1 + \frac{15}{T_1 - 2} - \frac{2}{T_1}\right)\beta^2 + \frac{4}{T_1} - \frac{3}{T_1 + 2} - \frac{1}{T_1^2}}}{\left(\beta^2 + \frac{1}{T_1}\right)\sqrt{2\beta^4 + \left(1 + \frac{2}{T_1 - 2}\right)\beta^2 + \frac{1}{T_1 - 2}}}.$$
(3.3)

We plot a heat map of this relationship in Figure 5. As the Sharpe ratio (equivalently β) increases, overfitting becomes less of an issue. Indeed, as the signal-to-noise ratio becomes arbitrarily large for $\beta \to \infty$, the replication ratio approaches 100% (SR_{EOOS}/SR_{EIS} \to 1). Similarly, the difference between in- and out-of-sample performance also vanishes as the training data set becomes arbitrarily large ($T_1 \to \infty$). Both these relationships are nonlinear, with large initial improvements that eventually saturate.

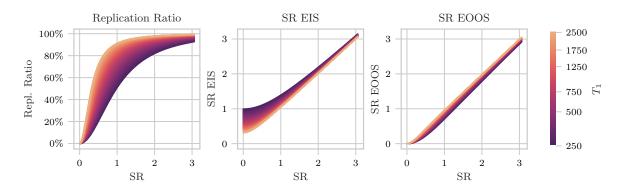


Figure 5: Replication ratio $\frac{SR_{EOOS}}{SR_{EIS}}$, SR_{EIS} and SR_{EOOS} in the 1-asset, 1-signal case as in (3.3), with varying SR, T_1 .

Multivariate Case We now turn to the case of multiple assets and signals (m, p > 1). To illustrate some of the basic mechanisms, we look at a representative example. To wit, we consider an in-sample Sharpe ratio of 2 with a ten-year backtest of daily returns $(T_1 = 2520)$, and assume that assets and signals are uncorrelated, $\Sigma_s = \mathbf{I}_p$, $\Sigma_{\epsilon} = \mathbf{I}_m$. For simplicity, we also assume that the signals include no intercept $(\mu_s = \mathbf{0})$ and that investments are directly proportional to the

signals without reweighing $(\mathbf{Z} = \mathbf{I}_m)$. Finally, to achieve the desired in-sample Sharpe ratio of 2, we let $\boldsymbol{\beta} = k \mathbf{1}_{m,p}$ for a suitable constant k and the matrix $\mathbf{1}_{m,p} \in \mathbb{R}^{m \times p}$ of all ones.

This is a kind of worst-case scenario for overfitting, because independent signals and returns introduce a maximal amount of freedom into the model. In this case the replication ratio becomes

$$\frac{\text{SR}_{\text{EOOS}}}{\text{SR}_{\text{EIS}}} = \frac{k^2 pm}{k^2 pm + \frac{pm}{T_1}} \frac{\sqrt{2(k^2 pm)^2 + (c_1 + \tilde{c}_1)k^2 pm + c_2 + \tilde{c}_2}}{\sqrt{2(k^2 pm)^2 + c_1k^2 pm + c_2}},$$
(3.4)

for the constants c_i , \tilde{c}_i defined in (3.1). Whence, for independent signals and residuals, we obtain a similar expression as in the univariate case (3.3).

Figure 6 shows that, conditional on constant in-sample Sharpe ratio $SR_{EIS} = 3$, the replication ratio (3.4) declines more quickly in p as the number of assets m increases. Why does this happen? For each additional asset or signal we increase the freedom in the model to overfit to the historical period, therefore reducing the replication ratio.

One way to address this is to reduce the number of model parameters, e.g., by running a pooled regression that assumes that, up to some suitable normalisations, a family of common signals affects all assets in the same manner. For example, [GP13] normalise their momentum signals for each asset to make them comparable in the cross-section and then only fit one parameter across all stocks for each of these signals, rather than many stock-specific coefficients. While it is entirely possible that the assets have differing true reactions to stimuli (for example, stocks in different sectors may respond differently to the same economic shock), in many cases this fact is overruled by the benefit that can be gained through reducing the number of parameters and avoiding overfitting. A concrete example of this is provided in Section 5 where we observe that a panel regression increases the replication ratio from 71% to 30% for a ten-year backtest.

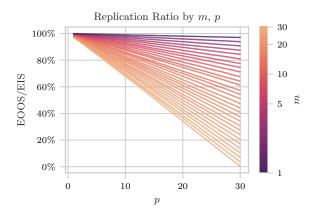


Figure 6: Replication ratio $\frac{\text{SR}_{\text{EOOS}}}{\text{SR}_{\text{EIS}}}$ for increasing assets m and signals p, and fixed in-sample Sharpe ratio $\text{SR}_{\text{EIS}} = 3$ with $T_1 = 252 \cdot 100$. Large T_1 is required to keep SR_{EIS} fixed at 3 as p, m increase, with smaller T_1 we cannot always find β such that $\text{SR}_{\text{EIS}} = 3$.

Importantly, the above discussion is linked to the Sharpe ratio being held constant as the number of assets m increases. Generally, we observe that Sharpe ratios increase as we add additional uncorrelated assets to strategies, primarily due to diversification. To test when this additional performance boost outstrips the overfitting effect we consider a simple case with one signal and an increasing number of assets that the signal works equally well on. Figure 7 shows that if these assets are completely independent, then adding additional assets results in both an increased replication ratio and out-of-sample Sharpe ratio. However, even with just 1% average correlation between assets we see that the replication ratio already becomes monotonically decreasing. For such a low level of correlation between the assets, the out-of-sample Sharpe ratio does initially increase as

⁶ This is reminiscent to the fitting of price impact models where, after normalising for each asset's volatility and trading volume, universal models also often outperform stock-specific models out of sample, as in [MWW24] and references therein.

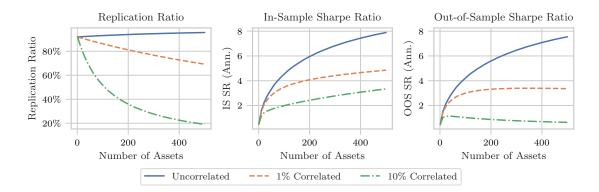


Figure 7: Replication ratio, IS Sharpe ratio and OOS Sharpe ratio as assets increase under different correlations

more assets are added because diversification more than compensates for overfitting. However, already with 10% correlation this effect is reversed. In summary, adding assets generally is bad for your replication ratio. But if you can find some lowly correlated assets with consistent predictive power, it is certainly a benefit to the out-of-sample performance to include them.

3.3 Unknown Covariances

In Proposition 3.1 we have assumed that the true covariance matrix Σ_{ϵ} is known when we construct the portfolio $\widehat{\mathbf{w}}_t$. Of course, in practice we must also estimate this parameter from the observed data. This is not an easy task, and much has been written on how one can improve your estimation of the covariance matrix of returns [BBP17]. However, the corresponding estimation errors are usually of secondary importance compared to misestimation of the regression parameter β .

When the covariance matrix is also estimated, it is no longer feasible to obtain analytical results on the replication ratio. However, it is straightforward to assess this in a simulation study. To this end we perform a Monte Carlo simulation with 30,000 samples where, as in Section 3.1, we select random parameters centred around the empirical point estimates from Section 5 (cf. Section D for more details) for varying numbers of signals p and assets m, along with varying lookback windows T_1 . We form portfolios using either the known precision matrix Σ_{ϵ}^{-1} or the estimated precision matrix from the observed residuals $\widehat{\Sigma}_{\epsilon}^{-1}$.

In the high-dimensional regime where the dimension m exceeds the effective sample size $T_{\rm eff} = T_1 - 1$, the sample covariance matrix $\hat{\Sigma}_{\epsilon}$ is rank-deficient and singular. To ensure the matrix is invertible we use a simple trace-preserving linear shrinkage estimator. We define the regularised covariance matrix as:

$$\widehat{\mathbf{\Sigma}}_{\text{reg}} = \alpha \widehat{\mathbf{\Sigma}}_{\epsilon} + \gamma \mathbf{I}_m,$$

where \mathbf{I}_m is the identity matrix. The shrinkage intensity α and the scaling factor γ are determined by the ratio $q = m/T_{\text{eff}}$. When q > 1, we set $\alpha = 1/q$ and $\gamma = (1 - 1/q)\bar{\lambda}$, where $\bar{\lambda} = \text{tr}(\widehat{\Sigma}_{\epsilon})/m$ represents the average eigenvalue. This approach preserves the trace (total variance of the residuals) while giving us an invertible covariance matrix Σ_{reg} . Note that when q < 1 we simply use the estimated covariance matrix directly.

Figure 8 displays the in-sample Sharpe ratio, the out-of-sample Sharpe ratio and the replication ratio for m=20 and m=500. In the first case, where the asset universe is small, results using the estimated and true covariance matrices are nearly identical. In the high-dimensional setting (m=500), the in-sample Sharpe ratio is inflated further when using the estimated covariance matrix instead of the true one. Regarding the out-of-sample Sharpe ratios, we see that when the number T of time steps observed is less than the number of assets (m=500), i.e., where the covariance matrix is rank-deficient) then the out-of-sample Sharpe ratio is much worse with the estimated covariance matrix (almost zero). However, once this is resolved (T>m), the estimated portfolio actually slightly outperforms the one constructed with the true covariance

matrix. Moreover, as long as T > m, the replication ratio is very well aligned even when m is large.

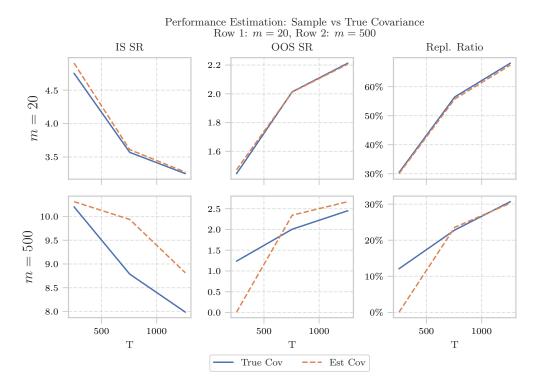


Figure 8: In- and out-of-sample Sharpe ratio and replication ratio using either the true or estimated covariance matrices Σ_{ϵ} , $\widehat{\Sigma}_{\epsilon}$ to form $\widehat{\mathbf{w}}_{t}$ in Monte Carlo simulations.

3.4 Autoregressive Signals

So far, we assumed for analytical tractability that signals and residuals are iid. A typical assumption (in [GP13] for example) is that instead the signal follows an AR(1) process, so that signals are persistent and investment opportunities in turn autocorrelated across time. Such persistent signals are not analytically tractable in our framework, but we can assess the corresponding replication ratios via simulations. We examine the simplest case with one asset and one AR(1) signal,

$$r_{t+1} = \beta s_t + \epsilon_{t+1}, \quad s_t = \phi s_{t-1} + u_t,$$

where $\epsilon_{t+1} \sim \mathcal{N}(0,1)$ and u_t is a centred white noise process with unit variance.⁷ In this case, the true Sharpe ratio is

$$SR = \frac{\mathbb{E}[w_t r_{t+1}]}{\sqrt{\mathbb{V}[w_t r_{t+1}]}} = \frac{\beta^2 \sigma_s^2}{\sqrt{2\beta^4 \sigma_s^4 + \beta^2 \sigma_s^2}} = \frac{\beta}{\sqrt{1 + 2\beta^2 - \phi^2}},$$
 (3.5)

where we have taken $w_t = \beta s_t$, κ_s is the kurtosis (not excess kurtosis) of the signal's noise and assumed $\beta > 0$ without loss of generality. For the second equality in (3.5), we have used that $\kappa_s = 3$ for u_t normally distributed, and $\sigma_s^2 = \frac{1}{1-\phi^2}$.

To understand the impact of the persistence of AR(1) signal compared to their iid counterparts, we proceed by Monte Carlo simulations. We let $T_2 = 1260$ and vary $T_1 \in \{252, 1260, 2520\}$, $\phi \in \{0, 0.5, 0.9, 0.99, 0.999\}$ and SR $\in [0, 2]$. Note that $\phi = 0.9$ corresponds to a signal half-life of 6.5 (daily) time-steps, whereas $\phi = 0.999$ implies a half-life of 690 time-steps. For a given

⁷ We assume that $|\phi| < 1$ so that (s_t) is weakly stationary thus $\mathbb{E}[s_t] = 0$ and $\mathbb{V}[s_t] = \frac{1}{1-\phi^2}$.

signal persistence ϕ we vary the parameter β to achieve the desired true Sharpe ratio in each case using (3.5).

In Figure 9 we compare the replication ratios (and their difference to their counterparts computed in our model with iid signals) for different signal persistences parametrised by ϕ . Unless ϕ is very close to 1 (> 0.99), the difference compared to the iid case $\phi = 0$ is surprisingly small. Thus, although our assumption of iid signals may initially seem like a strong one, it in fact has little impact on the replication ratio in practice.

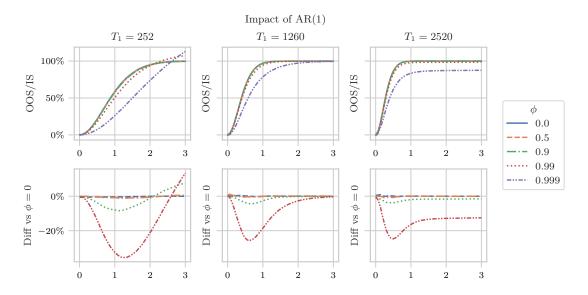


Figure 9: Impact of s_t following an AR(1) process. The top row shows the replication ratio, and the bottom row shows the simple difference of this ratio vs $\phi = 0$, corresponding to the iid case.

4 Comparison with the Literature

Most closely related to the present study is the work of Kan, Wang, and Zheng [KWZ24]. They compare the expected in- and out-of-sample Sharpe ratios of the Markowitz portfolio for m assets whose expected returns are estimated on a training dataset of size T. They find that the expected in-sample Sharpe ratio is

$$\mathbb{E}[\widehat{\theta}] = \frac{\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{T-m-1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{T-m}{2}\right)} {}_{1}F_{1}\left(-\frac{1}{2}; \frac{m}{2}; -\frac{m\theta^{2}}{2}\right),$$

where $\Gamma(\cdot)$ is the Gamma function, ${}_1F_1(\cdot;\cdot;\cdot)$ the confluent hypergeometric function, and $\theta = \sqrt{\mu^T \Sigma^{-1} \mu}$ the true Sharpe ratio of the Markowitz portfolio. Likewise, the expected out-of-sample Sharpe ratio can also be computed in closed form as

$$\mathbb{E}[\tilde{\theta}] = \frac{\theta^2 \sqrt{T} \Gamma\left(\frac{m+1}{2}\right) \Gamma\left(\frac{T-m+2}{2}\right) \Gamma\left(\frac{T}{2}\right)}{\sqrt{2} \Gamma\left(\frac{m+2}{2}\right) \Gamma\left(\frac{T-m+1}{2}\right) \Gamma\left(\frac{T+1}{2}\right)} {}_1F_1\left(\frac{1}{2}; \frac{m+2}{2}; -\frac{T\theta^2}{2}\right).$$

In the framework of [GP13], the setting of [KWZ24] corresponds to using a single, infinitely persistent trading signal $(\phi \to 1)$ for each asset. In contrast, we consider multiple signals per asset that decay quickly $(\phi \to 0)$. Incorporating multiple signals per asset allows us to cover many common trading strategies, such as momentum strategies based on several different moving averages of past returns. Figure 10 illustrates how overfitting becomes an increasingly important

concern when allowing for additional freedom in the model by considering multiple signals per asset. We fix $T = T_1 = T_2 = 2520$, SR = 1.5, vary the number of assets m in both models and set either p = 1 or p = 10 in our model. We observe that for a single signal (p = 1) our haircuts are similar to the ones of [KWZ24], but for p = 10 signals our haircut is much more severe. Note that this is conditional on a fixed true Sharpe ratio, and therefore the strength of each signal must be lower as p increases; in practice, adding signals involves a trade-off between increasing the true SR and raising the risk of overfitting.

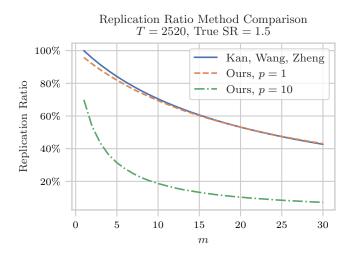


Figure 10: Comparison of the replication ratio for Kan, Wang, and Zheng and ours with $T = T_1 = T_2 = 2520$ and SR = 1.5 for varying m. For ours we set p = 10 or p = 1.

An important common conclusion of both [KWZ24] and the present study is the non-linear dependence of the replication ratio on the true Sharpe ratio. To wit, for small values of the true Sharpe ratio, the expected out-of-sample Sharpe ratio is much lower than the expected in-sample Sharpe ratio. A takeaway from this is that when developing investment strategies, it is preferable to develop a single high Sharpe strategy than many low Sharpe strategies. This aligns with the accepted wisdom of successful practitioners, for example Peter Muller (Founder, PDT Partners) who quipped [Mul01]:

In my opinion it is far better to refine an individual strategy...than to attempt to put together lots of weaker strategies. Depth is more important than breadth for investment 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.

5 Predicting Commodity Futures

To test our method we follow the approach of Gârleanu and Pedersen [GP13] in the empirical application section of their seminal paper on dynamic trading. We select the commodity futures in Table 1 and collect close prices for the closest-to-expiry contracts from May 1998 to December 2023. The daily returns are computed as close-to-close percent changes in prices.

As in [GP13], we compute the 5-day, 1-year and 5-year Sharpe ratios for each asset and use these as signals. In effect these are normalised versions of the moving averages typically used in momentum-style strategies. We additionally transform the signals to be mean zero and unit variance. Normalizing the variances enables us to easily apply L^2 -regularisation when fitting β to calibrate the model we later use as a simulation engine; the centring is required per our conditions on the signals in the theoretical model. This yields a total of 36 dynamic (cross-sectional) signals, which with the addition of the asset-specific intercept yields 37 regressors per asset.

Commodity	Venue	Ticker
Aluminium	LME	MAL3=LX
Copper	LME	MCU3=LX
Nickel	LME	MNI3=LX
Zinc	LME	MZN3=LX
Lead	LME	MPB3=LX
Gasoil	ICE	LGOc1
WTI	NYMEX	CLc1
Gold	COMEX	GCc1
Silver	COMEX	SIc1
Coffee	NYBOT	KCc1
Cocoa	NYBOT	CCc1
Sugar	NYBOT	SBc1

Table 1: Commodities selected for testing from the LME (London Metal Exchange), ICE (Intercontinental Exchange), NYMEX (New York Mercantile Exchange), COMEX (Commodity Exchange) and NYBOT (New York Board of Trade).

To assess the validity of our results we need to compare our analytical approximations for the expected in-sample and out-of-sample Sharpe ratios, SR_{EIS} and SR_{EOOS} , to similarly averaged observed equivalents. This requires more than the one realisation of the returns available historically. To resolve this, we simulate many realisations of a fitted model and use these to check our results.

The first step in enabling this simulation is to estimate a model for the returns, which naturally we take as

$$\mathbf{r}_{t+1} = \beta \mathbf{s}_t + \boldsymbol{\epsilon}_{t+1}.$$

We estimate $\beta \in \mathbb{R}^{37 \times 12}$ by ridge regression with $\gamma = 0.1$. We additionally estimate the covariance of the signals Σ_s and residuals Σ_ϵ from the historical data (noting that the means are zero by construction).

To simulate non-iid and non-Gaussian signals and residuals we also fit an AR(1) process to the historically observed signals,

$$\mathbf{s}_t' = \Phi \mathbf{s}_{t-1}' + \mathbf{u}_t,$$

where $\mathbf{s}_t' \in \mathbb{R}^{36}$ denotes the non-intercept signals, $\mathbf{u}_t \in \mathbb{R}^{36}$ denotes the shocks to the AR(1) process, and $\Phi \in \mathbb{R}^{36 \times 36}$ denotes the AR(1) parameter.

Additionally, we consider the impact of ϵ_{t+1} and \mathbf{u}_t not being multivariate Gaussian, by fitting a t-distribution to each dimension of the residuals and shocks and using these to simulate residuals and shocks with fat tails. To induce the original covariance structure of the residuals and shocks we scale the independent samples by the Cholesky transformation of the relevant covariance matrix, Σ_{ϵ} or Σ_{u} .

With the fitted models we simulate 10,000 realisations of the signals and residuals under both the Gaussian IID model, and with the AR(1) with fat tails model. All fitted parameter values from this section can be found on the authors' webpages. For context, we observe that the average correlation between signals is 20%, with most signals exhibiting very low correlation, and a few exhibiting very high correlation. The average correlation between residuals is 28% and the average excess kurtosis for the residuals is 9. The diagonal of the Φ matrix has average 0.88, and the off-diagonal has average 0.02.

As mentioned in the introduction, the haircuts we observe for these trials are quite steep, and they could be improved by reducing the amount of freedom in the model, for example by panel regression. Indeed, making the structural assumption that the normalised momentum signals have universal weights across assets reduces the number of model parameters for m=12 assets and $p=3\times 12+1=37$ signals each from 444 to 3. In simulations where the full model is used to generate the data but the panel regression is used for the portfolio construction, we see that the replication ratio is 70%, much higher than the 38% observed when the portfolio is constructed using the beta from the full regression. This is due to reduced overfitting, causing us to have

lower in-sample Sharpe ratios (0.9 vs 8.2) but more consistent out-of-sample Sharpe ratios (0.6 vs 3.1). On the real data, when we compare panel regression with the full model we observe that the panel portfolio not only has a better replication ratio than the full model (68% vs 9%), but it also has a higher out-of-sample Sharpe ratio (0.5 vs 0.3), a sharp reversal given the latter's substantially better in-sample performance (panel 0.7, full 3.6). This further cements our advice to prefer simpler models wherever possible.

In addition to the study below, we conduct an additional study on the impact of using the estimated covariance matrix $\hat{\Sigma}_{\epsilon}$ rather than the true covariance matrix Σ_{ϵ} to construct the portfolio, the results of which can be seen in Section E. Broadly, we find that the impact is minimal, which affirms our previous study in Section 3.3.

5.1 Gaussian and iid Signals and Residuals

We first test our approach in the case where the data-generating model satisfies our key assumptions, namely that the signals and residuals are Gaussian and iid.

We simulate 13 years, with 252 trading days per year, of new signals and returns, and use 10 years as in-sample and 3 years as out-of-sample data. For each in-sample simulation we compute the OLS estimator for $\hat{\boldsymbol{\beta}} = \mathbf{R}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}$, where $\mathbf{R} \in \mathbb{R}^{12 \times 252 \cdot 10}$ is the stacked matrix of returns and $\mathbf{S} \in \mathbb{R}^{37 \times 252 \cdot 10}$ is the stacked matrix of signals. We then form the portfolio $\hat{\mathbf{w}}_t = \mathbf{\Sigma}_{\epsilon}^{-1} \hat{\boldsymbol{\beta}} \mathbf{s}_t$ as described in Section 2, utilising the true covariance matrix from the historical data. (As noted in Section 3.3 estimating the covariance matrix as well only has a minor impact. This is confirmed also for this example, see Table 2). This portfolio is then tested on the in-sample and out-of-sample periods to compute the observed in-sample and out-of-sample Sharpe ratios for each simulation.

In Figure 11 we see that the simulated expected in-sample and out-of-sample Sharpe ratios align perfectly with our analytical results, and so does the replication ratio. This shows that with real-world parameter values our analytical approximations are highly accurate. We note here that we have plotted the sample Sharpe ratios directly, and are thus comparing $\mathbb{E}[SR_{IS}]$ with SR_{EIS} and similarly for the out-of-sample Sharpe ratios. For completeness, we report the average mean, variance and other metrics in Table 2. In the next section we additionally check the robustness of our results when our assumptions are violated.

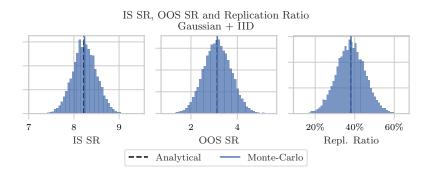


Figure 11: The expected in-sample Sharpe ratio, expected out-of-sample Sharpe ratio and replication ratio for the experiments in Section 5.1.

5.2 AR(1) Signals and t-distributed Residuals and Shocks

Next, we examine how our calculations hold up when the assumptions of the model are no longer true by simulating the returns and signals using the AR(1) model with fat-tail residuals and shocks.

In Figure 12 we see that even when our assumptions are violated, our analytical replication ratio is still remarkably close to its simulation observed counterpart. Relaxing the Gaussian assumption to allow for fat tails causes the in-sample Sharpe ratio to decrease, but also causes the out-of-sample Sharpe ratio to decrease. Additionally, allowing the signals to be AR(1) rather than iid

causes the out-of-sample Sharpe ratio to decrease, however this effect is rather small. Finally, as the out-of-sample and in-sample Sharpe ratio are affected by the non-Gaussian returns, and the AR(1) signals have a relatively small effect, the replication ratio with both is almost identical to our analytical value under the Gaussian and iid assumptions. This confirms the practical relevance of our analytical results.

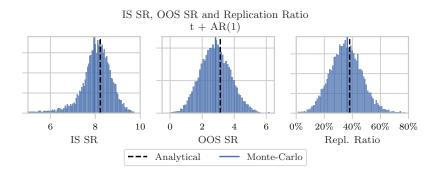


Figure 12: The expected in-sample Sharpe ratio, expected out-of-sample Sharpe ratio and replication ratio for the experiments in Section 5.2.

6 Empirical Study: Goyal, Welch, and Zafirov Dataset

Dataset

As another test of our approach, we use the extensive dataset compiled by Goyal, Welch, and Zafirov [GWZ24]. The authors have constructed monthly, quarterly and annual signals which have been proposed in the academic literature to predict the equity premium of the CRSP US Index, a broad US equity index. We utilise the monthly predictors, of which there are 39 with at least some data from 1926 to 2024, almost 1200 months. These predictors are highly diverse, ranging from macroeconomic indicators, measures of investor sentiment, and variance-related signals to stock cross-sectional data, technical indicators, and commodity prices. For further details, an extensive description of each predictor is given in [GWZ24]. To pre-process these signals and the corresponding returns we wish to predict, we follow the approach of [KMZ24], who also leverage this dataset, and normalise the returns by the rolling 12-month volatility and the signals by their respective expanding volatility.

Empirical Methodology

Our analytical results are derived for the *expected* replication ratio, so to check their validity we must perform some sort of averaging to mitigate the random out- or under-performance of any given signal and sample period. To achieve that on this dataset we proceed as follows.

We repeatedly select random groups of signals of sizes between 1 and 39. We then randomly draw total lengths between 10 and 70 years and split them into (chronologically ordered) in- and out-of-sample periods with both legs at least five years long. Specifically, we perform this sampling 5,000,000 times. We constrain the sampled signals and periods to be such that each signal has full data coverage over the in- and out-of-sample period (as not all signals have data coverage for the full 1926 to 2024 period).

For each draw, we centre the selected predictors over the entire $T_1 + T_2$ window and whiten them using a Cholesky decomposition of their sample covariance. This is useful later on when we wish to estimate β for our analytical values. We subsequently add a column of ones to this whitened design matrix to include an intercept in the regression model.

We then estimate the OLS parameter $\hat{\beta}$ using the sampled signals' in-sample data (per Equation (2.3)) and test this on the in-sample and out-of-sample periods using the estimated Markowitz

portfolio $\widehat{\mathbf{w}}_t = \widehat{\boldsymbol{\Sigma}}_{\epsilon}^{-1} \widehat{\boldsymbol{\beta}} \mathbf{s}_t$, where $\widehat{\boldsymbol{\Sigma}}_{\epsilon}^{-1}$ is also estimated from the observed in-sample residuals. In this case we cannot know the true covariance matrix $\boldsymbol{\Sigma}_{\epsilon}$, but as we are predicting just one asset (the CRSP Index), $\widehat{\boldsymbol{\Sigma}}_{\epsilon}$ is simply the variance of the residuals σ_{ϵ}^2 , which is relatively easy to estimate (in comparison to a large covariance matrix of many assets).

Parameter Estimation for Analytical Values

This process yields pairs of in-sample and out-of-sample Sharpe ratios SR_{IS} and SR_{OOS} . According to our analysis above, the primary drivers of their ratio SR_{OOS}/SR_{IS} should be: the number of signals p, in-sample periods T_1 , and true signal strength β for the strategy. Clearly, the number of signals p and backtest length T_1 are directly available here, but we will need to estimate β . From our analysis throughout the paper, we've shown that estimating β is quite difficult.

However, for our needs, the exact value of each entry of β is less important than having the correct magnitude of β and in turn the correct magnitude of the true Sharpe ratio (see Figure 5). We found in experimentation that using the $\hat{\beta}$ value directly led to the analytical values being poor estimates, and therefore devised the following method to calibrate β as an input to the analytical computations. Specifically, rather than using $\hat{\beta}$ directly, we compute an implied parameter β^* by minimising the difference between the observed out-of-sample Sharpe ratio SR_{OOS} and the expected out-of-sample Sharpe ratio SR_{EOOS}:

$$\boldsymbol{\beta}^{\star} = \arg\min_{\boldsymbol{\beta}} \left(\operatorname{SR}_{\operatorname{OOS}} - \operatorname{SR}_{\operatorname{EOOS}}(\boldsymbol{\mu}_{s}, \boldsymbol{\Sigma}_{s}, \boldsymbol{\Sigma}_{\epsilon}, T_{1}, \boldsymbol{\beta}) \right)^{2}.$$

We perform this calibration separately for each resample. We fix the intercept coefficient within β to zero (assuming we predict day-to-day price moves but not the unconditional drift). For the remaining active signals, $\mu_s = \mathbf{0}_p$ and $\Sigma_s = \mathbf{I}_p$ due to our whitening step. Additionally, Σ_{ϵ} is estimated by $\hat{\Sigma}_{\epsilon}$ from the particular sample, and T_1 is the length of the sample's in-sample period.

For the minimisation, we take a reduced form for β and assume equal predictive power for all the active signals, $\beta = (0, k, k, ...)$. We do this to further reduce the degrees of freedom. We use a root-finding routine to choose the value of k such that the analytical SR_{EOOS} matches the realised out-of-sample Sharpe ratio for each specific trial. Consequently, we view the resulting β^* as a local calibration of the true Sharpe ratio for that particular set of signals and period.

Once we have this implied β , we can use it to estimate the expected in- and out-of-sample mean, variance and Sharpe ratio, and expected replication ratio for that particular set of signals over that particular sampled time period. We note here that in the following analysis, all analytical values use this implied β .

Analysis

We recap what we now have in hand. From the process described above we have obtained many sets of observed in- and out-of-sample empirical and analytical metrics. As discussed, the analytical metrics are expectations, while the empirical metrics are simply observed data points for a particular set of signals over a particular period, which will naturally have measurement error and deviate from their (unknown) expected values.

To resolve this, we look at averages for both the empirical and analytical metrics as a function of our identified key variables: number of signals p, length of backtest T_1 and true Sharpe ratio SR. By computing the averages as a function of these variables across all the samples, the effect of selecting different signals over different periods, and the finite sample measurement error, should average out, and our average empirical values should approach our average analytical values.

The results of this procedure are displayed in Figure 13. In each panel, we plot both the expected ratio of the out-of-sample and in-sample Sharpe ratios, and the ratio of their expectations, and the analytical replication ratio. Generally, we see that all the estimates are very close to each other. In the first panel, we observe that indeed the replication ratio decreases considerably as the number of signals p increases.

In the second panel, we plot the replication ratio as a function of T_1 . We found that this was not smoothly monotonic (perhaps due to insufficient samples), so we plot both the means per

backtest length in our samples, and a smooth fit. We see that in general the three estimates agree, with the replication ratio increasing as the backtest length increases.

The final panel shows the replication ratio as a function of the implied true Sharpe ratio. Again, all three cases show the expected (increasing) behaviour, and the empirical value, which corresponds more closely to our analytical estimate (the ratio of the expectations) is indeed closer to our analytical values than the alternative (expectation of the ratios). In addition to the replication ratio displayed, we report the individual metrics (mean, variance and Sharpe ratio) in Section E.

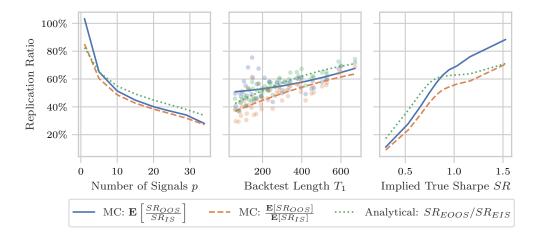


Figure 13: Empirical and implied analytical replication ratio by number of signals p, backtest length T_1 and implied true Sharpe ratio SR (annualised).

7 Conclusion

This paper derives analytical approximations for the in-sample and out-of-sample Sharpe ratios of portfolios constructed using linear prediction models. We show that increasing either the number of signals or assets too much makes this procedure susceptible to overfitting and thereby yields wildly overestimated in-sample Sharpe ratios.

We show that low true Sharpe ratio signals are particularly vulnerable to overfitting. Conversely, by extending the length of the in-sample period one can reduce the overfitting risk, and can produce a higher replication ratio out of sample.

We test our results on commodity futures using momentum-style signals and find that allowing AR(1) signals and non-Normal signals/residuals does not significantly impact the validity of our results. In particular, once we match the theoretical out-of-sample Sharpe ratio to the observed value, we see that the replication ratio is primarily a function of the out-of-sample Sharpe ratio and the curves for the AR(1) signals closely matches those of iid signals.

From this analysis, it seems that the best way to minimise the potential for overfitting is to minimise the number of signals and assets that are being used for any predictive model used to trade, and utilise the largest amount of data possible. To conclude, we quote Nick Patterson [Pat16] on his experience at Renaissance Technologies, which is very much in line with these results:

It's funny that I think the most important thing to do on 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.

They might be on to something.

References

- [BBP17] Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. "Cleaning Large Correlation Matrices: Tools from Random Matrix Theory". In: *Physics Reports*. Cleaning Large Correlation Matrices: Tools from Random Matrix Theory 666 (Jan. 2017), pp. 1–109. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2016.10.005.
- [BKW04] David A. Belsley, Edwin Kuh, and Roy E. Welsch. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity: 546. 1st edition. Hoboken, N.J.: Wiley-Interscience, July 2004. ISBN: 978-0-471-69117-4.
- [BL14] David H. Bailey and Marcos López de Prado. The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality. SSRN Scholarly Paper ID 2460551. Rochester, NY: Social Science Research Network, July 2014. DOI: 10. 2139/ssrn.2460551.
- [BL15] Chris Brightman and Feifei Li. Chasing Performance with ETFs. Tech. rep. Research Affiliates, Nov. 2015.
- [Bro20] Mike Brookes. "The Matrix Reference Manual". http://www.ee.imperial.ac.uk/hp/staff/dmb/matrix/intro.html. 2020.
- [CH88] Samprit Chatterjee and Ali S. Hadi. Sensitivity Analysis in Linear Regression: 190. 1st edition. New York: Wiley, Apr. 1988. ISBN: 978-0-471-82216-5.
- [DL18] Hadrien De March and Charles-Albert Lehalle. "Optimal Trading Using Signals". Nov. 2018. arXiv: 1811.03718 [q-fin].
- [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.
- [GP13] Nicolae Gârleanu and Lasse Heje Pedersen. "Dynamic Trading with Predictable Returns and Transaction Costs". In: *The Journal of Finance* 68.6 (Dec. 2013), pp. 2309–2340. ISSN: 00221082. DOI: 10.1111/jofi.12080.
- [Gup00] A. K. Gupta. *Matrix Variate Distributions*. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics 104. Boca Raton, FL: Chapman & Hall, 2000. ISBN: 978-1-58488-046-2.
- [GWZ24] Amit Goyal, Ivo Welch, and Athanasse Zafirov. "A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction". In: *The Review of Financial Studies* 37.11 (Nov. 2024), pp. 3490–3557. ISSN: 0893-9454. DOI: 10.1093/rfs/hhae044.
- [Har06] Frank E. Harrell. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. 2nd ed. edition. New York, NY: Springer-Verlag New York Inc., Jan. 2006. ISBN: 978-0-387-95232-1.
- [HL15] Campbell R. Harvey and Yan Liu. *Backtesting*. SSRN Scholarly Paper ID 2345489. Rochester, NY: Social Science Research Network, July 2015. DOI: 10.2139/ssrn. 2345489.
- [HLZ16] Campbell R. Harvey, Yan Liu, and Heqing Zhu. "... and the Cross-Section of Expected Returns". In: *The Review of Financial Studies* 29.1 (Jan. 2016), pp. 5–68. ISSN: 0893-9454. DOI: 10.1093/rfs/hhv059.
- [Ioa05] John P. A. Ioannidis. "Why Most Published Research Findings Are False". In: PLOS Medicine 2.8 (Aug. 2005), e124. ISSN: 1549-1676. DOI: 10.1371/journal.pmed. 0020124.
- [JKP23] Theis Ingerslev Jensen, Bryan Kelly, and Lasse Heje Pedersen. "Is There a Replication Crisis in Finance?" In: *The Journal of Finance* 78.5 (2023), pp. 2465–2518. ISSN: 1540-6261. DOI: 10.1111/jofi.13249.

- [Jou+24] Jacques Joubert et al. "The Three Types of Backtests". SSRN Scholarly Paper. Rochester, NY, July 2024. DOI: 10.2139/ssrn.4897573.
- [KMZ24] Bryan Kelly, Semyon Malamud, and Kangying Zhou. "The Virtue of Complexity in Return Prediction". In: *The Journal of Finance* 79.1 (Feb. 2024), pp. 459–503. ISSN: 1540-6261. DOI: 10.1111/jofi.13298.
- [KWZ24] Raymond Kan, Xiaolu Wang, and Xinghua Zheng. "In-Sample and out-of-Sample Sharpe Ratios of Multi-Factor Asset Pricing Models". In: *Journal of Financial Economics* 155 (May 2024), p. 103837. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2024. 103837.
- [LKJ09] Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". In: *Journal of Multivariate Analysis* 100.9 (Oct. 2009), pp. 1989–2001. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2009.04.008.
- [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.
- [MWW24] Johannes Muhle-Karbe, Zexin Wang, and Kevin Webster. "Stochastic Liquidity as a Proxy for Nonlinear Price Impact". In: *Operations Research* 72.2 (Mar. 2024), pp. 444–458. ISSN: 0030-364X. DOI: 10.1287/opre.2022.0627.
- [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.
- [PP12] Kaare Brandt Petersen and Michael Syskind Pedersen. "Matrix Cookbook". In: (Nov. 2012).
- [Rao09] C. Radhakrishna Rao. Linear Statistical Inference and Its Applications. John Wiley & Sons, Sept. 2009. ISBN: 978-0-470-31714-3.
- [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.
- [Wie+16] Thomas Wiecki et al. All That Glitters Is Not Gold: Comparing Backtest and Outof-Sample Performance on a Large Cohort of Trading Algorithms. SSRN Scholarly Paper ID 2745220. Rochester, NY: Social Science Research Network, Mar. 2016. DOI: 10.2139/ssrn.2745220.

A Derivation of Equation (2.2)

To derive (2.2) we need to derive the expressions for

$$\mathbb{E}[(\mathbf{Z}\beta\mathbf{s}_t)^\mathsf{T}(\beta\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})]$$
 and $\mathbb{V}[(\mathbf{Z}\beta\mathbf{s}_t)^\mathsf{T}(\beta\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})].$

We start with the former:

$$\mathbb{E}[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}(\boldsymbol{\beta}\mathbf{s}_{t} + \boldsymbol{\epsilon}_{t+1})] = \mathbb{E}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t} + \mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}],$$

$$= \mathbb{E}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}],$$

$$= \operatorname{tr}(\mathbf{G}\boldsymbol{\Sigma}_{s}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\mu}_{s},$$

where $\mathbf{G} := \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta}$. We have used $\mathbb{E}[\boldsymbol{\epsilon}_{t+1}] = 0$ in the second equality, and identity 5.24 from [Bro20] in the final equality.

Next we tackle the variance,

$$V[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}(\boldsymbol{\beta}\mathbf{s}_{t} + \boldsymbol{\epsilon}_{t+1})] = V[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t} + \mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}],$$

$$= V[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}] + V[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}] + \operatorname{Cov}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}, \mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}],$$

$$= V[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}] + V[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}],$$

$$= 2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + V[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}],$$

$$= 2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + \mathbb{E}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}] - \mathbb{E}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}]^{2},$$

$$= 2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + \mathbb{E}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\epsilon}_{t+1}],$$

$$= 2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + \mathbb{E}[\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}],$$

$$= 2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_{s}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s},$$

where $\mathbf{F} := \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta}$. We have used $\mathbb{E}[\boldsymbol{\epsilon}_{t+1}] = 0$ to deduce $\text{Cov}[\mathbf{s}_t^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta} \mathbf{s}_t, \mathbf{s}_t^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\epsilon}_{t+1}] = 0$, identity 379 from [PP12, sec. 8.2.2] to compute $\mathbb{V}[\mathbf{s}_t^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta} \mathbf{s}_t]$, identity 322 from [PP12, sec. 6.2.2] to compute the penultimate line and identity 5.24 from [Bro20] in the final equality.

Thus,

$$SR = \frac{\operatorname{tr}(\mathbf{G}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{G}\boldsymbol{\mu}_s}{\sqrt{2\operatorname{tr}((\mathbf{G}\boldsymbol{\Sigma}_s)^2) + 4\boldsymbol{\mu}_s^\mathsf{T}\mathbf{G}\boldsymbol{\Sigma}_s\mathbf{G}\boldsymbol{\mu}_s + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{F}\boldsymbol{\mu}_s}},$$

where

$$\mathbf{F} \coloneqq \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_\epsilon \mathbf{Z} \boldsymbol{\beta} \quad \text{and} \quad \mathbf{G} \coloneqq \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta}.$$

B General Case Result of Proposition 3.1

Proposition B.1. The expected average in-sample P&L is

$$\mathbb{E}\left[\widehat{\mathbb{E}}[(\widehat{\mathrm{PnL}}_t)_{t\in\mathcal{T}_1}]\right] = \mathrm{tr}(\mathbf{G}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{G}\boldsymbol{\mu}_s + \frac{p}{T_1}\,\mathrm{tr}(\mathbf{Z}\boldsymbol{\Sigma}_\epsilon).$$

The expected variance of the in-sample P&L is

$$\begin{split} &\mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\mathbf{PnL}}_{t})_{t\in\mathcal{T}_{1}}\right]\right] \\ &= 2\operatorname{tr}\left((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}\right) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_{s}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s} \\ &+ \frac{3}{T_{1} - p - 1}\operatorname{tr}\left(2\mathbf{F}(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s} \cdot (\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + p\mathbf{F}\boldsymbol{\Sigma}_{s}\right) \\ &+ \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})\left(\frac{p(p + T_{1} + 4)}{T_{1}(T_{1} + 2)} - \frac{2\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\left(p - T_{1} - 2\right)\left(p - T_{1}\right)}{T_{1}^{2}\left(T_{1} + 1\right)\left(T_{1} + 2\right)}\right) \\ &- \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})^{2} \frac{2(p - T_{1})\left(\left(p - 2\right)\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s} + T_{1}\left(p - \mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\right) + p\right)}{T_{1}^{2}\left(T_{1} + 1\right)\left(T_{1} + 2\right)} \\ &+ \frac{2}{T_{1} - p - 1}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\operatorname{tr}\left(2\mathbf{G}\left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\mu}_{s} \cdot (\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + p\mathbf{G}\boldsymbol{\Sigma}_{s}\right) \\ &- \frac{2p}{T_{1}}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\left(\operatorname{tr}(\mathbf{G}\boldsymbol{\Sigma}_{s}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\mu}_{s}\right) \\ &+ \varepsilon_{1} + \varepsilon_{2} + \varepsilon_{3}, \end{split}$$

where $\mathbf{F} = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z} \boldsymbol{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta}, \ \mathbf{G} = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z} \boldsymbol{\beta} \ and$

$$\varepsilon_{1} = \operatorname{tr}\left(\operatorname{Cov}\left[\left(\mathbf{S}\mathbf{S}^{\mathsf{T}}\right)^{-1}, \mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right]\right),
\varepsilon_{2} = \operatorname{tr}\left(\operatorname{Cov}\left[\left(\mathbf{S}\mathbf{S}^{\mathsf{T}}\right)^{-1}, \mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right]\right),
\varepsilon_{3} = \operatorname{tr}\left(\operatorname{Cov}\left[\left(\mathbf{S}\mathbf{S}^{\mathsf{T}}\right)^{-1}, \mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right]\right),$$

for any $t \in \mathcal{T}_1$ by the iid property of $(\mathbf{s}_t)_{t \in \mathcal{T}_1}$. The expected average out-of-sample P&L is

$$\mathbb{E}\left[\widehat{\mathbb{P}}\left[(\widehat{\mathrm{PnL}}_t)_{t\in\mathcal{T}_2}\right]\right] = \operatorname{tr}(\mathbf{G}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^{\mathsf{T}}\mathbf{G}\boldsymbol{\mu}_s,$$

The expected variance of the out-of-sample P EL is

$$\begin{split} \mathbb{E}\left[\widehat{\mathbb{V}}\left[(\widehat{\mathbf{P}}\widehat{\mathbf{nL}}_t)_{t\in\mathcal{T}_2}\right]\right] \\ &= 2\operatorname{tr}\left((\mathbf{G}\boldsymbol{\Sigma}_s)^2\right) + 4\boldsymbol{\mu}_s^\mathsf{T}\mathbf{G}\boldsymbol{\Sigma}_s\mathbf{G}\boldsymbol{\mu}_s + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_s) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{F}\boldsymbol{\mu}_s \\ &+ \boldsymbol{\mu}_s^\mathsf{T}\left(2\mathbf{D}\boldsymbol{\Sigma}_s^\mathsf{T}\mathbf{F} + \operatorname{tr}\left(\mathbf{F}\boldsymbol{\Sigma}_s^\mathsf{T}\right)\mathbf{D} + \operatorname{tr}\left(\boldsymbol{\Sigma}_s^\mathsf{T}\mathbf{D}\right)\mathbf{F}\right)\boldsymbol{\mu}_s + \operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_s^\mathsf{T}\mathbf{F}\boldsymbol{\Sigma}_s\right) + \operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_s\right)\operatorname{tr}\left(\mathbf{F}\boldsymbol{\Sigma}_s^\mathsf{T}\right) \\ &+ \operatorname{tr}\left((\mathbf{Z}\boldsymbol{\Sigma}_\epsilon)^2\right)\left(\operatorname{tr}\left(\boldsymbol{\Sigma}_s^\mathsf{T}\mathbf{D}\right) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{D}\boldsymbol{\mu}_s\right), \end{split}$$

$$where \ \mathbf{D} \coloneqq \frac{\left(\boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s\boldsymbol{\mu}_s^\mathsf{T}\right)^{-1}}{T_1 - n - 1}. \end{split}$$

C Proof of Proposition B.1

Recall that the classical Ordinary Least Square method yields the estimator

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

where $\mathbf{E} \in \mathbb{R}^{m \times T_1}$ is the matrix of stacked residuals. This is a more convenient expression to work with than (2.3), and we will use it to evaluate the expectations of the sample means, variances and covariances. Since $(\mathbf{s}_t)_t$ and $(\boldsymbol{\epsilon}_{t+1})_t$ are iid sequences, the expectation of the sample expectation, variance and covariance are simply the true expectation, variance and covariance respectively.

C.1 In-sample expected average

From the definition (2.4), the in-sample expected $P\&L \mathbb{E}[\widehat{\mathbb{E}}[(\widehat{PnL}_t)_{t\in\mathcal{T}_1}]]$ can be evaluated as

$$\begin{split} & \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}(\boldsymbol{\beta}\mathbf{s}_{t}+\boldsymbol{\epsilon}_{t+1})\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ & = \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\{\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t}\}^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] + \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\{\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t}\}^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right], \\ & = \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\left\{\mathbf{Z}\left(\boldsymbol{\beta}+\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right)\mathbf{s}_{t}\right\}^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right], \\ & + \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\left\{\mathbf{Z}\left(\boldsymbol{\beta}+\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right)\mathbf{s}_{t}\right\}^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right], \\ & = \mathbb{E}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right] + \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left\{\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right\}_{t\in\mathcal{T}_{1}}\right]\right] \\ & + \mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\left\{\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right\}^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right], \end{split}$$

since the sample mean is an unbiased estimator, and since $(\mathbf{s}_t)_t$ and $(\epsilon_{t+1})_t$ are respectively iid, and $\mathbb{E}[\epsilon_{t+1}] = \mathbf{0}$.

Since we can "extract" the t-th column from a matrix \mathbf{X} by multiplying it by a vector \mathbf{e}_t which has 1 in the t-th position and zero everywhere else, then

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\left(\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t\right)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_1}\right]\right] = \mathbb{E}\left[\mathbf{e}_t^\mathsf{T}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{S}\mathbf{E}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_t\right] = 0, \quad \text{for all } t\in\mathcal{T}_1,$$

since **E** has a Matrix Normal distribution $\mathcal{MN}(\mathbf{0}, \mathbf{\Sigma}_{\epsilon}, \mathbf{I}_{T_1})$ by assumption and since **Z** is assumed symmetric. As previously mentioned, the first equality is true for all $t \in \mathcal{T}_1$ by the iid property of $(\mathbf{s}_t)_t$, and the fact the sample mean is an unbiased estimator.

The next term reads

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left(\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] = \mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\mathbf{E}\mathbf{e}_{t}\right], \quad \text{for any } t\in\mathcal{T}_{1},$$

$$= \mathbf{e}_{t}^{\mathsf{T}}\mathbb{E}[\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}]\mathbb{E}[\mathbf{E}^{\mathsf{T}}\mathbf{Z}\mathbf{E}]\mathbf{e}_{t},$$

$$= \operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\mathbf{e}_{t}^{\mathsf{T}}\mathbb{E}[\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}]\mathbf{e}_{t} = \frac{p}{T_{1}}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}),$$

where we have used the Matrix Normal identity

$$\mathbb{E}[\mathbf{X}\mathbf{A}\mathbf{X}^\mathsf{T}] = \mathbf{U}\operatorname{tr}(\mathbf{A}^\mathsf{T}\mathbf{V}) + \mathbf{M}\mathbf{A}\mathbf{M}^\mathsf{T},$$

with $\mathbf{X} \sim \mathcal{MN}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ [Gup00, Theorem 2.3.5]. We also used that $\operatorname{tr}(\mathbb{E}[\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}]) = p$ since the projection matrix $\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}$ is idempotent and thus its trace equals its rank, and

since $(\mathbf{s}_t)_t$ are iid, then the expectation of each element of the diagonal of $\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}$ are equal, thus $\mathbf{e}_t^{\mathsf{T}}\mathbb{E}[\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}]\mathbf{e}_t = \frac{p}{T_1}$. Therefore, the in-sample expected P&L is given by

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[(\widehat{\mathrm{PnL}}_t)_{t\in\mathcal{T}_1}\right]\right] = \operatorname{tr}\left(\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\right) + \boldsymbol{\mu}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_s + \frac{p}{T_1}\operatorname{tr}\left(\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\right).$$

C.2 In-sample expected variance

We turn next to the expected variance $\mathbb{E}[\widehat{\mathbb{V}}[(\widehat{PnL}_t)_{t\in\mathcal{T}_1}]]$:

$$\begin{split} &\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}(\boldsymbol{\beta}\mathbf{s}_{t}+\boldsymbol{\epsilon}_{t+1})\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &= \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left(\left(\mathbf{Z}\left(\boldsymbol{\beta}+\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right)\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t} + \left(\mathbf{Z}\left(\boldsymbol{\beta}+\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right)\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right], \\ &= \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t} + (\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t} + (\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1} + (\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right], \\ &= \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right] + \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] + \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] + \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{\mathsf{T}}\mathbf{s}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right)_{t\in\mathcal{T}_{1}}\right)\right] \\ &+ \mathbb{E}\left[\widehat{2}\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}$$

Since the signals and residuals are iid, and the sample variance is an unbiased estimator for the population variance, we can replace all the expected sample variance terms with the true variance. We are thus left with computing these true variances. The first two terms are

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] = \mathbb{V}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right] = 2\operatorname{tr}\left(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\right) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s},$$

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] = \mathbb{V}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right] = \operatorname{tr}\left(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\right) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s},$$

where the first equality in both expressions is for all $t \in \mathcal{T}_1$, and the first equality comes from the iid assumption and the fact the sample variance is an unbiased estimator.

The next term is

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] = \mathbb{V}\left[\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right], \quad \text{for all } t\in\mathcal{T}_{1}$$

$$= \mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{2}\right] - \mathbb{E}\left[\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right]^{2}$$

$$= \mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{2}\right],$$

since $\mathbb{E}[\mathbf{E}] = 0$ and \mathbf{E} and \mathbf{S} are independent. To evaluate this further we again use $\mathbf{s}_t = \mathbf{S}\mathbf{e}_t$ and let $\mathbf{H} := \mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{S}$ (the so called 'hat' matrix in linear regression), so that

$$\begin{split} \mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{2}\right] &= \mathbb{E}\left[\left(\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\right)^{2}\right] \\ &= \mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\right] \\ &= \mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}(\mathbf{I}_{T_{1}}\mathbf{H}^{\mathsf{T}}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\right] \\ &= \mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\right] , \\ &= \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\right) \right) \\ &= \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\right) \right) \\ &= \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right]\mathbb{E}\left[\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\right] + \mathcal{E}_{1}\right), \end{split}$$

where

$$\mathcal{E}_1 := \mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1} \mathbf{S} \mathbf{e}_t \mathbf{e}_t^\mathsf{T} \mathbf{S}^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_\epsilon \mathbf{Z} \boldsymbol{\beta} \mathbf{S} \mathbf{e}_t \mathbf{e}_t^\mathsf{T} \mathbf{S}^\mathsf{T} \right] - \mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1} \right] \mathbb{E}\left[\mathbf{S} \mathbf{e}_t \mathbf{e}_t^\mathsf{T} \mathbf{S}^\mathsf{T} \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_\epsilon \mathbf{Z} \boldsymbol{\beta} \mathbf{S} \mathbf{e}_t \mathbf{e}_t^\mathsf{T} \mathbf{S}^\mathsf{T} \right].$$

The \mathcal{E}_1 term results from the fact that $(\mathbf{SS}^{\mathsf{T}})^{-1}$ is of course not independent of the other terms, however experimentally we find that $\operatorname{tr}(\mathcal{E}_1)$ is negligible compared to the trace of the expectation product and decreases rapidly in the ratio p/T_1 . Provided we are in the regime where $T_1 \gg p$, then this approximation is reasonable as can be seen in Figure 3.

With this, we can finish the evaluation. To simplify the expression it is convenient to swap \mathbf{Se}_t with \mathbf{s}_t again, and define the matrix $\mathbf{F} := \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta}$, so that

$$\begin{split} & \mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{2}\right] \\ & = \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right]\mathbb{E}\left[\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right] + \mathcal{E}_{1}\right) \\ & = \operatorname{tr}\left[\frac{\left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)^{-1}}{T_{1} - p - 1}\left(2\left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)\mathbf{F}\left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s}\left(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right) + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_{s})\left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)\right)\right] + \varepsilon_{1} \\ & = \frac{1}{T_{1} - p - 1}\operatorname{tr}\left(2\mathbf{F}\left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s} \cdot \left(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)^{-1}\left(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right) + p\mathbf{F}\boldsymbol{\Sigma}_{s}\right) + \varepsilon_{1}, \end{split}$$

with $\varepsilon_1 := \operatorname{tr}(\mathcal{E}_1)$. Thus, the expected variance term is given by

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right]$$

$$=\frac{1}{T_{1}-p-1}\operatorname{tr}\left(2\mathbf{F}\left(\boldsymbol{\Sigma}_{s}+\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)+\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s}\cdot\left(\boldsymbol{\Sigma}_{s}+\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)^{-1}\left(\boldsymbol{\Sigma}_{s}-\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}\right)+p\mathbf{F}\boldsymbol{\Sigma}_{s}\right)+\varepsilon_{1}.$$

Moving on, we tackle the next term in the expected variance similarly:

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right] = \mathbb{V}\left[(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right], \text{ for all } t\in\mathcal{T}_{1},$$

$$= \mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)^{2}\right] - \mathbb{E}\left[(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right]^{2}$$

$$= \mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)^{2}\right] - \left(\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})\frac{p}{T_{1}}\right)^{2}.$$

We can evaluate the first term as

$$\mathbb{E}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)^{2}\right] \\
= \mathbb{E}\left[\mathbf{e}_{t}\mathbf{H}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\mathbf{E}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\mathbf{E}\mathbf{H}^{\mathsf{T}}\mathbf{e}_{t}\right] \\
= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})\left(\mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{e}_{t}\right] + \mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{e}_{t}\right]\right) + \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})^{2}\mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{e}_{t}\right] \\
= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})\left(\frac{p}{T_{1}} + \left(\frac{p(p+2)}{T_{1}(T_{1}+2)} - \mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\frac{2(T_{1}-p)(T_{1}-p+2)}{T^{2}(T+1)(T+2)}\right)\right) \\
+ \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})^{2}\left(\frac{p(p+2)}{T_{1}(T_{1}+2)} - \mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\frac{2(T_{1}-p)(T_{1}-p+2)}{T^{2}(T+1)(T+2)}\right),$$

where we have used the fact that when there is no intercept term the diagonal entries of the hat matrix \mathbf{H} are distributed $\beta\left(\frac{p}{2},\frac{(T-p)}{2}\right)$ [BKW04, Appendix 2A], and for the case where there is an intercept term, Chatterjee and Hadi [CH88, Section 2.3.7] derive an approximate distribution for the diagonal elements of the hat matrix,

$$\frac{T_1 - p}{p - 1} \frac{h_{tt} - \frac{1}{T_1}}{1 - h_{tt}} \dot{\sim} F(p - 1, T_1 - p),$$

where $F(p-1, T_1-p)$ denotes the F distribution with p-1 and T_1-p degrees of freedom. However, according to Rao [Rao09, p. 542] this is an exact distribution, rather than an approximate one when $\mu_s = \mathbf{0}$, which holds in our assumptions. With a little work this yields

$$\mathbb{E}[h_{tt}^2] = \frac{p(p+2)}{T_1(T_1+2)} - \frac{2(T_1-p)(T_1-p+2)}{T^2(T+1)(T+2)},$$

and noting that $\mathbf{1}_p^\mathsf{T} \boldsymbol{\mu}_s$ is zero with no intercept, and one with an intercept, this yields the final line of the derivation above. Putting it all together we obtain the expected variance term

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right]$$

$$=\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})\left(\frac{p}{T_{1}}+\left(\frac{p(p+2)}{T_{1}(T_{1}+2)}-\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\frac{2(T_{1}-p)(T_{1}-p+2)}{T^{2}(T+1)(T+2)}\right)\right)$$

$$+\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})^{2}\left(\frac{p(p+2)}{T_{1}(T_{1}+2)}-\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\frac{2(T_{1}-p)(T_{1}-p+2)}{T^{2}(T+1)(T+2)}\right)-\left(\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\frac{p}{T_{1}}\right)^{2}$$

$$=\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})\frac{pT_{1}\left(T_{1}+1\right)\left(p+T_{1}+4\right)-2\left(p-T_{1}-2\right)\left(p-T_{1}\right)\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}}{T_{1}^{2}\left(T_{1}+1\right)\left(T_{1}+2\right)}$$

$$-\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})^{2}\frac{2\left(p-T_{1}\right)\left(\left(p-2\right)\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}+T_{1}\left(p-\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\right)+p\right)}{T_{1}^{2}\left(T_{1}+1\right)\left(T_{1}+2\right)}.$$

Next we begin to tackle the covariance terms, again we use the fact that the sample covariance is an unbiased estimator and the iid property. The first two covariance terms are zero:

$$\mathbb{E}\left[2\widehat{\text{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}},\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right]=2\widehat{\text{Cov}}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t},\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right]=0,$$

since $\mathbb{E}[\boldsymbol{\epsilon}_{t+1}] = \mathbf{0}$, and

$$\mathbb{E}\left[2\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_1},\left((\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_1}\right]\right] = 2\mathrm{Cov}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t,(\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right]$$

is null for all $t \in \mathcal{T}_1$ as well since $\mathbb{E}[\epsilon_{t+1}] = \mathbf{0}$. Regarding the next covariance term,

$$\mathbb{E}\left[2\widehat{\operatorname{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}},\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right]$$

$$= 2\widehat{\operatorname{Cov}}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t},(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right], \text{ for all } t\in\mathcal{T}_{1},$$

$$= 2\left(\mathbb{E}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right] - \mathbb{E}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right]\mathbb{E}\left[(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right]\right).$$

Starting with the first expectation,

$$\begin{split} & \mathbb{E}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1} \right] \\ &= \mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{E}^{\mathsf{T}}\mathbf{Z}\mathbf{E}\mathbf{e}_{t} \right] \\ &= \operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\mathbb{E}\left[\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{H}\mathbf{e}_{t} \right] \\ &= \operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1} \right] \mathbb{E}\left[\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}} \right] + \mathcal{E}_{2} \right) \\ &= \frac{\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\operatorname{tr}\left(2\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s}(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + \boldsymbol{p}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s} \right)}{T_{1} - p - 1} + \varepsilon_{2}, \end{split}$$

where we have proceeded as before with a mixed expectation of $(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}$ and \mathbf{S} , and

$$\mathcal{E}_2 = \mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t\mathbf{s}_t^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t\mathbf{s}_t^\mathsf{T} \right] - \mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1} \right] \mathbb{E}\left[\mathbf{s}_t\mathbf{s}_t^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t\mathbf{s}_t^\mathsf{T} \right],$$

and $\varepsilon_2 := \operatorname{tr}(\mathcal{E}_2)$. Next, the second expectation reads

$$\mathbb{E}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t \right] \mathbb{E}\left[(\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1} \right] = \left(\operatorname{tr}(\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}) + \boldsymbol{\mu}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_s \right) \frac{p}{T_1} \operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_\epsilon).$$

Putting this together yields

$$\mathbb{E}\left[2\widehat{\text{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}},\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right]$$

$$=\frac{2}{T_{1}-p-1}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\operatorname{tr}\left(2\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}(\boldsymbol{\Sigma}_{s}+\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})+\boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s}\cdot(\boldsymbol{\Sigma}_{s}+\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s}-\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})+p\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\right)$$

$$-\frac{2p}{T_{1}}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\left(\operatorname{tr}(\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z})+\boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s}\right)+\varepsilon_{2}.$$

The next covariance term is similar, namely

$$\mathbb{E}\left[2\widehat{\operatorname{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}},\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right] \\
= 2\widehat{\operatorname{Cov}}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1},\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right], \text{ for all } t\in\mathcal{T}_{1}, \\
= 2\left(\mathbb{E}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right] - \mathbb{E}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right]\mathbb{E}\left[\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right] \\
= 2\mathbb{E}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right],$$

since $\mathbb{E}[\epsilon_{t+1}] = \mathbf{0}$. The first expectation can be evaluated as

$$\begin{split} & \mathbb{E}\left[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1} (\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t \right] \\ &= \mathbb{E}\left[\mathbf{e}_t^\mathsf{T}\mathbf{S}^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\operatorname{tr}(\mathbf{e}_t^\mathsf{T}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{S}\mathbf{e}_t)\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_t \right] \\ &= \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1} \right] \mathbb{E}\left[\mathbf{s}_t\mathbf{s}_t^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\mathbf{s}_t^\mathsf{T} \right] + \mathcal{E}_3 \right) \\ &= \frac{1}{T_1 - p - 1}\operatorname{tr}\left(2\mathbf{F}(\boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s\boldsymbol{\mu}_s^\mathsf{T}) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{F}\boldsymbol{\mu}_s \cdot (\boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s\boldsymbol{\mu}_s^\mathsf{T})^{-1}(\boldsymbol{\Sigma}_s - \boldsymbol{\mu}_s\boldsymbol{\mu}_s^\mathsf{T}) + p\mathbf{F}\boldsymbol{\Sigma}_s \right) + \varepsilon_3, \end{split}$$

where again we handle the $(SS^T)^{-1}$ terms as before, and

$$\mathcal{E}_3 = \mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t\mathbf{s}_t^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\mathbf{s}_t^\mathsf{T} \right] - \mathbb{E}\left[(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1} \right] \mathbb{E}\left[\mathbf{s}_t\mathbf{s}_t^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\mathbf{s}_t^\mathsf{T} \right],$$

and $\varepsilon_3 := \operatorname{tr}(\mathcal{E}_3)$. Thus,

$$\mathbb{E}\left[2\widehat{\text{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}},\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{1}}\right]\right]$$

$$=\frac{2}{T_{1}-p-1}\operatorname{tr}\left(2\mathbf{F}(\boldsymbol{\Sigma}_{s}+\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})+\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s}\cdot(\boldsymbol{\Sigma}_{s}+\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s}-\boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})+p\mathbf{F}\boldsymbol{\Sigma}_{s}\right)+\varepsilon_{3},$$

where again $\mathbf{F} = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z} \boldsymbol{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta}$. We turn now to the penultimate term:

$$\mathbb{E}\left[2\widehat{\operatorname{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}},\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{1}}\right]\right]$$

$$= 2\widehat{\operatorname{Cov}}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1},\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right], \text{ for all } t\in\mathcal{T}_{1},$$

$$= 2\left(\mathbb{E}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right] - \mathbb{E}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right]\mathbb{E}\left[\left(\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\right)^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right] = 0,$$

as the derivation yields a cubic in **E** and $\mathbb{E}[\epsilon_{t+1}] = \mathbf{0}$.

For the final term, we can write

$$\begin{split} & \mathbb{E}\left[2\widehat{\text{Cov}}\left[\left((\mathbf{ZES}^\mathsf{T}(\mathbf{SS}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_1},\left((\mathbf{ZES}^\mathsf{T}(\mathbf{SS}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_1}\right]\right] \\ & = 2\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left((\mathbf{ZES}^\mathsf{T}(\mathbf{SS}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t(\mathbf{ZES}^\mathsf{T}(\mathbf{SS}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_1}\right] \\ & -\widehat{\mathbb{E}}\left[\left((\mathbf{ZES}^\mathsf{T}(\mathbf{SS}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_1}\right]\widehat{\mathbb{E}}\left[\left((\mathbf{ZES}^\mathsf{T}(\mathbf{SS}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_1}\right] = 0, \end{split}$$

which again yields terms with only **E** or cubics in **E**, and from the identity for cubic expectations we see that every term in the expectation includes μ_{ϵ} which is zero, and thus the full expectation is zero. We can now collate all the terms for the expected variance $\mathbb{E}[\widehat{\mathbb{V}}[(\widehat{\text{PnL}}_t)_{t\in\mathcal{T}_t}]]$:

$$\begin{split} &\mathbb{E}[\widehat{\mathbb{V}}[(\widehat{\mathbf{P}nL}_{t})_{t\in\mathcal{T}_{1}}]] \\ &= 2\operatorname{tr}\left((\mathbf{G}\boldsymbol{\Sigma}_{s})^{2}\right) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Sigma}_{s}\mathbf{G}\boldsymbol{\mu}_{s} + \operatorname{tr}(\mathbf{F}\boldsymbol{\Sigma}_{s}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s} \\ &+ \frac{3}{T_{1} - p - 1}\operatorname{tr}\left(2\mathbf{F}(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{F}\boldsymbol{\mu}_{s} \cdot (\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + p\mathbf{F}\boldsymbol{\Sigma}_{s}\right) \\ &+ \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}) \frac{pT_{1}\left(T_{1} + 1\right)\left(p + T_{1} + 4\right) - 2\left(p - T_{1} - 2\right)\left(p - T_{1}\right)\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}}{T_{1}^{2}\left(T_{1} + 1\right)\left(T_{1} + 2\right)} \\ &- \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z})^{2} \frac{2\left(p - T_{1}\right)\left(\left(p - 2\right)\mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s} + T_{1}\left(p - \mathbf{1}_{p}^{\mathsf{T}}\boldsymbol{\mu}_{s}\right) + p\right)}{T_{1}^{2}\left(T_{1} + 1\right)\left(T_{1} + 2\right)} \\ &+ \frac{2}{T_{1} - p - 1}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\operatorname{tr}\left(2\mathbf{G}(\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\mu}_{s} \cdot (\boldsymbol{\Sigma}_{s} + \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}})^{-1}(\boldsymbol{\Sigma}_{s} - \boldsymbol{\mu}_{s}\boldsymbol{\mu}_{s}^{\mathsf{T}}) + p\mathbf{G}\boldsymbol{\Sigma}_{s}\right) \\ &- \frac{2p}{T_{1}}\operatorname{tr}(\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})\left(\operatorname{tr}(\mathbf{G}\boldsymbol{\Sigma}_{s}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{G}\boldsymbol{\mu}_{s}\right) \\ &+ \varepsilon_{1} + \varepsilon_{2} + \varepsilon_{3}, \end{split}$$

where

$$\mathbf{F} := \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \mathbf{\Sigma}_{\epsilon} \mathbf{Z} \boldsymbol{\beta} \quad \text{and} \quad \mathbf{G} := \boldsymbol{\beta}^\mathsf{T} \mathbf{Z} \boldsymbol{\beta},$$

and

$$\varepsilon_{1} = \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\right] - \mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right]\mathbb{E}\left[\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\mathbf{S}\mathbf{e}_{t}\mathbf{e}_{t}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\right] \\
\varepsilon_{2} = \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right] - \mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right]\mathbb{E}\left[\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right] , \\
\varepsilon_{3} = \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\mathsf{T}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right] - \mathbb{E}\left[(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\right]\mathbb{E}\left[\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\mathsf{T}\boldsymbol{\beta}\mathbf{s}_{t}\mathbf{s}_{t}^{\mathsf{T}}\right] .$$

C.3 Out-of-sample expected average

Next we derive the expected out-of-sample expected return and variance. We now have that $\widehat{\beta}$ is independent of the realisations of the signal and the residuals, thus the derivation is simplified significantly. Firstly, the expected out-of-sample P&L is given by

$$\mathbb{E}\left[\widehat{\mathbb{E}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_t)^\mathsf{T}(\boldsymbol{\beta}\mathbf{s}_t + \boldsymbol{\epsilon}_{t+1})\right)_{t \in \mathcal{T}_2}\right]\right] = \mathbb{E}[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t] = \operatorname{tr}(\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}) + \boldsymbol{\mu}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_s,$$

where we have no $\mathbb{E}[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}]$ term since $\mathbb{E}[\boldsymbol{\epsilon}_{t+1}] = \mathbf{0}$ and by independence.

C.4 Out-of-sample expected variance

Next the variance is given by

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}(\boldsymbol{\beta}\mathbf{s}_{t}+\boldsymbol{\epsilon}_{t+1})\right)_{t\in\mathcal{T}_{2}}\right]\right]$$

$$=\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}\right]+\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right]+2\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}},\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right]\right].$$

The first term can be evaluated as

$$\begin{split} \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}\right]\right] &= \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}\right] + \widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}\right] \\ &+ 2\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}\right]\right], \end{split}$$

where we must be careful to note that $\mathbf{ES}^{\mathsf{T}}(\mathbf{SS}^{\mathsf{T}})^{-1}$, i.e. the error in $\widehat{\boldsymbol{\beta}}$, are the errors and signals from the first period, not the second. The first term of this is simply given by

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_2}\right]\right] = \mathbb{V}[(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t] = 2\operatorname{tr}(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}) + 4\boldsymbol{\mu}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_s,$$

where the first equality holds for all $t \in \mathcal{T}_2$ since \mathbf{s}_t are iid. The covariance term is zero since $\mathbb{E}[\boldsymbol{\epsilon}_{t+1}] = \mathbf{0}$, so we are left with $\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_2}\right]\right]$. With $\boldsymbol{\varepsilon}_{\beta} \coloneqq \mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}$,

$$\begin{split} & \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_2}\right]\right] = \mathbb{V}\left[(\mathbf{Z}\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right], \quad \text{for all } t\in\mathcal{T}_2, \\ & = \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\left(\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}_s\right) + \boldsymbol{\mu}_s^\mathsf{T}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}_s\left(\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}^\mathsf{T}\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\varepsilon}_{\boldsymbol{\beta}}\right)\boldsymbol{\mu}_s\right]. \end{split}$$

To evaluate this we require some intermediate expectations,

$$\begin{split} \mathbb{E}\left[\boldsymbol{\varepsilon}_{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\varepsilon}_{\beta}^{\mathsf{T}}\mathbf{Z}\right] &= \mathbf{D}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}, \\ \mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\varepsilon}\right] &= \mathbf{D}\operatorname{tr}\left(\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\right), \\ \mathbb{E}\left[\boldsymbol{\varepsilon}_{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\varepsilon}_{\beta}^{\mathsf{T}}\right] &= \boldsymbol{\Sigma}_{\epsilon}\operatorname{tr}\left(\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\mathbf{D}\right), \end{split}$$

where

$$\mathbf{D} \coloneqq rac{\left(\mathbf{\Sigma}_s + oldsymbol{\mu}_s oldsymbol{\mu}_s^\mathsf{T}
ight)^{-1}}{T_1 - p - 1}.$$

Putting this all together yields

$$\begin{split} & \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^\mathsf{T}(\mathbf{S}\mathbf{S}^\mathsf{T})^{-1}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\beta}\mathbf{s}_t\right)_{t\in\mathcal{T}_2}\right]\right] \\ & = \operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\right) + \operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_s\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\right) \\ & + \boldsymbol{\mu}_s^\mathsf{T}\left(2\mathbf{D}\boldsymbol{\Sigma}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta} + \operatorname{tr}\left(\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\right)\mathbf{D} + \operatorname{tr}\left(\boldsymbol{\Sigma}_s^\mathsf{T}\mathbf{D}\right)\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta}\right)\boldsymbol{\mu}_s \end{split}$$

Thus, the first term for the expected variance out-of-sample is

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\beta}\mathbf{s}_{t}\right)_{t\in\mathcal{T}_{2}}\right]\right]$$

$$= 2\operatorname{tr}(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}) + 4\boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s}$$

$$+ \operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\right) + \operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_{s}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\right)$$

$$+ \boldsymbol{\mu}_{s}^{\mathsf{T}}\left(2\mathbf{D}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta} + \operatorname{tr}\left(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\right)\mathbf{D} + \operatorname{tr}\left(\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\mathbf{D}\right)\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\right)\boldsymbol{\mu}_{s}.$$

We can proceed similarly for the second term,

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right]\right] = \mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right] + \widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right] + 2\widehat{\mathrm{Cov}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}, \left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right]\right].$$

Again the first term is simply

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_2}\right]\right] = \mathbb{V}\left[\left(\mathbf{Z}\boldsymbol{\beta}\mathbf{s}_t\right)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}\right] = \operatorname{tr}(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon) + \boldsymbol{\mu}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_s,$$

and the final term is zero as **E** is from the in-sample period, and therefore independent of ϵ_{t+1} from the out-of-sample period, and $\mathbb{E}[\epsilon_{t+1}] = \mathbf{0}$. Thus, we need only to find the second term,

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\mathbf{E}\mathbf{S}^{\mathsf{T}}(\mathbf{S}\mathbf{S}^{\mathsf{T}})^{-1}\mathbf{s}_{t})^{\mathsf{T}}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_{2}}\right]\right] = \mathbb{E}\left[\operatorname{tr}(\mathbf{Z}\boldsymbol{\varepsilon}_{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\varepsilon}_{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\varepsilon}_{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\varepsilon}_{\beta}\boldsymbol{\mu}_{s}\right] \\ = \operatorname{tr}\left((\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})^{2}\right)\left(\operatorname{tr}(\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\mathbf{D}) + \boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{D}\boldsymbol{\mu}_{s}\right).$$

And therefore, the full second term is given by

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_t)^\mathsf{T}\boldsymbol{\epsilon}_{t+1}\right)_{t\in\mathcal{T}_2}\right]\right] = \operatorname{tr}(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_s\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon) + \boldsymbol{\mu}_s^\mathsf{T}\boldsymbol{\beta}^\mathsf{T}\mathbf{Z}\boldsymbol{\Sigma}_\epsilon\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_s + \operatorname{tr}\left((\mathbf{Z}\boldsymbol{\Sigma}_\epsilon)^2\right)\left(\operatorname{tr}(\boldsymbol{\Sigma}_s^\mathsf{T}\mathbf{D}) + \boldsymbol{\mu}_s^\mathsf{T}\mathbf{D}\boldsymbol{\mu}_s\right).$$

The third term, the covariance term, is necessarily zero by $\mathbb{E}[\epsilon_{t+1}] = \mathbf{0}$ and independence, thus the expected out-of-sample variance is given by

$$\mathbb{E}\left[\widehat{\mathbb{V}}\left[\left((\mathbf{Z}\widehat{\boldsymbol{\beta}}\mathbf{s}_{t})^{\mathsf{T}}(\boldsymbol{\beta}\mathbf{s}_{t}+\boldsymbol{\epsilon}_{t+1})\right)_{t\in\mathcal{T}_{2}}\right]\right]$$

$$=2\operatorname{tr}(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}})+4\boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s}$$

$$+\operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\right)+\operatorname{tr}\left(\mathbf{D}\boldsymbol{\Sigma}_{s}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\right)$$

$$+\boldsymbol{\mu}_{s}^{\mathsf{T}}\left(2\mathbf{D}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}+\operatorname{tr}\left(\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\right)\mathbf{D}+\operatorname{tr}\left(\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\mathbf{D}\right)\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\right)\boldsymbol{\mu}_{s}$$

$$+\operatorname{tr}(\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\Sigma}_{s}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})+\boldsymbol{\mu}_{s}^{\mathsf{T}}\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon}\mathbf{Z}\boldsymbol{\beta}\boldsymbol{\mu}_{s}+\operatorname{tr}\left((\mathbf{Z}\boldsymbol{\Sigma}_{\epsilon})^{2}\right)\left(\operatorname{tr}(\boldsymbol{\Sigma}_{s}^{\mathsf{T}}\mathbf{D})+\boldsymbol{\mu}_{s}^{\mathsf{T}}\mathbf{D}\boldsymbol{\mu}_{s}\right).$$

D Monte Carlo Details

As previously mentioned, we use Monte Carlo simulations both to check that the approximations (3.2) are reasonable ($\varepsilon_i = 0$ and ignoring the convexity adjustment) and to support and provide additional analysis. To do this, we observe the point estimates of the parameters from the experiments using commodity futures in Section 5 and select distributions to draw new realisations within a cloud around these point estimates. Specifically,

- For the signals we let the standard deviation of the signals be 1, which is unrestrictive as we can always standardise any given signal.
- For the residuals we sample the variances from a χ^2 distribution, which is well known to be the sample distribution for variance.
- For both signals and residuals we draw correlation matrices from the LJK distribution [LKJ09], commonly used as a prior for correlation matrices in Bayesian modelling, and which can be considered the correlation equivalent of the Wishart distribution used for covariance matrices.

- We can then construct the covariance matrix $\Sigma = \text{diag}(\sigma) \mathbf{P} \text{diag}(\sigma)$, where σ is the vector of variances per signal/residual and \mathbf{P} is the sampled correlation matrix for the signal/residuals.
- For β we sample entries from a Laplace distribution, which provided the best fit when looking at the entries of real β matrices from Section 5.

With β and Σ_{ϵ} , Σ_{s} coming from real data we would expect certain structures such as clusters with the covariance matrix (e.g. similarity between stocks in the same sector), or similar parameters for certain stocks; this random sampling unfortunately misses out on this. However, from studying the expressions for the replication ratio it does not appear that these characteristics would impact our results. This is further confirmed by noting that when using real assets and signals in Section 5, the results remain valid. To achieve a desired true Sharpe ratio we simply scale the matrix β appropriately. Finally, we let $\mu_{s} = \mathbf{0}$ and do not include a drift term.

We note here that to simulate even 100 assets requires a significant amount of compute, particularly with large T, which realistically will almost always be the case for a strategy of this style. Thus, the analytical approximations are useful to enable quick comparative adjustments for strategies, without requiring lengthy simulations.

E Additional Values for Section 5 and Section 6

Section 5 - Estimated Covariance

In Section 5 we have used the true covariance matrix when constructing our portfolios. To further confirm that the values are robust to the estimation of the covariance matrix (as previously shown in Section 3.3), we run the same simulation again, but this time estimating the covariance matrix for each sample from the observed in-sample data, and using that to construct the portfolio $\hat{\mathbf{w}} = \hat{\Sigma}_{\epsilon}^{-1} \hat{\beta} \mathbf{s}_{t}$.

We collect all the metrics for the analytical, Gaussian + IID simulated estimates, and t + AR(1) simulated estimates into Table 2. Overall, we see that replacing the true residual covariance matrix with its in-sample estimate has only a negligible impact on both in-sample and out-of-sample performance, in line with the sensitivity analysis in Section 3.3. Additionally, we see that the average realised Sharpe ratio is extremely close to the estimate which does not account for convexity (yielded by dividing the average mean by the square root of the average variance), which is in turn close to our analytical value. Thus, to conclude this section, we see that even when the data generating process departs from our idealised assumption, the analytical estimates are still a good estimator for all the metrics of interest.

Section 6 - Full Metrics

We report here the expected metrics in the same style as the previous section. Note that the results reported are averages over all 5,000,000 samples. For example, the interpretation of the empirical expected in-sample mean is: "The average observed in-sample mean for all combinations of signals over all sub-periods", and the interpretation of the analytical expected in-sample mean is: "The average expected in-sample mean for all combinations of signals over all sub-periods".

We observe that while the raw mean and variance estimates differ significantly between the analytical and empirical results, this discrepancy is primarily attributable to a scale factor. Most notably, the large empirical out-of-sample variance of 44.58 (compared to the analytical 0.27) is driven by significant kurtosis in the empirical values; for comparison, the modal values are 0.05 and 0.09 and the median values are 0.49 and 0.12 respectively. In particular, some resamples generate very large realised variances for portfolios whose corresponding out-of-sample Sharpe ratios are essentially zero. These extreme observations have little impact on the average Sharpe ratio but inflate the average variance. As the portfolio weights can be multiplied by a constant to scale both the mean and variance without altering the risk-adjusted return, the scale-invariant Sharpe ratio provides a more robust metric for comparison.

		${\bf Gaussian + IID}$		t + AR(1)	
	Analytical	True Cov	Est. Cov	True Cov	Est. Cov
In-Sample					
Mean μ	0.28	0.28	0.29	0.31	0.32
Variance σ^2	0.30	0.30	0.30	0.41	0.38
Sharpe Ratio SR_{IS}	N/A	8.24	8.26	7.95	8.12
$\mu/\sigma~(\mathrm{SR}_{\mathrm{EIS}})$	8.22	8.24	8.26	7.72	8.10
Out-of-Sample					
Mean μ	0.11	0.11	0.11	0.14	0.14
Variance σ^2	0.29	0.29	0.30	0.59	0.62
Sharpe Ratio SR_{OOS}	N/A	3.14	3.13	2.83	2.83
$\mu/\sigma \; (\mathrm{SR}_{\mathrm{EOOS}})$	3.13	3.14	3.13	2.79	2.79
Replication Ratio					
$ m SR_{OOS}/SR_{IS}$	N/A	38%	38%	36%	35%
SR_{EOOS}/SR_{EIS}	38%	38%	38%	36%	34%

Table 2: Simulation Results comparing Analytical and Estimated values under Gaussian IID and AR(1)+t assumptions for Section 5. Sharpe ratios are annualised assuming 252 trading days.

In this regard, the analytical and empirical estimates align remarkably well. The in-sample expected analytical Sharpe ratio $\mathbb{E}[SR_{EIS}] = 1.16$ is comparable to the derived empirical ratio of 1.21. Similarly, the out-of-sample estimates are similar for the analytical and empirical values: the analytical model predicts $\mathbb{E}[SR_{EOOS}] = 0.18$, which matches the average observed $\mathbb{E}[SR_{OOS}] = 0.18$. This alignment results in consistent replication ratios across the different aggregation methods, ranging between 46% and 54%.

	Analytical	Empirical
In-Sample		
Mean $\mathbb{E}[\mu]$	0.19	0.33
Variance $\mathbb{E}[\sigma^2]$	0.32	0.87
Sharpe Ratio $\mathbb{E}[SR_{IS}]$	N/A	1.33
$\mu/\sigma \; (\mathbb{E}[SR_{EIS}])$	1.16	1.21
Out-of-Sample		
Mean $\mathbb{E}[\mu]$	0.09	0.41
Variance $\mathbb{E}[\sigma^2]$	0.27	44.58
Sharpe Ratio $\mathbb{E}[SR_{OOS}]$	N/A	0.18
$\mu/\sigma \ (\mathbb{E}[SR_{EOOS}])$	0.18	0.21
Replication Ratio		
$\mathbb{E}[\mathrm{SR}_{\mathrm{OOS}}/\mathrm{SR}_{\mathrm{IS}}]$	N/A	52%
$\mathbb{E}[\mathrm{SR}_{\mathrm{OOS}}]/\mathbb{E}[\mathrm{SR}_{\mathrm{IS}}]$	N/A	46%
$\mathbb{E}[\mathrm{SR}_{\mathrm{EOOS}}/\mathrm{SR}_{\mathrm{EIS}}]$	54%	N/A
$\mathbb{E}[SR_{EOOS}]/\mathbb{E}[SR_{EIS}]$	52%	N/A

Table 3: Statistics from resampling study, comparing analytical (implied) and empirical values for Section 6. Sharpe ratios are annualised from monthly values.

Disclaimer

Unless specifically indicated otherwise, the views and opinions expressed in this communication are those of the author(s) and do not reflect the official policy or position of [Qube Research & Technologies Ltd]. The information provided herein is for informational purposes only and should not be construed as advice, financial or otherwise, or opinion of [Qube Research & Technologies Ltd]. [Qube Research & Technologies Ltd] has neither approved nor disapproved of such content and makes no representations as to the accuracy, completeness, or reliability of any information contained herein.

Recipients are advised not to rely solely on this communication for any purpose. [Qube Research & Technologies Ltd] shall not be liable for any errors or omissions, nor for any damages resulting from reliance on this information.