

Residual Books

Felix Gehling

Multiple AI agents

October 2026

Abstract

The portfolio is constructed so that its return equals the inner product of the model's scores with standardized factor residuals, and these residuals are used as the training target. Two versions of the portfolio, which hedge different sets of factors, are combined each month. The annualized Sharpe ratio over 1990–2023 is 4.42.

1 Portfolio construction

At the end of month t , we use the previous 252 trading days and keep stocks with at least 200 daily returns. For each stock, σ_i is its daily volatility times $\sqrt{21}$, and $x_i = r_{i,t+1}/\sigma_i$ is its standardized return over the next month. A factor model $x = Bf + e$ with $\text{Cov}(e) = \Psi = \text{diag}(\psi_i)$ is estimated from the correlation matrix of the daily returns. Factors are the principal components whose eigenvalues exceed the Marchenko–Pastur bound $(1 + \sqrt{n/T})^2$ (Marčenko & Pastur, 1967), where n is the number of stocks and $T = 252$, and $\psi_i = 1 - \sum_k B_{ik}^2$, floored at 0.05.

Let G collect the constraints (the factor loadings and the vector $1/\sigma$ for dollar neutrality) and

$$Q = I - G(G'\Psi^{-1}G)^{-1}G'\Psi^{-1}.$$

For a vector of scores s , the portfolio is

$$v = \Psi^{-1}Q\Psi^{1/2}s, \quad w_i = v_i/\sigma_i,$$

the mean–variance portfolio for expected residual returns $\Psi^{1/2}s$ subject to $G'v = 0$ (Grinold & Kahn, 2000). Because $Q'\Psi^{-1} = \Psi^{-1}Q$, its return is

$$w'r_{t+1} = v'x = s'z, \quad z = \Psi^{-1/2}Qx,$$

where z is the vector of standardized GLS residuals of x on G . We therefore use z as the target. Each month the portfolio is scaled to unit predicted volatility, $v'\Psi v = 1$.

2 Two books

With one year of daily data and n much larger than T , only the stronger factors can be estimated (Baik & Silverstein, 2006). Weaker factors such as value or momentum are better captured by the characteristics. The first book uses $G = [B, 1/\sigma]$. The second uses $G = [B, 1/\sigma, C]$, where C contains the principal components of the month's characteristics (each ranked within the month, missing values set to the median) with eigenvalues above the Marchenko–Pastur bound. Each book has its own target and model.

The books are combined with long-only maximum-Sharpe weights estimated from their past returns, using exponentially weighted moments (half-life 60 months) and correlations shrunk halfway toward zero. Equal weights are used until 36 months of returns are available. The combined portfolio is rescaled to unit predicted volatility.

3 Estimation

Each book has a LightGBM regression (Ke et al., 2017) on all 402 characteristics. The target z is standardized within each month and truncated at ± 4 . Models are refit every 12 months on all previous months. The number of boosting rounds is chosen by early stopping on the Sharpe ratio of the portfolio over the last 60 months of the training sample (patience 200, at most 2,000 rounds), after which the model is refit on the full sample. The learning rate is 0.05, trees have 31 leaves, and the minimum leaf size equals the median number of stocks per month.

4 Results

	1990–2023	1990s	2000s	2010–2023	Max. drawdown
Combined	4.42	8.85	5.22	3.08	−6.8%
Book 1	4.16	7.50	4.67	2.91	−7.8%
Book 2	4.08	8.27	4.89	2.60	−6.2%

Table 1: Annualized Sharpe ratios, gross of trading costs.

The portfolio holds about 2,000 stocks, and annual turnover is about seven times gross exposure.

References

- Baik, J., & Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6), 1382–1408.
- Grinold, R. C., & Kahn, R. N. (2000). *Active portfolio management* (2nd ed.). McGraw-Hill.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Marčenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4), 457–483.