

Markowitz- ML Portfolio for the JKP Common Task Framework

By: Carolina Pennington, Tuva Hatlegjerde, Mikkel Shiffman, Eric Perczek

We created a Markowitz style long-short US equity strategy built from the WRDS data and Barra style industry and theme risk model, submitted to the JKP Common Task Framework. Our sample from 1990-2023 (408 months) gives us an annualized Sharpe ratio of 2.32 with a 10% ex-ante volatility target.

Data:

We used two of the three WRDS tables provided `ctff_chars` (402 firm characteristics and next-month returns) and `ctff_features` (the characteristic list), and did not use `ctff_daily_ret`.

Stock-months	1,369,324
Period	Dec-1951–Nov-2023 (864 months)
Stocks per month	~360 (1950s) to ~2,100–2,400 (1990s onward)
Characteristics	402
Target	<code>ret_exc_lead1m</code> , next month's excess return
Test period	Dec-1989–Nov-2023 (408 months, flagged <code>ctff_test == 1</code>)
Industry	2-digit SIC code

Method:

Each month we hold a portfolio in which the weights are proportional to the inverse return covariance matrix times the expected-return forecast. The forecast predicts next month's stock returns while the covariance matrix predicts how those returns move together. Both predictions use characteristics known at that month and the portfolio is then rescaled to 10% annualized ex-ante volatility. This is the equation we used:

$$w_t \propto \Sigma_t^{-1} \mu_t$$

Pre-processing: The characteristics are rank transformed within each month to $[-0.5, 0.5]$, with missing values set to 0. The forecast target is next month's return minus the cross-sectional average for that month.

Expected returns: Using the below equation, we used a single ridge regression to forecast next month's return on all 402 ranked characteristics, fit once on the pre-test months.

$$\beta = \left(\frac{X'X}{n} + \lambda I \right)^{-1} \frac{X'y}{n}, \quad \mu_t = X_t \beta$$

Risk model: B_t are factor exposures, F the covariance of monthly factor returns, D_t idiosyncratic variances. Factors are 74 industry dummies (2-digit SIC) plus 40 themes, where a theme is a cluster of characteristics from hierarchical clustering of their correlation matrix and a stock's exposure is its average rank within the cluster. D_t is each stock's trailing 252-day idiosyncratic volatility squared, times one calibrated constant. F , the clustering and the constant are fixed once; B_t and D_t update monthly.

$$\Sigma_t = B_t F B_t' + D_t$$

Splits the equation into:

- B_t = Each stock's risk factor exposures
- F = Covariance of the monthly factor returns,
- D_t = Idiosyncratic variance

Development and Model Selection:

We kept three static blocks. We fit the model in the beginning and did not update it during the test period, it was a static and simple model.

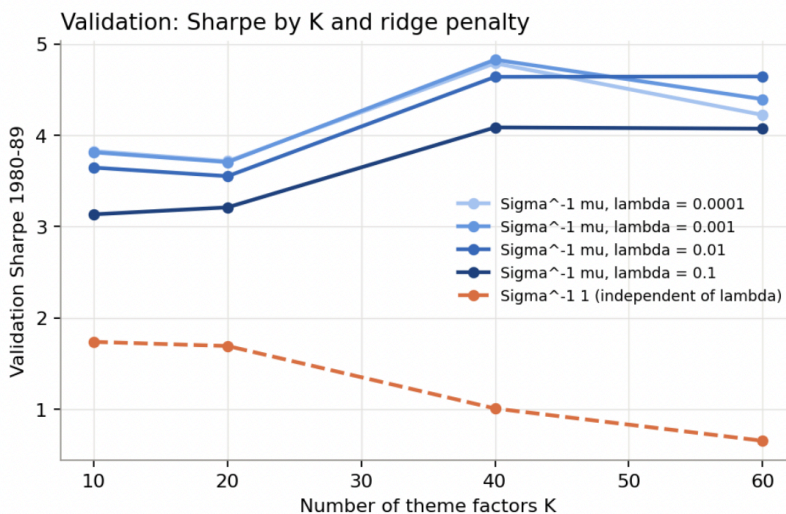
Block	Months	Stock-months	Used for
Train	Dec-1951-Dec-1979 (337)	282,997	fit every candidate model
Validation	Jan-1980-Nov-1989 (119)	200,629	choose knobs (K , λ , industry factors, rule) by Sharpe
Test	Dec-1989-Nov-2023 (408)	885,698	scored once, after refitting on train + validation

We kept four portfolio rules in line with the 10% ex-ante volatility to keep the Sharpe ratios comparable. Forecast, min-variance, Markowitz, and Blend (half Markowitz, half forecast).

Step 1: Created a baseline model. Our first risk model only used a single market factor, an intercept and the K themes($K=40$ themes and a ridge penalty of 0.001). Through this model we learned that forecasts and risk are complementary and combining them gives a Sharpe near 2. Next, unconstrained Markowitz can under forecast its own risk. Finally, the ridge penalty matters little below 0.01 while the theme count matters a lot.

Test, 1990-2023	Forecast	Min-variance	Markowitz
Sharpe	1.54	1.40	1.98

Below is the data for the baseline risk model (single market factor): The Sharpe rises sharply from K=10 to K=40, while the min-variance rule declines. This gives us evidence for choosing K=40.



Step 2:

Added industry neutrality and alternate regressions. We added industry risk factors, lasso regression, blend rule, and an industry forecast target. The table below shows the results from these changes.

Change	Validation Sharpe change	Test Sharpe change	Decision
Industry risk factors (off → on)	+1.09 (better in 100% of pairs)	+0.38 (100%)	keep
K = 20 → 40 themes	+1.09 (100%)	+0.36 (100%)	keep
Industry-demeaned forecast target	-0.23 (better in 10%)	-0.11 (17%)	drop
Lasso instead of ridge (best ridge penalty each)	+0.06	-0.07	drop, keep ridge
Elastic net / ensemble	no gain	no gain	drop
Blend instead of pure Markowitz	+0.08 (65%)	+0.03 (68%)	keep as an option

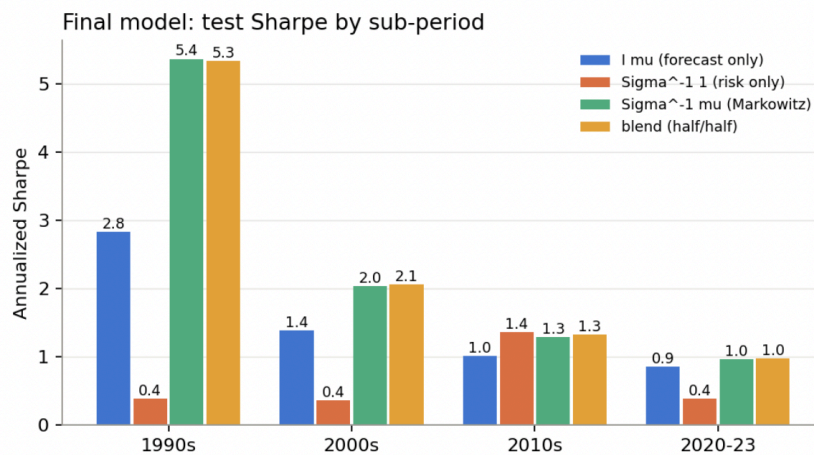
All decisions were made on validation; the test column was computed afterwards and only confirms the ranking held. Industry neutrality belongs in Σ : once there, Markowitz already hedges industry exposure, so demeaning the forecast removes it twice and discards signal. Ridge beat lasso on simplicity, as lasso's 0.06 validation edge is within noise.

Results:

We chose the final model based on validation and testing every combination of industry risk factors on or off, two theme-cluster sizes, and two ridge penalties. The final model included: industry risk factors on, K=40, ridge penalty of 0.001, and ridge regression.

	Forecast	Min-variance	Markowitz	Blend
Sharpe	1.53	0.63	2.32	2.31
Realized vol (10% target)	17.6%	33.1%	43.1%	35.3%

In comparison, the competition's "Markowitz-ML" benchmark scores 2.52, and top leaderboard entries score around 4. Average gross leverage of the Markowitz weights is 18.8x. Markowitz and Blend perform equally well and outperform the Forecast only and Min-Variance. The minimum variance rule actually gets worse once industry factors are added.



Performance decays by decade: 5.4 in the 1990s, 2.0, 1.3, then 1.0 in 2020 to 2023. The full-sample 2.32 overstates current performance; an investor should plan around a Sharpe of 1.0 to 1.3.

Robustness Check:

To see how much the Sharpe ratio depends on which historical years we test, we ran a 50-world bootstrap. The results are shown below:

Rule	Real history	Bootstrap mean	Bootstrap range	sd
Forecast (forecasts only)	1.53	1.45	0.60-2.16	0.29
Min-variance (risk only)	0.63	0.56	-0.02-1.00	0.19
Markowitz (forecast + risk)	2.32	2.26	1.55-2.92	0.26
Blend (half Markowitz, half Forecast)	2.31	2.24	1.42-2.84	0.26

Conclusion:

We created a Markowitz portfolio that is built from ridge regression return forecasts and a Barra style industry and the risk model. It delivers an annualized Sharpe ratio of 2.32 out of sample over the period 1990-2023, which is well above what a forecast or risk model alone achieves (1.52 and 0.63 respectively). The bootstrap puts 2.32 near the center of a 1.8 to 2.8 band, so it is not a lucky draw, but the decade trend means the realistic forward expectation is a Sharpe near 1.0 to 1.3. Even so, pairing a ridge forecast with a factor risk model delivers more than the forecast alone (1.53).