Universität Konstanz

Testing Out-of-Sample Portfolio Performance

Ekaterina Kazak[1] Winfried Pohlmeier[2]

[1] University of Konstanz, GSDS

[2] University of Konstanz, CoFE, RCEA

Econometric Research in Finance Workshop 2017
SGH Warsaw School of Economics
September 15, 2017

Example: Mean Variance Portfolio

Universität Konstanz

- Consider an investor who chooses a portfolio among N financial assets
- $r_t \in \mathbb{R}^N$, $\mathrm{E}[r_t] = \mu$, $\mathrm{V}[r_t] = \Sigma$
- Efficient (norm constrained) portfolio:

$$\max_{\omega} CE(\omega) = \max_{\omega}\{\mu'\omega - \frac{\gamma}{2}\omega'\Sigma\omega\} \quad \text{such that} \quad \iota'\omega = 1$$

$$\omega(m) = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma} \cdot \left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\mu$$

- number parameters $\sim \mathcal{O}(N^2)$.

Empirical Portfolio Models: Stylized Facts

Universität Konstanz

- Stylized facts of empirical portfolio weights:
  - high standard error highly instable across time
  - bad predictive quality
  - some with pathological distribution with no finite first and second moments
- The naive equally weighted portfolio is often a very strong competitor in within-sample and out-of-sample competitions

Illustration: Estimation Noise

Universität Konstanz

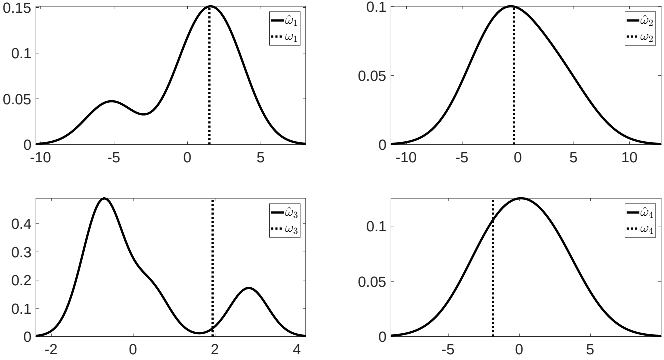


Figure: Frequency distribution of estimated portfolio weights vs the true theoretical value based on the Monte-Carlo study, $r \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma)$, number of assets $N = 5$, estimation window length $T = 120$ (i.e. 10 years of monthly data). (μ, Σ) set as in the Kenneth R. French data.

Empirical Portfolio Models: Performance Testing

Universität Konstanz

- Evaluation of Portfolio allocation strategies: Sharpe Ratio or Certainty Equivalent
- Testing the difference in performance measures by a z-test
- Related literature:
  - Under iid-normality for Sharpe Ratio: Jobson and Korkie (1981), Memmel (2003)
  - Bootstrap test: Ledoit and Wolf (2008), Ledoit and Wolf (2011)
  - Delta method for Certainty Equivalent: DeMiguel et al. (2009)
  - Test the difference in performance measures of two arbitrary asset returns
- *Portfolio* performance testing quality is under-researched

Goals of the Paper

Universität Konstanz

- Investigation of the stochastic nature of out-of-sample portfolio returns underlying the performance tests
- Analysis of the size and power properties of portfolio performance tests
- Give guidance on how to deal with the low power
- providing an lternative way of using the information of performance test within an algorithmic pre-testig strategy

Universität Konstanz

- Introduction
- Out-of-Sample Returns and Performance Measures
- Monte Carlo Design
- Test properties
- Power Optimal Pretest Portfolios
- Summary

Certainty Equivalent

Universität Konstanz

- Consider a portfolio of N financial assets $r_t \in \mathbb{R}^N$ with $\mathrm{E}[r_t] = \mu$ and $\mathrm{V}[r_t] = \Sigma$
- Let $\omega(s)$ denote an $N \times 1$ vector of portfolio weights of a strategy s
- Portfolio return $\quad r_t^p(s) = \omega(s)' r_t$
- Portfolio mean $\quad \mu_p(s) = \mathrm{E}[r_t^p(s)] = \omega(s)'\mu$
- Portfolio variance $\quad \mathrm{V}[r_t^p(s)] = \omega(s)'\Sigma\omega(s)$
- Certainty Equivalent $\quad CE(\omega(s)) = \omega(s)'\mu - \frac{\gamma}{2}\,\omega(s)'\Sigma\,\omega(s)$
- CE difference $\quad \Delta_0(s, \tilde{s}) = CE(\omega(s)) - CE(\omega(\tilde{s}))$

Certainty Equivalent

Universität Konstanz

- Out-of-sample portfolio return of strategy s:

  $\hat{r}^p_{t+1}(s) = \hat{\omega}_{t+1|t}(s)' r_{t+1} = \hat{\omega}_t(s)' r_{t+1}$

- $\mu_{op}(s) = \mathrm{E}\left[\hat{r}^p_{t+1}(s)\right] = \mathrm{E}\left[\hat{\omega}_t(s)\right]' \mu$

- $\sigma^2_{op}(s) = \mathrm{V}\left[\hat{r}^p_{t+1}(s)\right] = \mathrm{E}\left[\hat{\omega}_t(s)' \Sigma \, \hat{\omega}_t(s)\right] + \mu' \, \mathrm{V}\left[\hat{\omega}_t(s)\right] \mu$

- Out-of-sample CE:

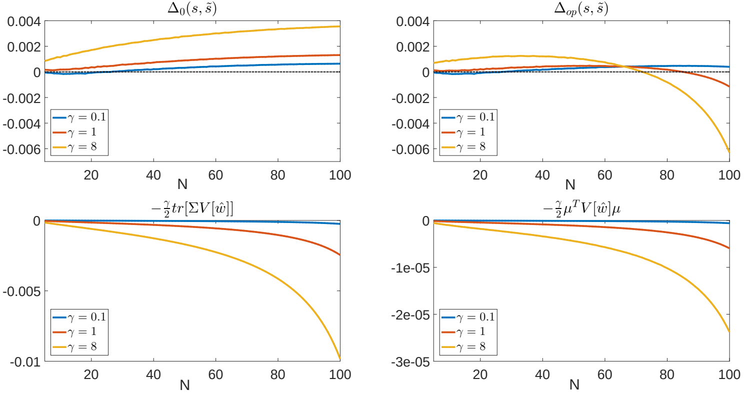
  $CE_{op}(\hat{\omega}_t(s)) = \mu_{op}(s) - \frac{\gamma}{2}\sigma^2_{op}(s) = CE(\omega(s)) - \frac{\gamma}{2} tr\big(\Sigma \, \mathrm{V}\left[\hat{\omega}(s)\right]\big) - \frac{\gamma}{2}\mu' \, \mathrm{V}\left[\hat{\omega}_t(s)\right] \mu$

- Out-of sample difference:

  $\Delta_{op}(s, \tilde{s}) \equiv$

  $\Delta_0(s, \tilde{s}) - \frac{\gamma}{2}\Big[tr(\Sigma \, \mathrm{V}\left[\hat{\omega}_t(s)\right]) - tr(\Sigma \, \mathrm{V}\left[\hat{\omega}_t(\tilde{s})\right])\Big] - \frac{\gamma}{2}\mu' \Big[\mathrm{V}\left[\hat{\omega}_t(s)\right] - \mathrm{V}\left[\hat{\omega}_t(\tilde{s})\right]\Big]\mu$

CE difference GMVP vs 1/N



Average CE differences over 5 000 random N out of 100 asset combinations for GMVP and equally weighted portfolio by dimension of the asset universe N for different values of the risk aversion parameter γ. The estimation window length T is set to 120 (10 years of monthly observations). Upper-left plot: difference in theoretical CE, $\Delta_0(s, \tilde{s})$. Lower-left plot: estimation noise penalty. Lower-right plot: out-of-sample risk penalty. Upper-right plot: overall out-of-sample CE difference.

Result 1: the null hypothesis

Universität Konstanz

- Applied research: comparison is based on $\widehat{CE}_{op}(\hat{\omega}_t(s))$

$$H_0: \quad \Delta_{op}(s, \tilde{s}) = \mathrm{E}\left[\widehat{CE}_{op}(\hat{\omega}_t(s)) - \widehat{CE}_{op}(\hat{\omega}_t(\tilde{s}))\right] = 0$$

- $\mathrm{E}\left[\widehat{CE}_{op}(\hat{\omega}_t(s))\right] = CE(\omega(s)) - \frac{\gamma}{2} tr\left[\Sigma\, \mathrm{V}\left[\hat{\omega}(s)\right]\right] - \frac{\gamma}{2}\mu'\, \mathrm{V}\left[\hat{\omega}_t(s)\right]\mu$
- Takes into account estimation risk $-\frac{\gamma}{2} tr\left[\Sigma\, \mathrm{V}\left[\hat{\omega}(s)\right]\right]$
- Takes into account forecasting risk $-\frac{\gamma}{2}\mu'\, \mathrm{V}\left[\hat{\omega}_t(s)\right]\mu$
- Evaluate test properties by a Monte Carlo Study

Universität Konstanz

- Introduction
- Out-of-Sample Returns and Performance Measures
- Monte Carlo Design
- Test properties
- Power Optimal Pretest Portfolios
- Summary

Distribution of the out-of-sample returns

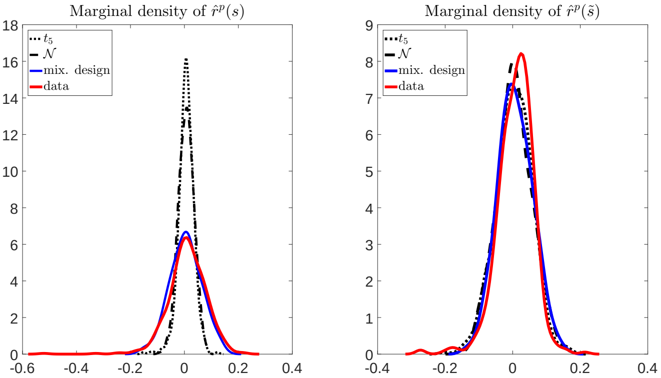
Universität Konstanz

- In the following the out-of-sample portfolio returns $r^p_{t+1}(s)$ are based on the estimated weights and $r^p_{t+1}(\tilde{s})$ are based on the non-stochastic weights
- Assume $r_t \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma)$
- Okhrin and Schmid (2006): estimated Global Minimum Portfolio (GMVP) weights follow a multivariate elliptical t-distribution
- Mixture design:

$$f\left(\hat{r}^p_{t+1}(s), r^p_{t+1}(\tilde{s})\right) = f\left(\hat{\omega}_t(s)' r_{t+1}, \omega_t(\tilde{s})' r_{t+1} | \hat{\omega}_t(s)\right) \cdot g\left(\hat{\omega}_t(s)\right)$$

$$\begin{pmatrix} \hat{r}^p_{t+1}(s) \\ r^p_{t+1}(\tilde{s}) \end{pmatrix}\Bigg|_{\hat{\omega}_t(s)} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\omega}_t(s)'\mu \\ \omega_t(\tilde{s})'\mu \end{bmatrix}, \begin{bmatrix} \hat{\omega}_t(s)'\Sigma\hat{\omega}_t(s) & \hat{\omega}_t(s)'\Sigma\,\omega_t(\tilde{s}) \\ \omega_t(\tilde{s})\Sigma\hat{\omega}_t(s)' & \omega_t(\tilde{s})'\Sigma\,\omega_t(\tilde{s}) \end{bmatrix}\right)$$

Distribution of the out-of-sample returns



Marginal distribution of the out-of-sample portfolio returns for the GMVP ($\hat{r}^p(s)$) and equally weighted portfolio ($\hat{r}^p(\tilde{s})$): based on the real data (in red), simulated from bivariate t_5, simulated from bivariate normal and simulated from the proposed mixture design (in blue). The mean and standard deviation of the simulated returns are adjusted to be the same as of the empirical portfolio returns.

Universität Konstanz

- Introduction
- Out-of-Sample Returns and Performance Measures
- Monte Carlo Design
- Test properties
- Power Optimal Pretest Portfolios
- Summary

Test Properties

Universität Konstanz

The following comparison is based on:

- Type of the test
  - Delta Method
  - Bootstrap (percentile and t-statistic)
- Out-of-sample evaluation horizon (H)
- Risk aversion (γ)
- Estimation noise N/T, with $N = 30$ and T denoting the size of in-sample window
- Benchmark strategy (correlatedness of the out-of-sample returns)

Size ($\hat{\alpha}$)

Universität Konstanz

Table: Empirical rejection probabilities under H_0 for GMVP vs 1/N. $\alpha = 5\%$. $N/T = 0.01$

		Two-sided			One-sided		
		Delta method	Bootstrap Percentile	Bootstrap t-statistic	Delta method	Bootstrap Percentile	Bootstrap t-statistic
H = 100	$\gamma = 0.5$	0.0539	0.0505	0.0542	0.0520	0.0503	0.0520
	$\gamma = 1.0$	0.0544	0.0508	0.0547	0.0522	0.0502	0.0523
	$\gamma = 3.0$	0.0547	0.0504	0.0549	0.0515	0.0504	0.0519
H = 500	$\gamma = 0.5$	0.0540	0.0532	0.0541	0.0514	0.0511	0.0517
	$\gamma = 1.0$	0.0536	0.0531	0.0540	0.0522	0.0518	0.0523
	$\gamma = 3.0$	0.0543	0.0538	0.0545	0.0512	0.0513	0.0514
H =1000	$\gamma = 0.5$	0.0564	0.0564	0.0566	0.0537	0.0535	0.0538
	$\gamma = 1.0$	0.0561	0.0561	0.0565	0.0530	0.0528	0.0529
	$\gamma = 3.0$	0.0563	0.0559	0.0564	0.0519	0.0525	0.0523

Figures in the table correspond to the share of Monte Carlo draws where the null hypothesis was rejected (out of 50 000 draws). H denotes the out-of-sample evaluation window length and γ denotes risk aversion coefficient.

Power $(1 - \hat{\beta})$

Universität Konstanz

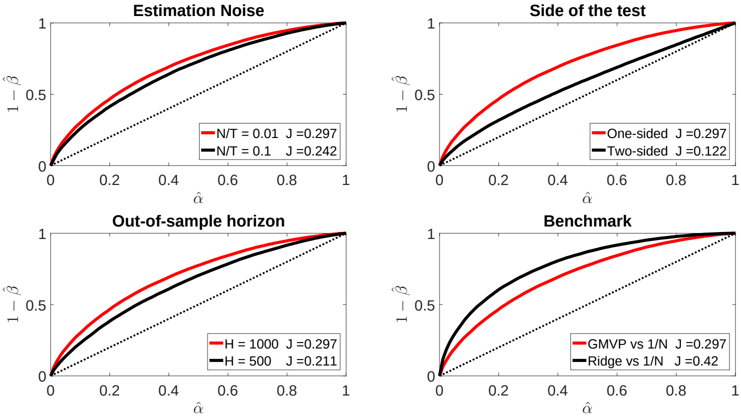
Table: Power at 1% expected CE difference for GMVP vs 1/N. $\alpha = 5\%$. $N/T = 0.01$

		Two-sided			One-sided		
		Delta method	Bootstrap Percentile	Bootstrap t-statistic	Delta method	Bootstrap Percentile	Bootstrap t-statistic
H = 100	$\gamma = 0.5$	0.0625	0.0588	0.0630	0.0871	0.0845	0.0875
	$\gamma = 1.0$	0.0622	0.0581	0.0624	0.0851	0.0825	0.0855
	$\gamma = 3.0$	0.0618	0.0587	0.0625	0.0830	0.0818	0.0842
H = 500	$\gamma = 0.5$	0.0875	0.0868	0.0874	0.1399	0.1396	0.1404
	$\gamma = 1.0$	0.0903	0.0896	0.0906	0.1405	0.1405	0.1408
	$\gamma = 3.0$	0.0876	0.0875	0.0877	0.1368	0.1375	0.1375
H =1000	$\gamma = 0.5$	0.1280	0.1276	0.1283	0.1993	0.1992	0.1999
	$\gamma = 1.0$	0.1271	0.1267	0.1272	0.1952	0.1956	0.1960
	$\gamma = 3.0$	0.1227	0.1238	0.1229	0.1922	0.1933	0.1928

Figures in the table correspond to the share of Monte Carlo draws where the null hypothesis was rejected (out of 50 000 draws). H denotes the out-of-sample evaluation window length and γ denotes risk aversion coefficient.

ROC curves: Delta Method

Universität Konstanz



ROC curves for a two-sided Delta Method. Risk aversion is set to $\gamma = 1$ for the asset space $N = 30$. Left panel: GMVP vs 1/N for different estimation noise N/T ratios. Right panel: GMVP combined with the ridge covariance matrix estimator vs 1/N with $N/T = 0.01$.

Result 2: properties of tests

Universität Konstanz

- The tests are heavily influenced by:
  - out-of-sample horizon length
  - estimation noise
  - correlation degree among the out-of-sample returns of the two strategies
- The power of the tests is very low

Result 2: properties of tests

Universität Konstanz

- The tests are heavily influenced by:
  - out-of-sample horizon length
  - estimation noise
  - correlation degree among the out-of-sample returns of the two strategies
- The power of the tests is very low

  However
- One-sided tests have better testing properties
- Tests can be used in pretesting

Universität Konstanz

- Introduction
- Out-of-Sample Returns and Performance Measures
- Monte Carlo Design
- Test properties
- Power Optimal Pretest Portfolios
- Summary

Pretest estimation

Universität Konstanz

Deciding between two portfolio strategies in the presence of low power:

- The pretest estimator depends either on strategy s in case H_0 is rejected or on $\tilde{s}$ otherwise:

$$H_0: \quad \Delta_{op}(s,\tilde{s}) \leq 0 \qquad \text{and} \qquad H_1: \quad \Delta_{op}(s,\tilde{s}) > 0$$

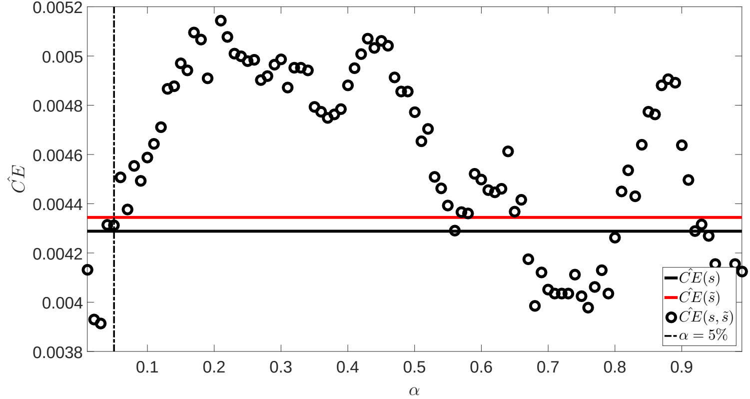
$$\omega_{t+h}(s,\tilde{s}) = \mathbb{1}\left(\hat{\Delta}_{op}(s,\tilde{s}) > \Delta^*(\alpha)\right)\left(\omega_{t+h}(s) - \omega_{t+h}(\tilde{s})\right) + \omega_{t+h}(\tilde{s}), \quad h = 1, ..., H$$

$$\begin{aligned} \mathrm{E}\left[\widehat{CE}_{op}(s,\tilde{s}) \middle| \Delta_{op}(s,\tilde{s}) > 0\right] &= \pi(\alpha) \cdot \mathrm{E}\left[\hat{\Delta}_{op}(s,\tilde{s}) \middle| \hat{\Delta}_{op}(s,\tilde{s}) > \Delta^*(\alpha)\right] \\ &+ \mathrm{E}\left[\widehat{CE}_{op}(\tilde{s})\right] \\ \mathrm{E}\left[\widehat{CE}_{op}(s,\tilde{s}) \middle| \Delta_{op}(s,\tilde{s}) \leq 0\right] &= \alpha \cdot \mathrm{E}\left[\hat{\Delta}_{op}(s,\tilde{s}) \middle| \hat{\Delta}_{op}(s,\tilde{s}) > \Delta^*(\alpha)\right] \\ &+ \mathrm{E}\left[\widehat{CE}_{op}(\tilde{s})\right] \end{aligned}$$

- How to choose an optimal α?

Pretesting strategy: infeasible

Universität Konstanz



Pretesting strategy: GMVP vs $1/N$, $\gamma = 1$. Kenneth R.French data on 5 industry portfolios with estimation window length of $T = 60$ (5 years of monthly observations), corresponding to $N/T = 0.08$ ratio. Dashed line corresponds to the conventional $\alpha = 5\%$.

Pretesting strategy

Universität Konstanz

- Feasible solution: the in-sample CE optimizing significance level α_t^* is chosen for the test, determining the strategy for the next period $t+1$:

$$\alpha_{t+1}^* = \arg\max_{\alpha} CE_{in}^*(\alpha, s, \tilde{s}|t-T, ..., t).$$

- Shrinking α_{t+1}^* towards a target α_0, e.g. to the conventional 5% level:

$$\alpha_{t+1}^s = (1-\lambda)\alpha_{t+1}^* + \lambda\alpha_0,$$

- Adaptive smoothing the series according to

$$\alpha_{t+1}^m = (1-\lambda)\alpha_{t+1}^* + \lambda\alpha_t^m,$$

Pretesting strategy: out-of-sample CE

Universität Konstanz

	T=60			T=120		
	γ=0.1	γ=1	γ=3	γ=0.1	γ=1	γ=3
Strategy	N=5					
GMVP	0.0898	0.0755	0.0470	0.0858	0.0722	0.0427
1/N	0.0929	0.0777	0.0417	0.0885	0.0721	0.0367
In-sample	0.0957	0.0806	0.0496	0.0905	0.0756	0.0438
Shrinking	0.0948	0.0794	0.0456	0.0908	0.0749	0.0405
Smoothing	**0.0984**	**0.0830**	**0.0516**	**0.0931**	**0.0778**	**0.0455**
	N=30					
GMVP	0.0839	0.0662	0.0355	0.0829	0.0733	0.0492
1/N	0.0930	0.0784	0.0459	0.0893	0.0738	0.0407
In-sample	0.0962	0.0793	0.0468	**0.0969**	0.0821	0.0517
Shrinking	0.0938	0.0782	0.0460	0.0926	0.0783	0.0481
Smoothing	**0.0997**	**0.0837**	**0.0525**	**0.0968**	**0.0832**	**0.0536**

The numbers in the table correspond to the annualized average out-of-sample CE over 1000 randomly formed portfolios of the specified size. T denotes the estimation window length, γ denotes risk aversion coefficient and N is the number of assets. The numbers in bold correspond to the largest CE obtained for a given γ, N, T combination. The evaluation window length $H = 600$. The tuning parameter λ for both shrinking and smoothing the α^* series is set to be 0.5.

Refinements of the pretesting strategy

- adaptive smoothing: optimizing over the smoothing papameter
- bagging the indicator
- pretesting of $\hat{\omega}_t(g)$

Concluding remarks

Universität Konstanz

- Why is the $1/N$ strategy performing so well?
  - no estimation risk
  - low power of performance test
- Tests (incl. various implementations) have very similar properties
- One can improve on testing properties by choosing another benchmark and longer evaluation horizon
- One-sided tests do better in terms of power
- Choosing a lower α reduces the probability of random selection

Universität Konstanz

Thank you!

References

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

Jobson, J. D. and B. M. Korkie (1981): "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *The Journal of Finance*, 36, 889–908.

Ledoit, O. and M. Wolf (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2011): "Robust Performances Hypothesis Testing With the Variance," *Wilmott*, 2011, 86–89.

Markowitz, H. (1952): "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

Memmel, C. (2003): "Performance Hypothesis Testing with the Sharpe Ratio," *Finance Letters*, 1, 21–23.

Okhrin, Y. and W. Schmid (2006): "Distributional properties of portfolio weights," *Journal of Econometrics*, 134, 235–256.