General portfolio optimization

Assume you have n assets with expected returns μ and variance-covariance matrix Σ. What is the minimum variance portfolio that delivers expected return not lower than μ_0 ?

$$
\begin{gathered}
w^* = Min\, w'\Sigma w \;\; s.t. \\
w'\mathbf{1} = 1 \\
w'\boldsymbol{\mu} \geq \mu_0
\end{gathered}
$$

The solution to this problem is well-known has the standard interpretation of the two-fund separation theorem (or use software):

$$w^* = \alpha w_{min.var.} + (1-\alpha) w_{tangency}$$

s.t. $w_{min.var.} = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}$ is the minimum variance portfolio and

$w_{tangency} = \frac{1}{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}\Sigma^{-1}\boldsymbol{\mu}$ is the tangency (highest SR) one

$$\alpha = \frac{E\left(w'_{tangency}\mu\right) - \mu_0}{E\left(w'_{tangency}\mu\right) - E\left(w'_{min.var.}\mu\right)} = \frac{Tangency\, ret. - target\, ret.}{Tangency\, ret. - min.\, var.\, ret.}$$

General portfolio optimization

Practical implementation requires the estimates of expected returns and variance-covariance matrix for the max SR portfolio:

$$w_{tangency} = \frac{1}{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}\Sigma^{-1}\boldsymbol{\mu}$$

Expected returns: $\boldsymbol{\mu}$, a vector of n parameters

Variances: a vector of n parameters

Covariances: n(n-1)/2 parameters

The more assets you have, the less reliable are the in-sample portfolio weights.

A simple out-of-sample exercise

49 portfolios of stocks sorted by their industry classification

Estimate the mean and variance-covariance matrix on 180 monthly observations (15 years), form optimal portfolio for 60 months.

Data: US equities, July 1969 – June 2019

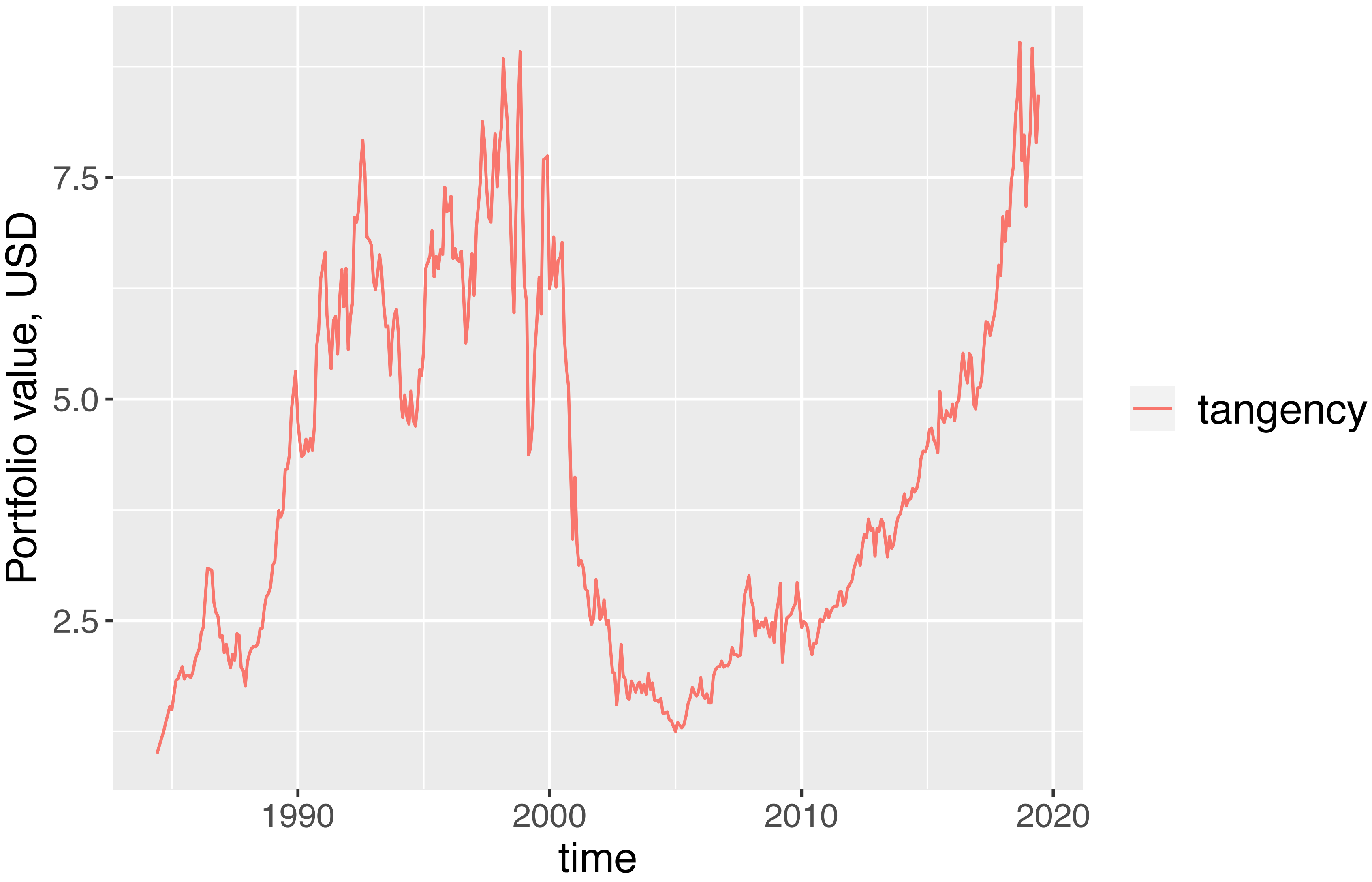
First out-of-sample month: July 1984

Non-overlapping out-of-sample 5 year investment periods

Start with $1 invested at the end of June 1984

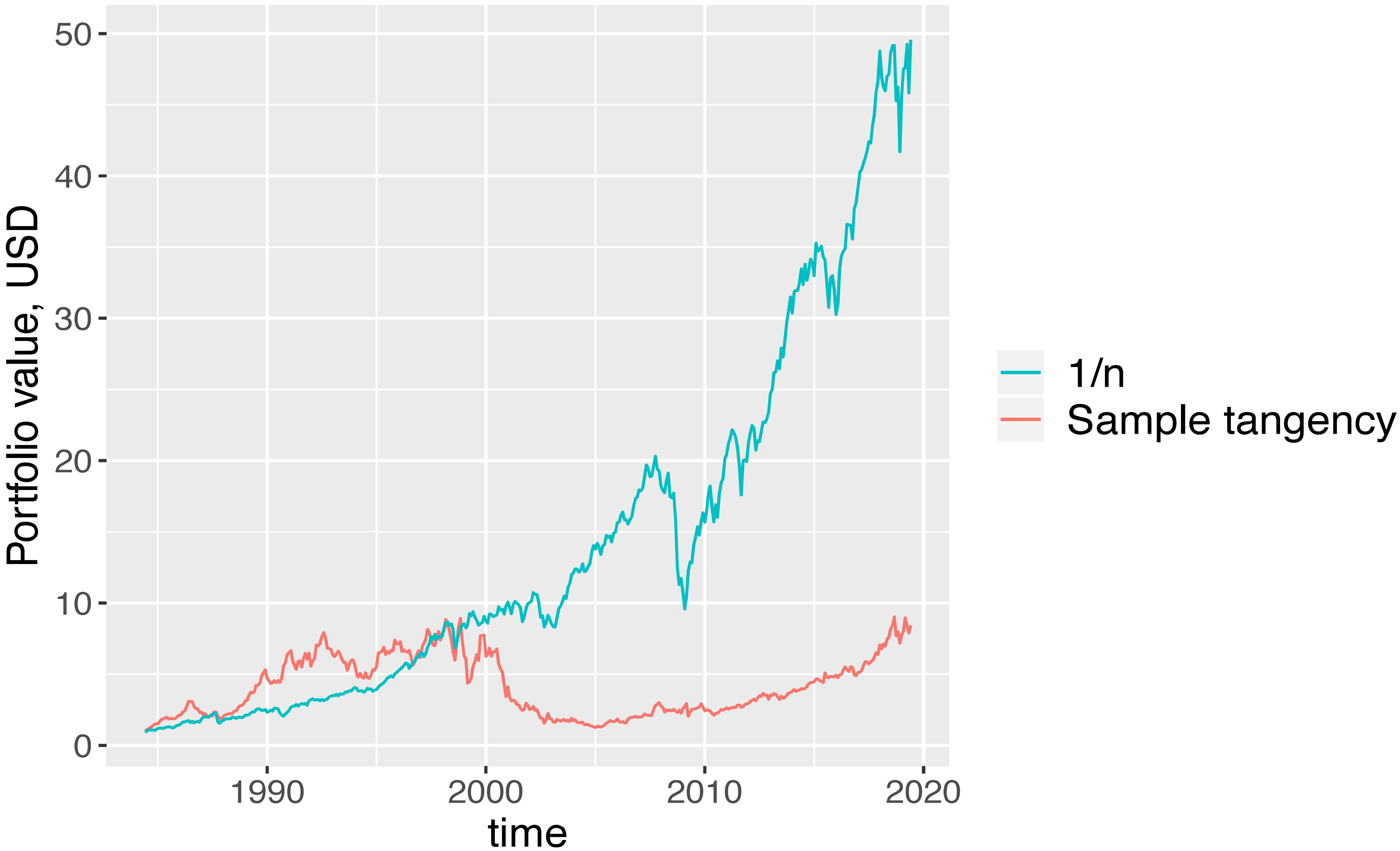
A simple out-of-sample exercise

$1 in 1984 grows to $8.43 by July 2019: 6.3% annual growth



Equally-weighted portfolio

A 1/n portfolio grew to $49.56 (on average, 11.8% annual return)



What happened?

In sample, optimization tends to overfit the data.

Out-of-sample, strategy results are often worse

- Estimation errors on the mean and variances
- Structural breaks, etc

Generally, estimation errors in expected returns are worse

3 basic tools to improve portfolio performance, **shrinkage**:

- Ridge
- Shrinkage to the mean
- Lasso

I will introduce ridge and lasso in the context of linear regressions (OLS), and then move to portfolios.

Ridge

A standard linear regression (OLS) tries to estimate the beta coefficients using the values that minimize:

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

Ridge regression is very similar to the OLS, but has an additional component:

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \mathrm{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2$$

Here $\lambda \geq 0$ is a *tuning parameter*, which determines the degree of shrinkage.

Ridge

An equivalent way to write the problem is:

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 \le t,$$

All the coefficients are shrunk towards 0 (but not exactly).

Can be used when the number of predictors (p) is larger than the number of observations (n).

Works well with highly correlated predictors.

Why? Smaller coefficients will have smaller standard errors, and with many parameters this is crucial.

Ridge

In matrix form the problem looks as follows:

$$RSS(\lambda) = (y - X\beta)'(y - X\beta) + \lambda\beta'\beta$$

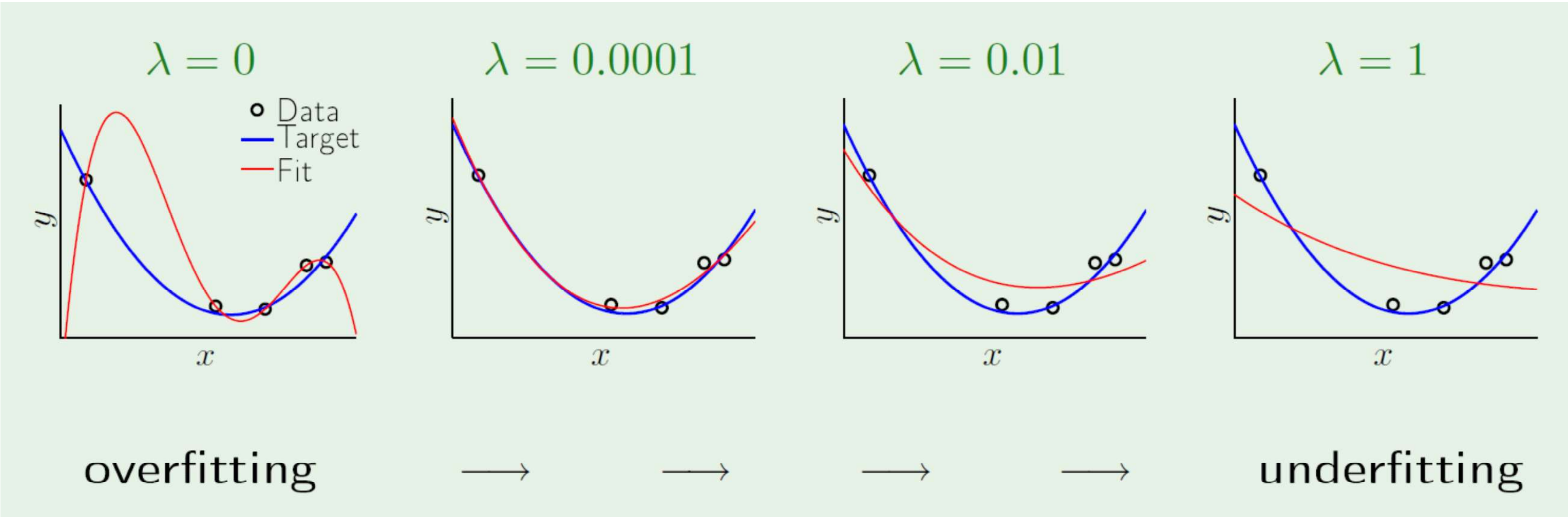
Note that the solution has an easy form:

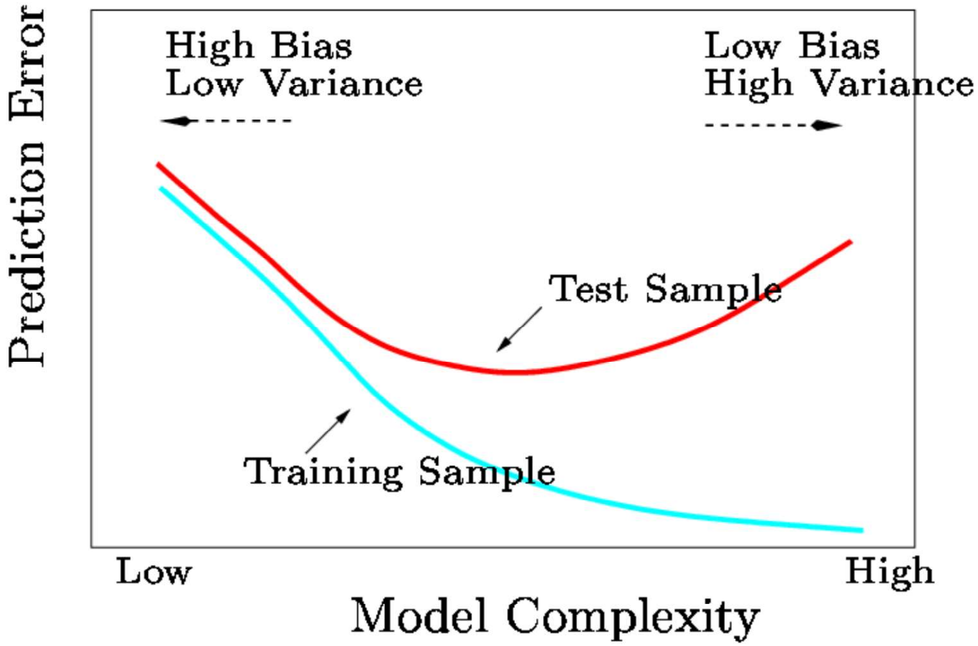
$$\beta^{ridge} = (X'X + \lambda \mathrm{I})^{-1}X'y$$

That is, the solution adds a positive number to all the diagonal of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ before inverting the matrix.

- Very convenient, when your optimization involves many predictors that are highly correlated
- Tends to stabilize the loadings
- Unlike OLS, produces a biased parameter estimate, but with a much smaller varance

Tradeoffs between bias and variance of the prediction





Back to portfolio optimization:

Ridge version of the portfolio optimization:

$$RSS(\lambda) = w'\Sigma\,\mathrm{w} + \lambda w'w + \lambda_1(w'1\ - 1) + \lambda_2(w'\mu\ - \mu_0).$$
$$= w'(\Sigma + \lambda I)\,w + \lambda\beta'\beta + \lambda_1(w'1\ - 1)$$

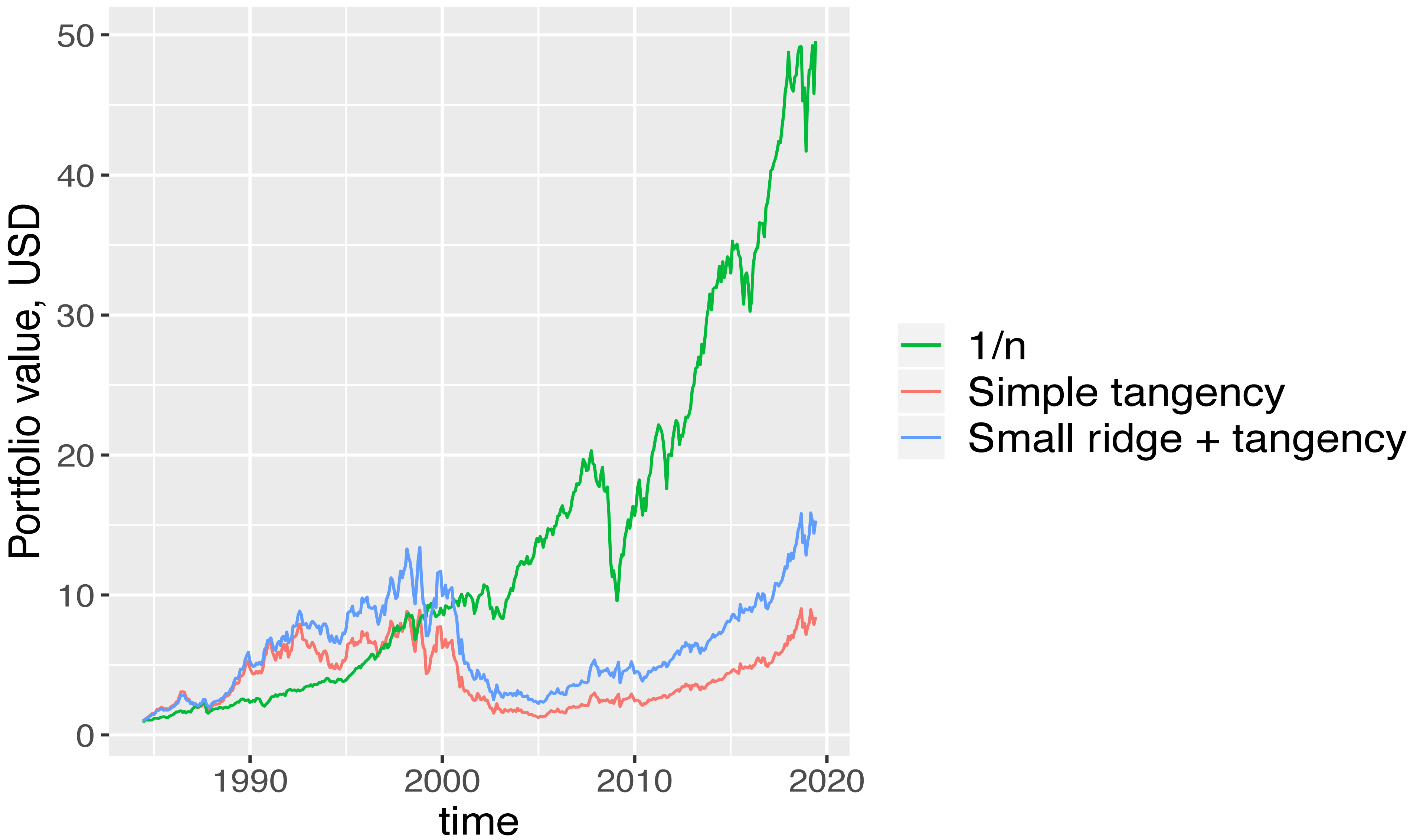
Each portfolio corresponding to the target return μ_0, is the same as in the original optimization, except for using $(\Sigma + \lambda I)$ instead of the sample variance-covariance matrix

Variance-covariance matrix is shrunk towards a diagonal one: less instability due to inversion.

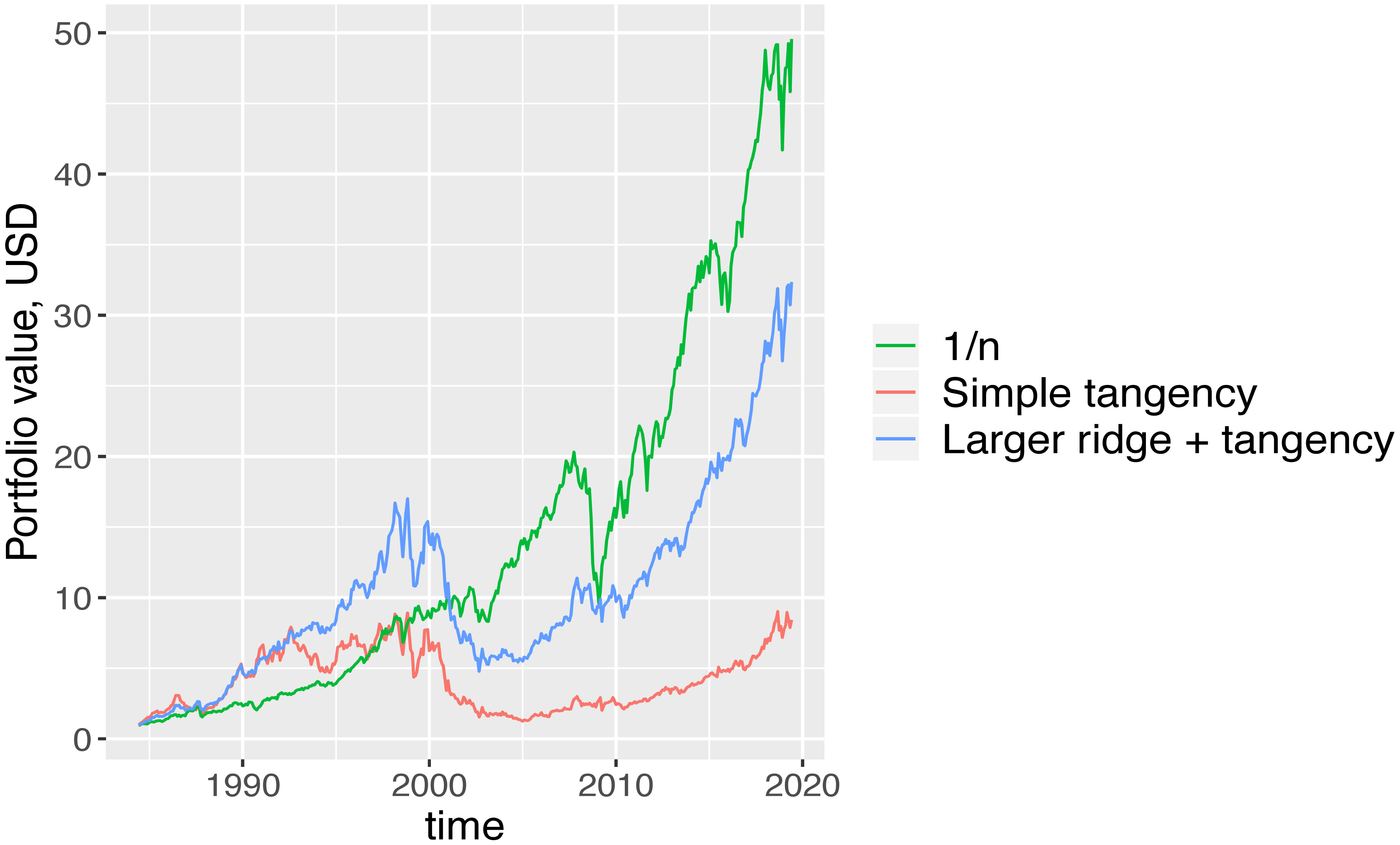
Maximum Sharpe ratio (tangency) portfolio with ridge:

$$w_{tangency} = \frac{1}{\mathbf{1}'(\Sigma + \lambda I)^{-1}\boldsymbol{\mu}}(\Sigma + \lambda I)^{-1}\boldsymbol{\mu}$$

Back to portfolio optimization: a little bit of ridge



Back to portfolio optimization: more ridge?



Shrinkage to the mean

Suppose that all the stocks truly have the same expected return, μ.

In sample, however, they all seem different.

The larger is the number of portfolios, the higher is the probability that the highest rate of return is actually overestimating the truth, and the lowest rate of return is underestimating it.

Example:

- under the null hypothesis of 0 true return, there is still 5% chance to get a significant t-stat.
- The higher is the number of draws, the more likely you are to observe something really large, or really small (multiple testing issue)

Idea: shrink sample expected returns towards their cross-sectional average

Shrinkage to the mean

Shrinking returns towards their cross-sectional average is equivalent to shrinking tangency portfolio weights to that of the minimum variance portfolio:

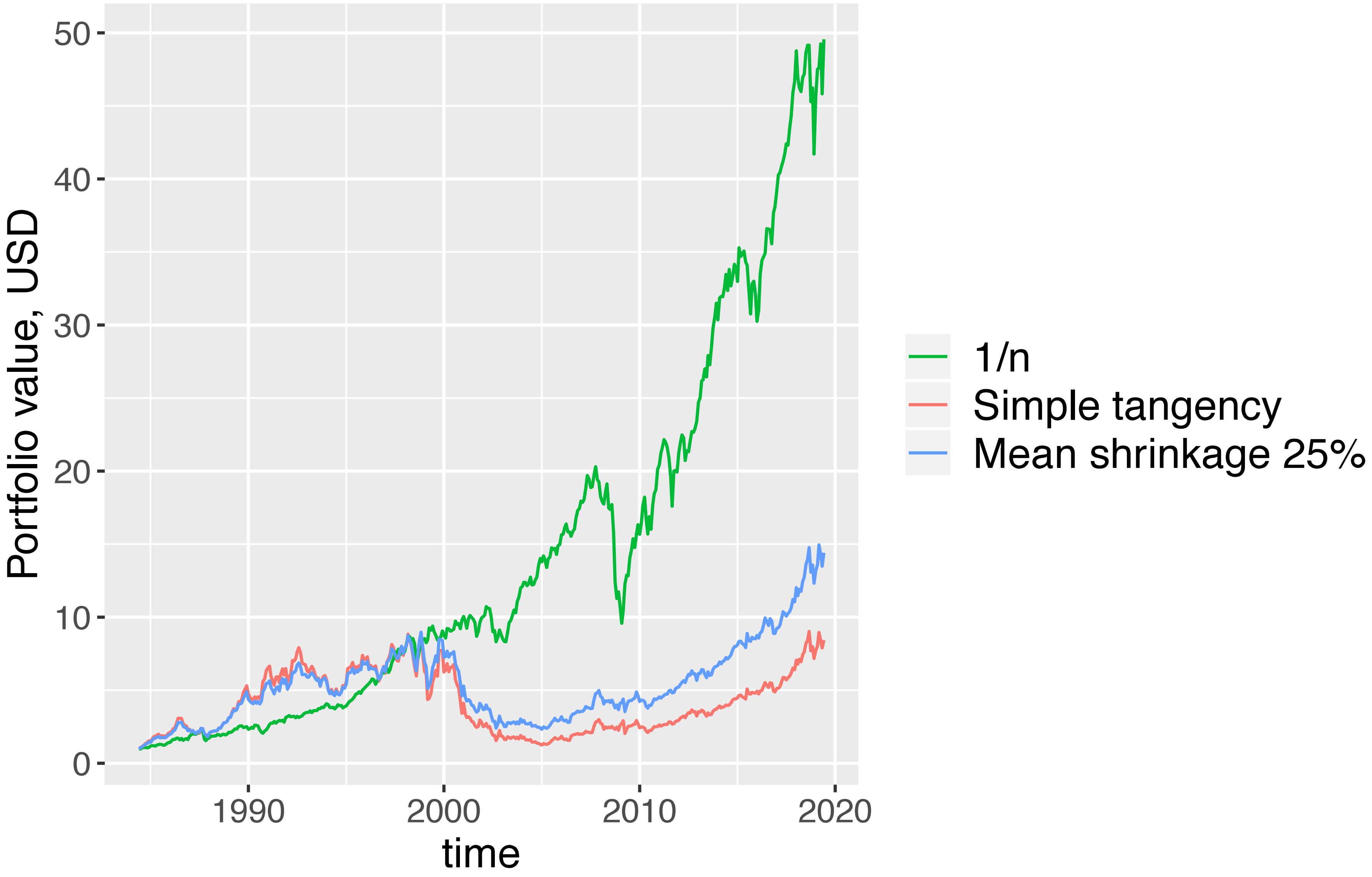
$$RSS(\lambda) = w'\Sigma\, \mathrm{w} + \lambda_1(w'1 \; - 1) + \lambda_2(w'(\alpha\mu + (1-\alpha)\mu_{average} \; - \mu_0).$$

Maximum Sharpe ratio (tangency) portfolio with shrinkage to the mean:

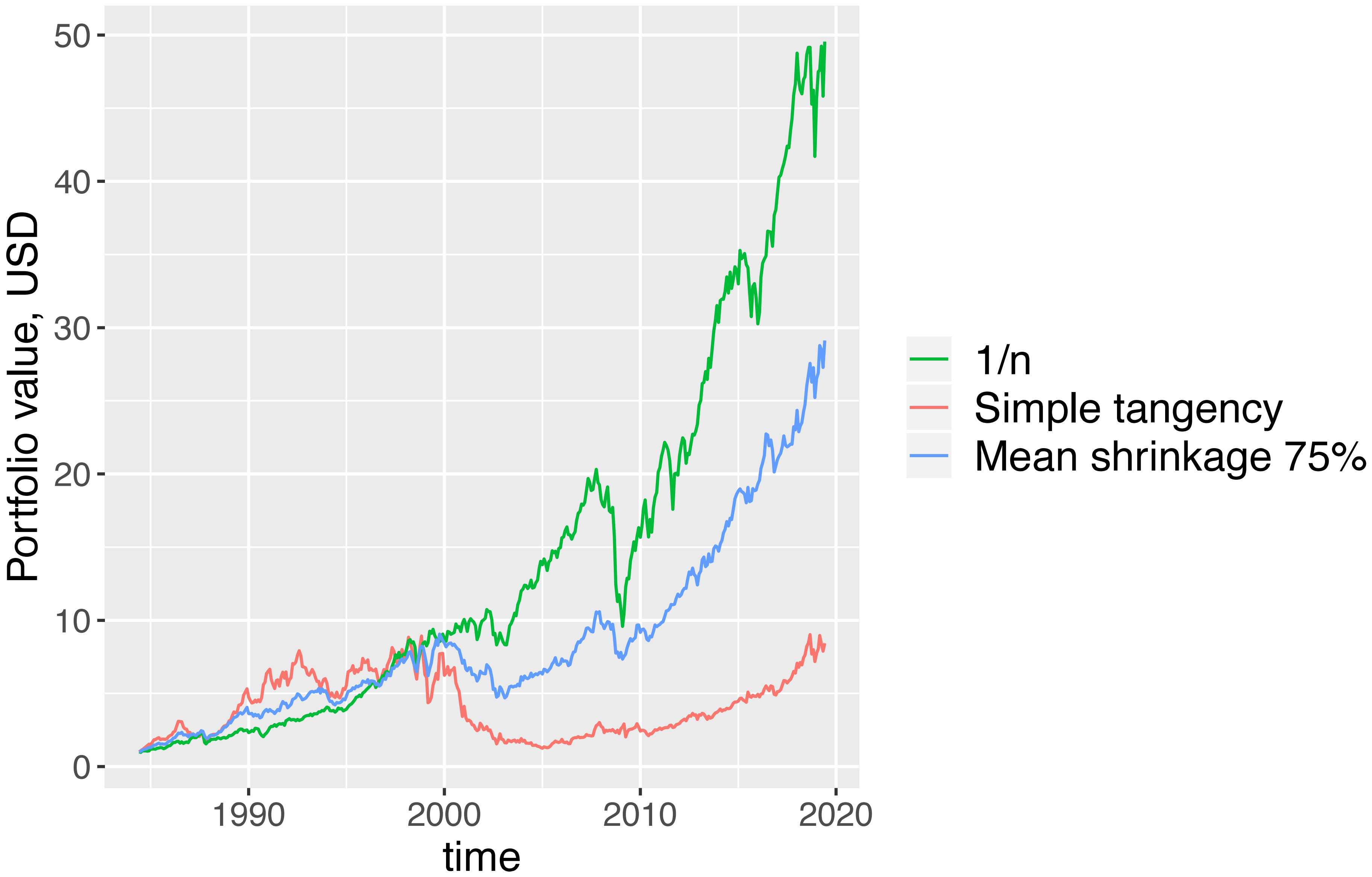
$$w_{tangency} = \alpha \frac{1}{\mathbf{1}'(\Sigma + \lambda I)^{-1}\boldsymbol{\mu}} \Sigma^{-1}\boldsymbol{\mu} + (1-\alpha)\frac{1}{\mathbf{1}'(\Sigma + \lambda I)^{-1}\mathbf{1}} \Sigma^{-1}\mathbf{1}$$

$$w_{tangency} = \alpha w_{naive\ tangency} + (1-\alpha) w_{min.variance}$$

Shrinkage to the mean: 25%



Shrinkage to the mean: 75%



LASSO

Least Absolute Shrinkage and Selection Operator is also a way of shrinking coefficients, but unlike ridge, it can set some parameters exactly to 0:

The lasso coefficients minimize the following object in a linear model:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|$$

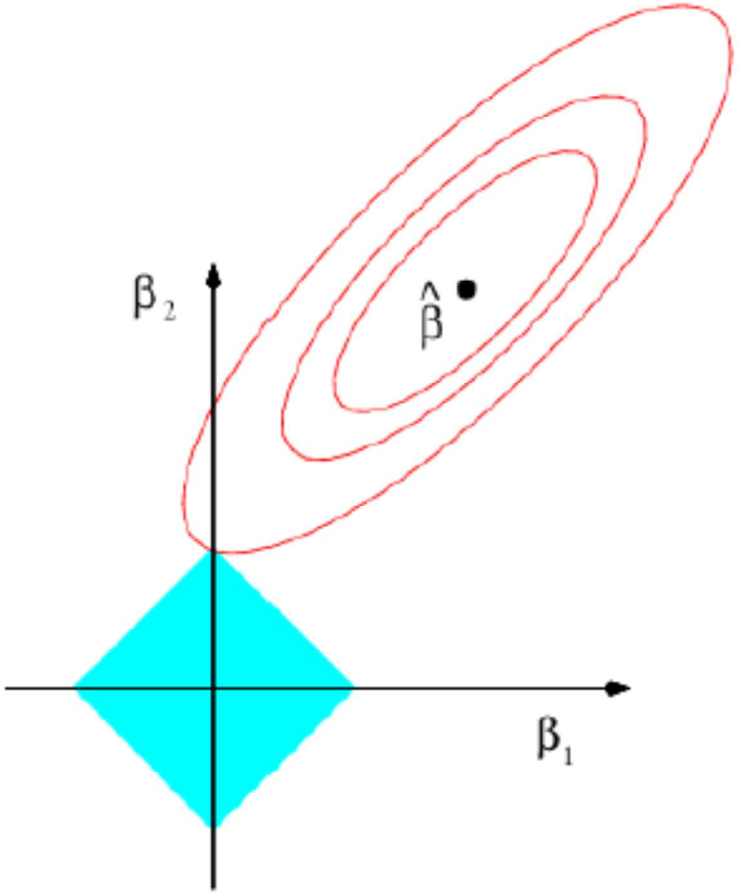
Note that the penalty relies on L1 norm, instead of L2 , like in ridge.

Generally, no closed-form solution.

When the tuning parameter is large, it will force some of the coefficients to be exactly 0. Hence, lasso provides simultaneous parameter selection and estimation.

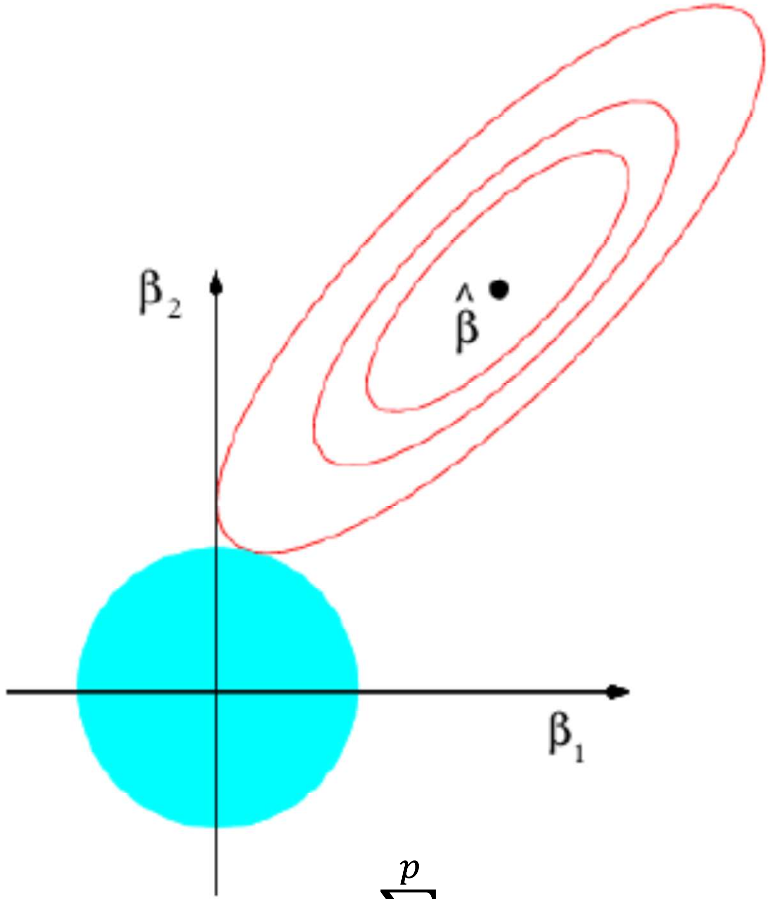
Lasso vs ridge

OLS contour + Lasso penalty



$$\sum_{i=1}^{p} |\beta_i| \leq t$$

OLS contour + ridge penalty



$$\sum_{i=1}^{p} \beta_i^2 \leq t$$

Portfolio optimization as a linear regression

Note that the standard tangency portfolio is the solution to the following problem:

$$w = \arg\min (y - Xw)' \, (y \ - Xw)$$

$$w = \arg\min (\Sigma^{-0.5}\mu \ - \Sigma^{0.5}w)' \, (\Sigma^{-0.5}\mu \ - \Sigma^{0.5}w)$$

where $\Sigma^{0.5}$ is the square root of the variance-covariance matrix, Σ, i.e.

$\Sigma^{0.5} = B$, s.t. $BB = \Sigma$

Check:

$$w = (X'X)^{-1} \, X'y = (\Sigma^{0.5'}\Sigma^{0.5})^{-1} \, \Sigma^{0.5}\Sigma^{-0.5}\mu = \Sigma^{-1}\mu$$

Hence, we can write the following OLS-lasso version for portfolio optimization:

$$w = \arg\min (\Sigma^{-0.5}\mu \ - \Sigma^{0.5}w)' \, (\Sigma^{-0.5}\mu \ - \Sigma^{0.5}w) + \lambda \sum_{i=1}^{p} |w_i| + \lambda_0 (1 - w'\mathbf{1})$$

Cross-validation to pick shrinkage

Selecting tuning parameters with a 5-fold cross-validation (CV)

1. Divide the data used for estimation into 5 parts
2. Select 4/5 of the data, and estimate the model for each λ, save the coefficients
3. Fit the model on the 5^{th} part of the data, compute prediction loss (SR gain, etc) = CV(1)

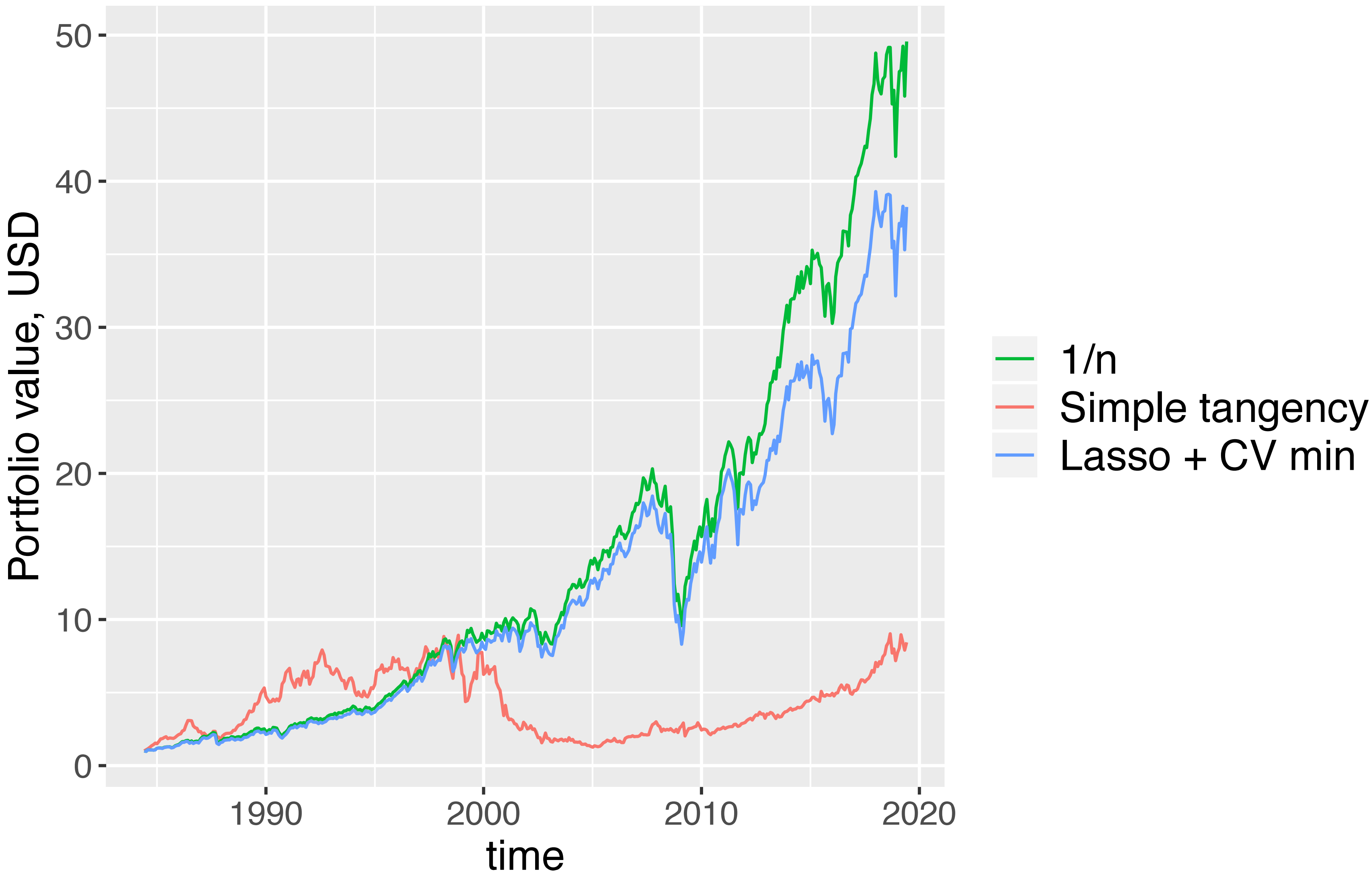
4. Repeat steps 1-3 for other 4 parts, and average over 5 numbers = CV(5)

5. Pick the value of the tuning parameter that leads to smaller CV, λ_{CVmin}

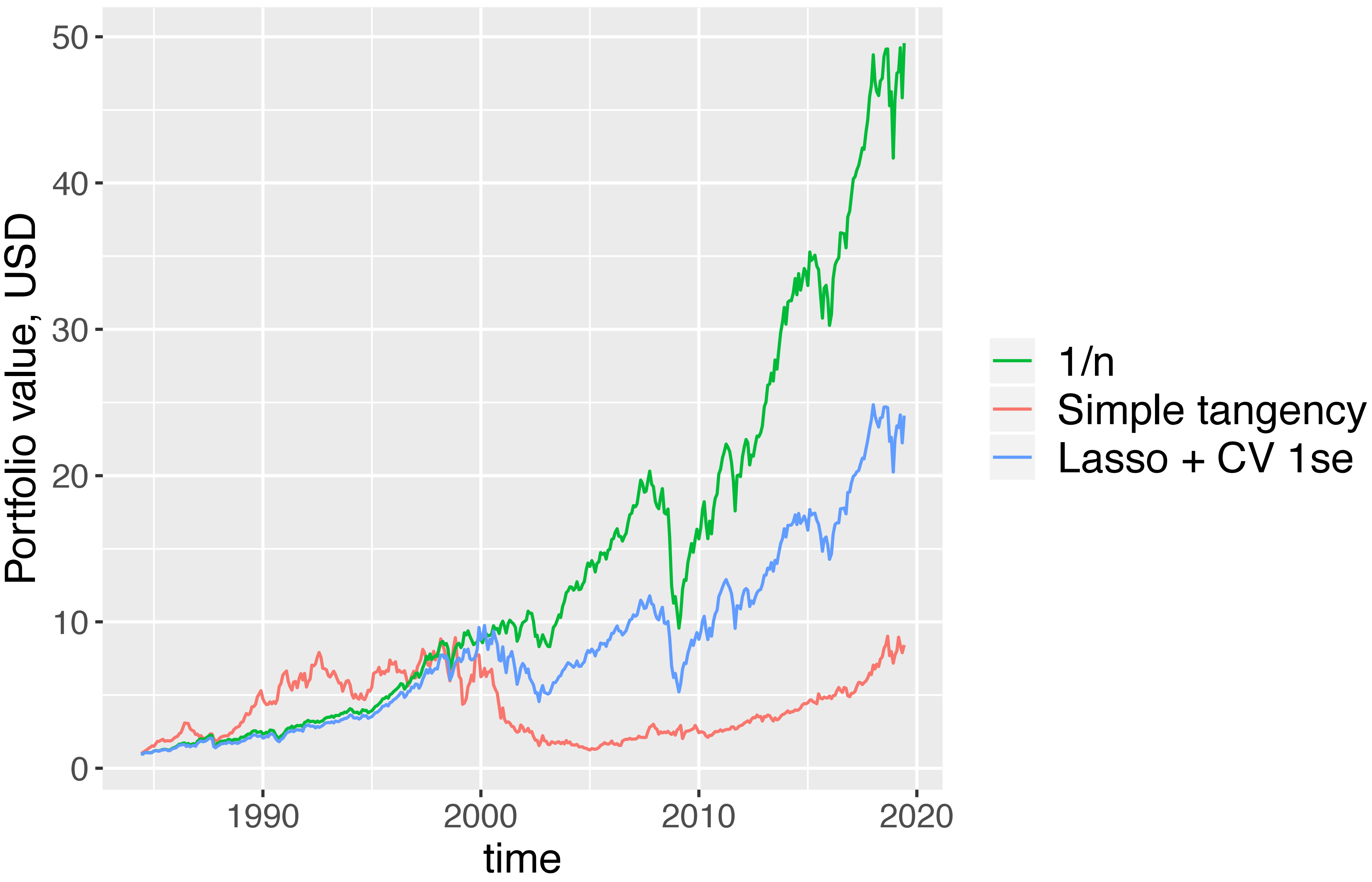
NB: it is also widespread to pick the smallest tuning parameter within 1 standard deviation of λ that minimizes CV: λ_{1se} .

- Leads to sparser, more conservative models that are still performing well.

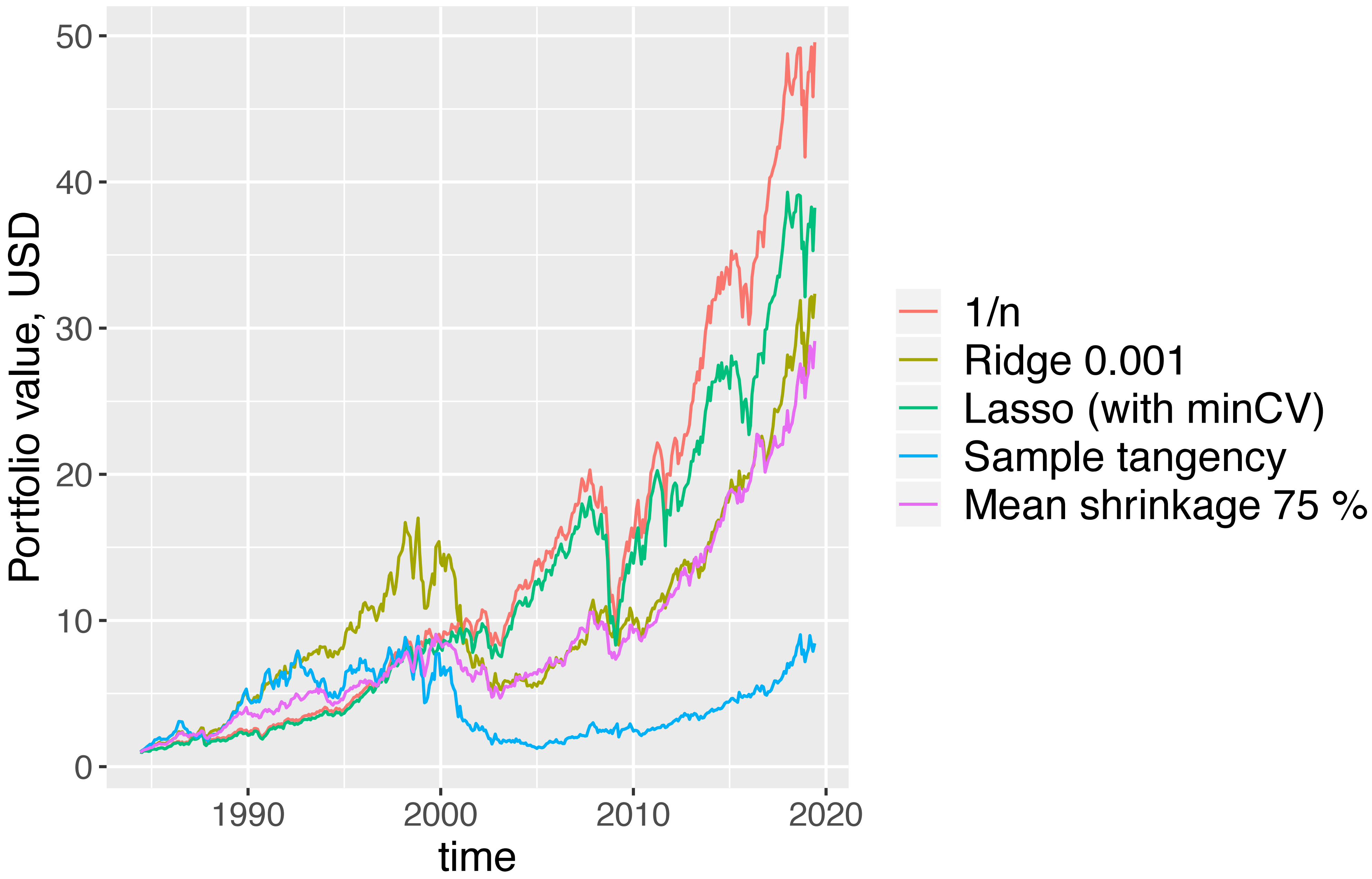
Lasso + λ that minimizes CV(5)



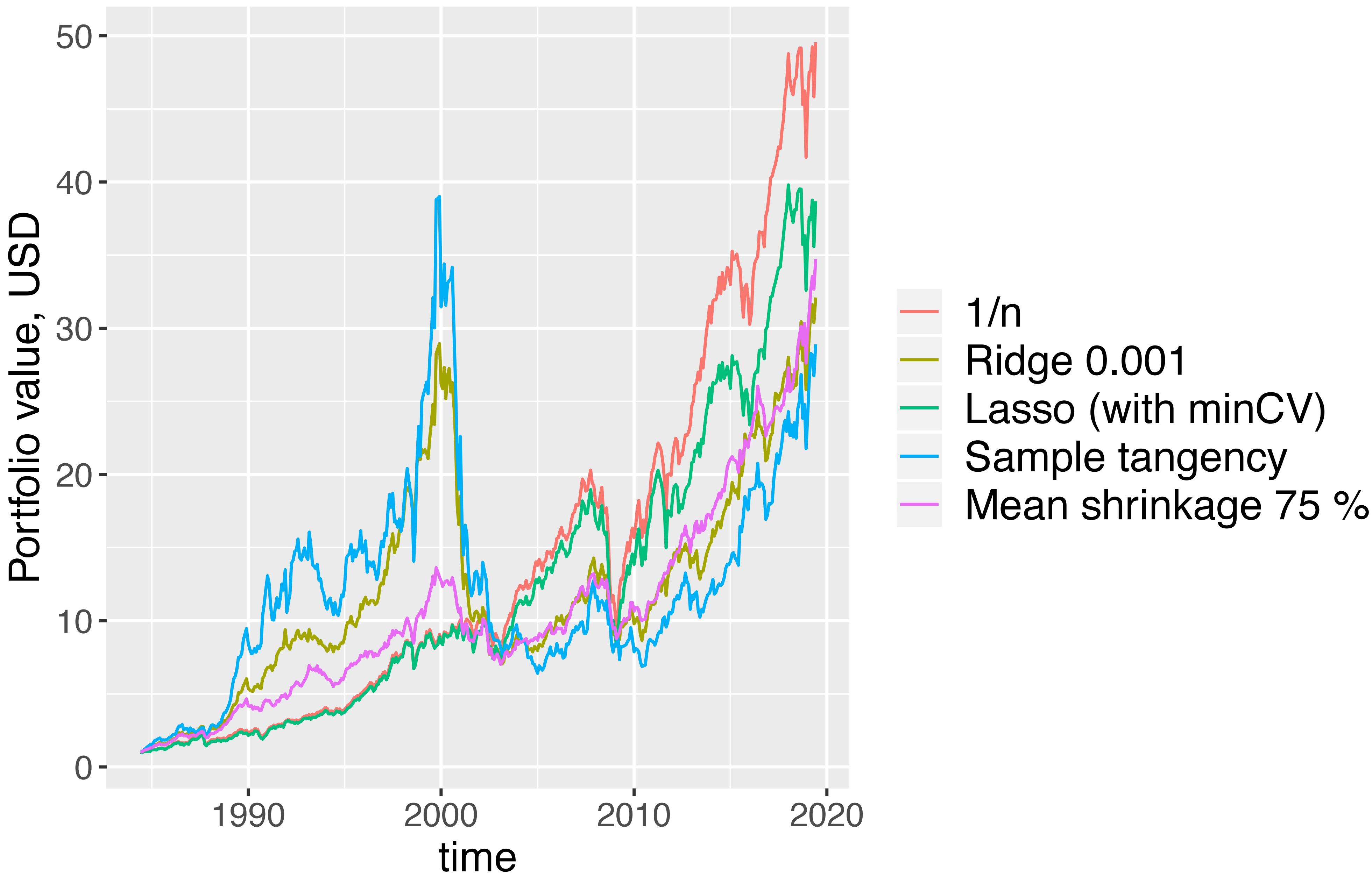
Lasso + λ within 1 s.d. of CV from λ_{CVmin}



Putting it all together: OOS investment for 5 years



Putting it all together: OOS investment for 1 year



OOS performance

Sample optimization tends to overfit the data, and hence leads to

- Extreme short and long positions
- Large turnover
- Substantial tail risk: e.g. after riding the dot com bubble, any portfolio would be wiped out
- Ridge stabilizes the loadings, but does not take into account expected returns
- Lasso and mean shrinkage focus primarily on expected returns
- Optimal solution: a mix of different approaches

- Is 1/n really so good?

Turnover and costs

- Equally weighted portfolios need to be rebalanced every time a stock appreciates/depreciates relative to the benchmark, 1/n weight.
- This often leads to turnovers 10-100 times larger that value-weighted portfolios
  - E.g. EWI S&P 500 is rebalanced every quarter
- Pro:
  - Protection against downturns in the biggest sectors
- Con:
  - Largely exposed to small caps and microcaps
  - Low liquidity
  - High volatility
  - Large transaction costs
  - Higher fees (e.g. 0.2% expense ratio for Invesco S&P 500 EWI ETF)
  - Short-term capital gains are often s.t. higher taxes
  - A lot of the return is loading on reversal and other short-term risk factors (later on this!): this is beta, not alpha.

Final thoughts

- The ideas of the standard portfolio optimization are universal
- Diversification matters!
- For a large number of assets, there are additional concerns:
  - Parameter stability
  - Overfitting
  - Turnover
  - Extreme long-short positions
  - Costs, liquidity, etc
- Shrinkage allows to substantially alleviate these concerns:
  - Lasso is used to select portfolios and estimate their weights
  - Ridge helps to stabilize loadings over time, especially for highly correlated assets
  - Shrinkage to the mean deals with cross-sectional outliers
  - Choice of the tuning parameters matters