Research Question

Sample eigenvectors are used throughout the sciences as proxies for unobservable population eigenvectors

How good are these estimates?

The answer is, it depends on a lot of things:

1. the aspect ratio, p/n
2. the underlying data generation process

Why Simulations?

We look at eigenvectors in simulation because

1. we know the ground truth and can measure the error
2. we can adjust across different regimes

But we must keep in mind that simulations simplify empirical behavior

Our simulation is loosely crafted on empirical results for empirical robustness

Simulation Specifications

For p > 1 we will develop an estimated p-dimensional covariance matrix assuming returns follow a latent one-factor model:

$$X = \beta f + z$$

Only X, the simulated returns, are observed

If you make a covariance matrix in this form, your true eigenvalues will not be at 1

In the empirical data we have seen presented lambda seems to depend linearly on p

Parameters Used in the Simulation

β is randomly drawn from a multivariate normal distribution with mean 1 and standard deviation of 0.25

f (factor returns) are constructed from n random draws from a normal distribution with mean 0 and standard deviation σ = 0.16

z (idiosyncratic asset-specific returns) are constructed from a normal distribution with mean 0 and standard deviation δ = 0.2

S (sample covariance matrix) is equal to XX^T / n

Metrics of the Simulation

We are looking at the leading eigenvalue, λ^2, and the leading eigenvector

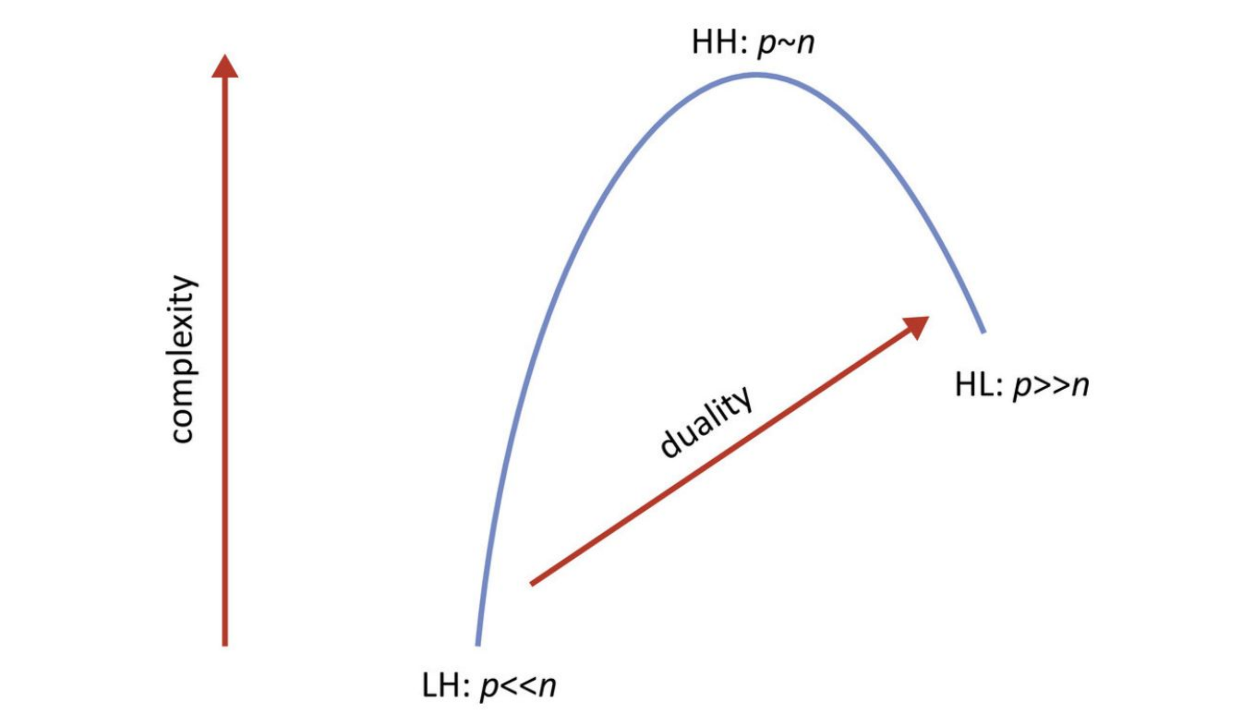
Our measure of how close our estimate, b = β / |β|, is to the truth, h (the leading eigenvector of S), is the absolute dot product of our estimate and h the leading eigenvector

An absolute dot product closer to 1 signals that our estimate is more accurate

Our measure of the leading eigenvalue is at times scaled by p to help visualize the eigenvalues linear relationship to p

Matrix Theory Regimes

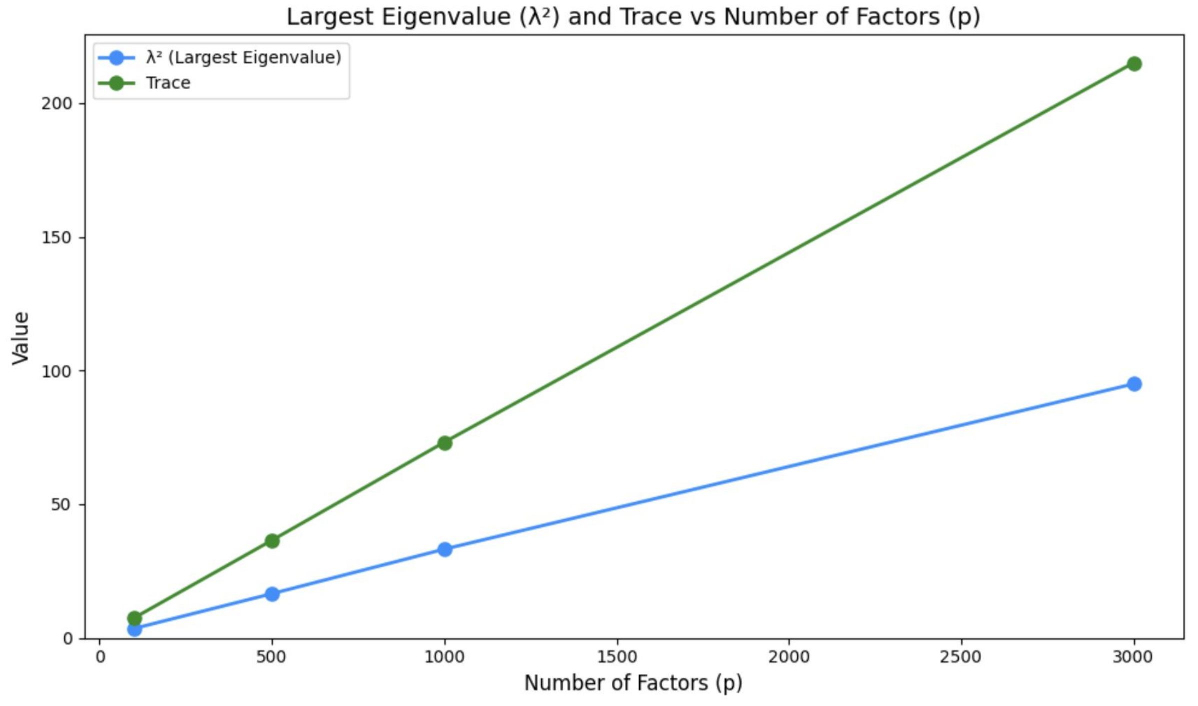
- Classical Statistics: p = 10, $n \to \infty$
- Classical Random Matrix Theory: $p \to \infty$, $n \to \infty$, and $p/n \to q$
- New Random Matrix Theory: n = 10, $p \to \infty$



Factor Models

In the data that we care about, empirical data, lambda seems to depend linearly on p

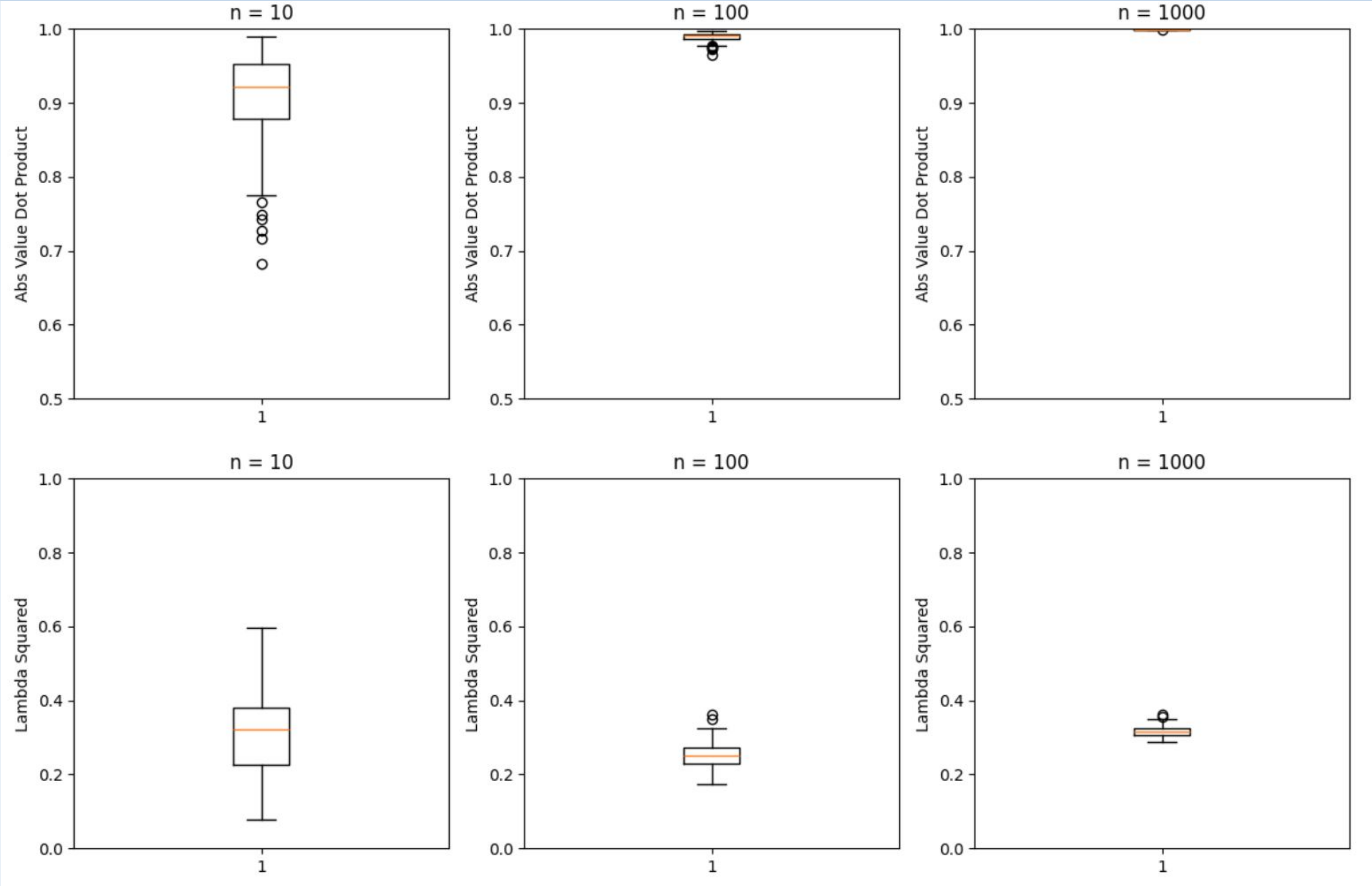
True Covariance Matrix: $\sigma^2 \beta\beta^T + \delta^2 I$



Classical Statistics,

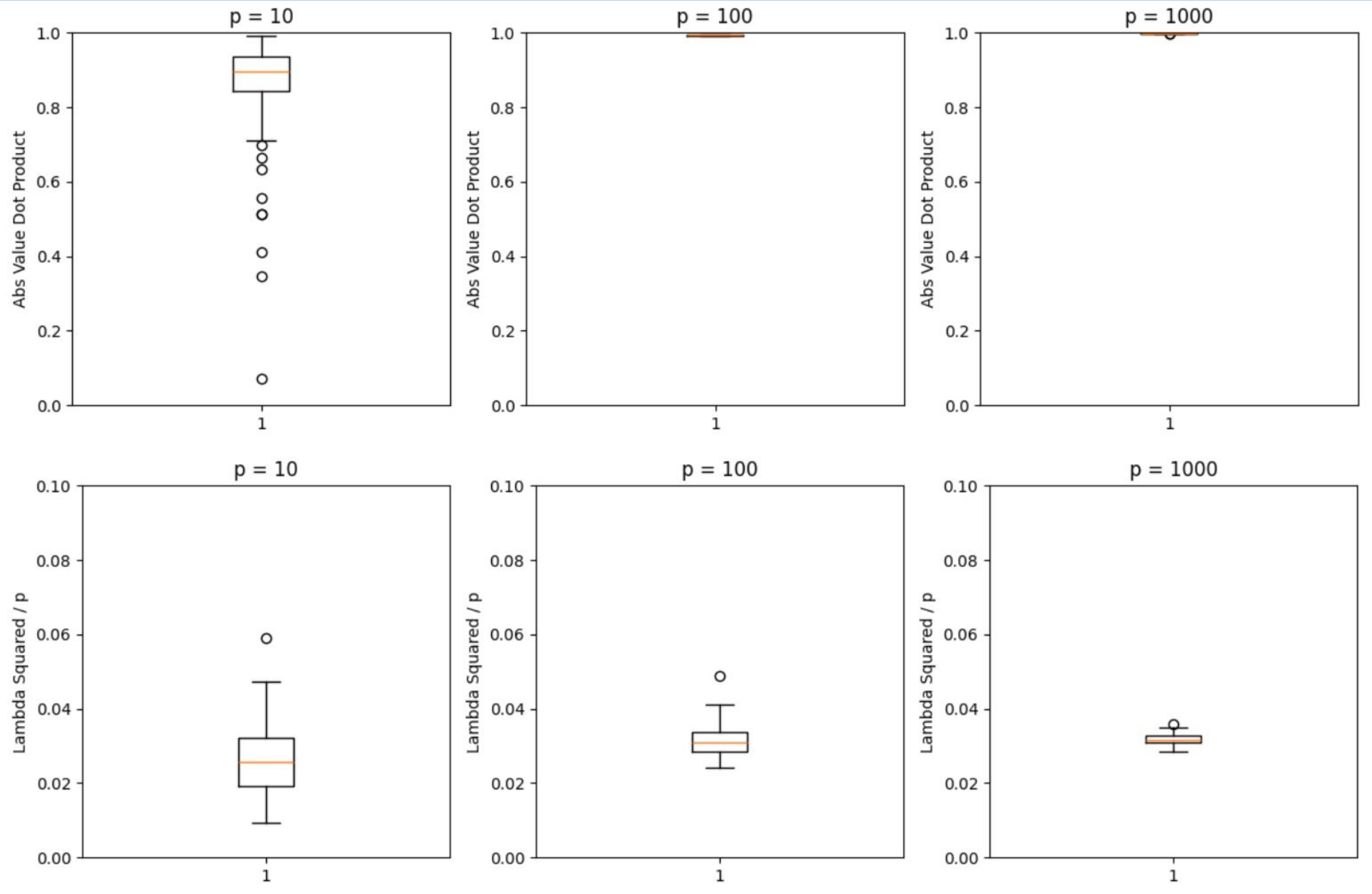
$p = 10, n \to \infty$

Medium Specific Volatility (δ = 0.20)



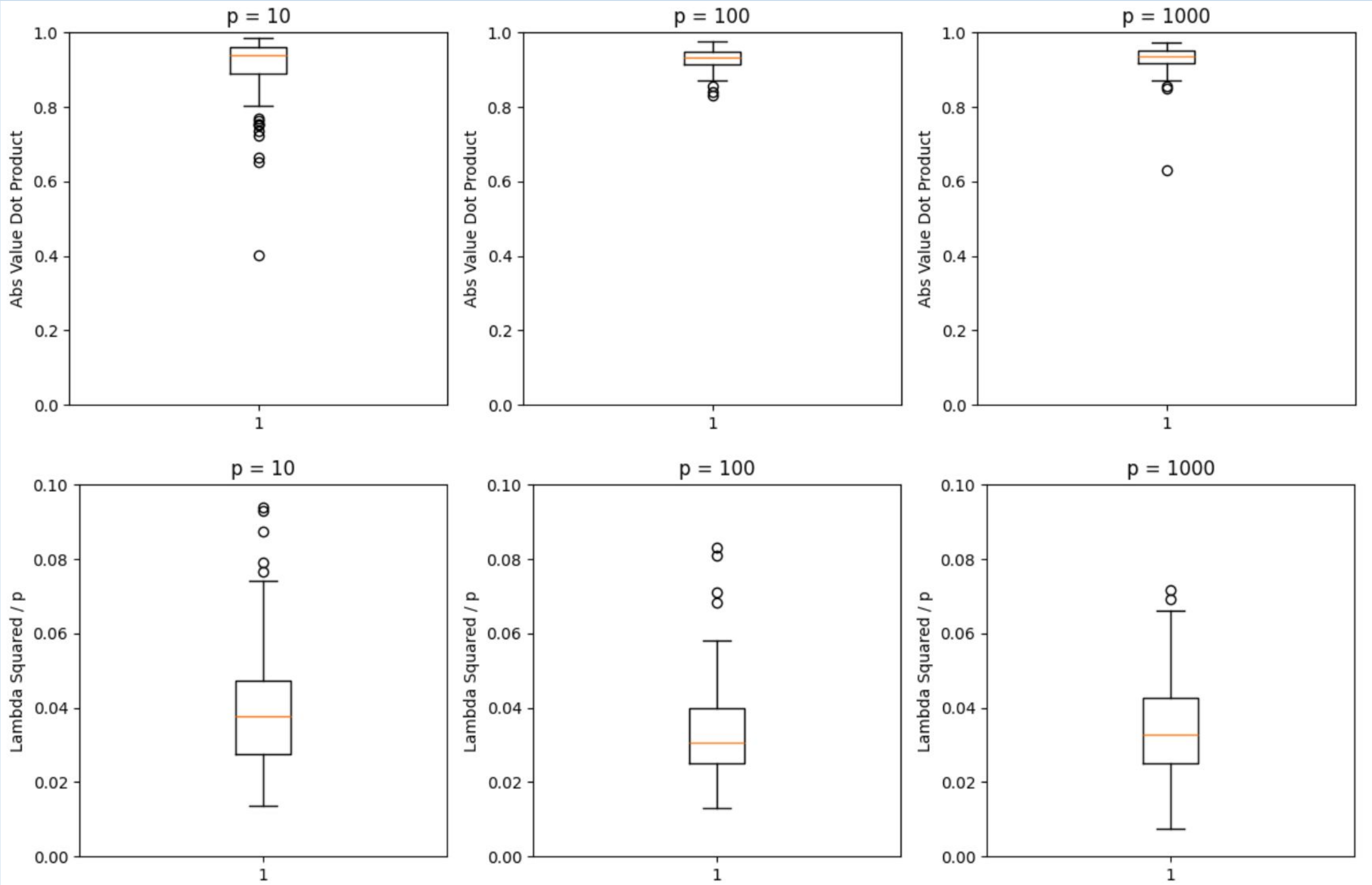
Classical Random Matrix Theory,
lambda = n/p $\rightarrow$ 1

Medium Specific Volatility (δ = 0.20)



New Random Matrix Theory,
$n = 10$, $p \to \infty$

Medium Specific Volatility (δ = 0.20)



Consistency Ratios

Simulation	n →	p →	λ^2 →	CR = λ^2 (n / p) →
Classical Statistics	∞	10	λ^2	∞
Classical Random Matrix Theory	p	n	∞	∞
Johnstone Spiked	p	n	λ^2	finite non-zero
New Random Matrix Theory (JSE)	10	∞	∞	finite non-zero