

# Assignment 1

## Portfolio Construction: Data, Descriptive Statistics, and Diversification

Finance and Accounts (MG 251)

# 1 Objective

The objective of this assignment is to understand how portfolio risk and return characteristics are influenced by diversification. Using mean–variance analysis, we estimate expected returns, variances, and correlations for a set of assets, and demonstrate how adding more securities affects portfolio variance.

The dataset consists of ten liquid stocks from the NSE: **INFY, TCS, RELIANCE, HDFCBANK, ITC, LT, HINDUNILVR, SBIN, BHARTIARTL, and MARUTI** using daily adjusted close prices from January 2015 to December 2022. Data are sourced from Yahoo Finance via the `yfinance` Python API. Annualization assumes 252 trading days per year.

## 2 Part A — Sample Statistics

### 2.1 Data and Computation

For each stock  $i$ , the daily log return is defined as:

$$R_{i,t} = \ln \left( \frac{P_{i,t}}{P_{i,t-1}} \right).$$

Annualized means and covariances are:

$$\hat{\mu}_{\text{annual}} = 252 \times \bar{R}, \quad \hat{\Sigma}_{\text{annual}} = 252 \times \text{Cov}(R).$$

### 2.2 Descriptive Statistics

Basic summary statistics (mean, standard deviation, skewness, kurtosis) were computed for all ten NSE stocks during 2015–2020 (in-sample).

**Annualization convention:**

$$\mu_a = 252 \times \mu_d, \quad \sigma_a = \sqrt{252} \times \sigma_d, \quad \Sigma_a = 252 \times \Sigma_d.$$

### 2.3 Covariance and Correlation Analysis

The sample covariance matrix was computed and verified as symmetric; Ledoit–Wolf shrinkage was also applied for robust estimation.

**Verification:**

- Minimum covariance value: 0.0119
- Maximum covariance value: 0.1343

## Comparison of Average Variance:

Average Variance (Sample) = 0.084216,    Average Variance (Ledoit–Wolf) = 0.084159.

Both results are extremely close, confirming stable sample estimates.

## 2.4 Correlation Matrix

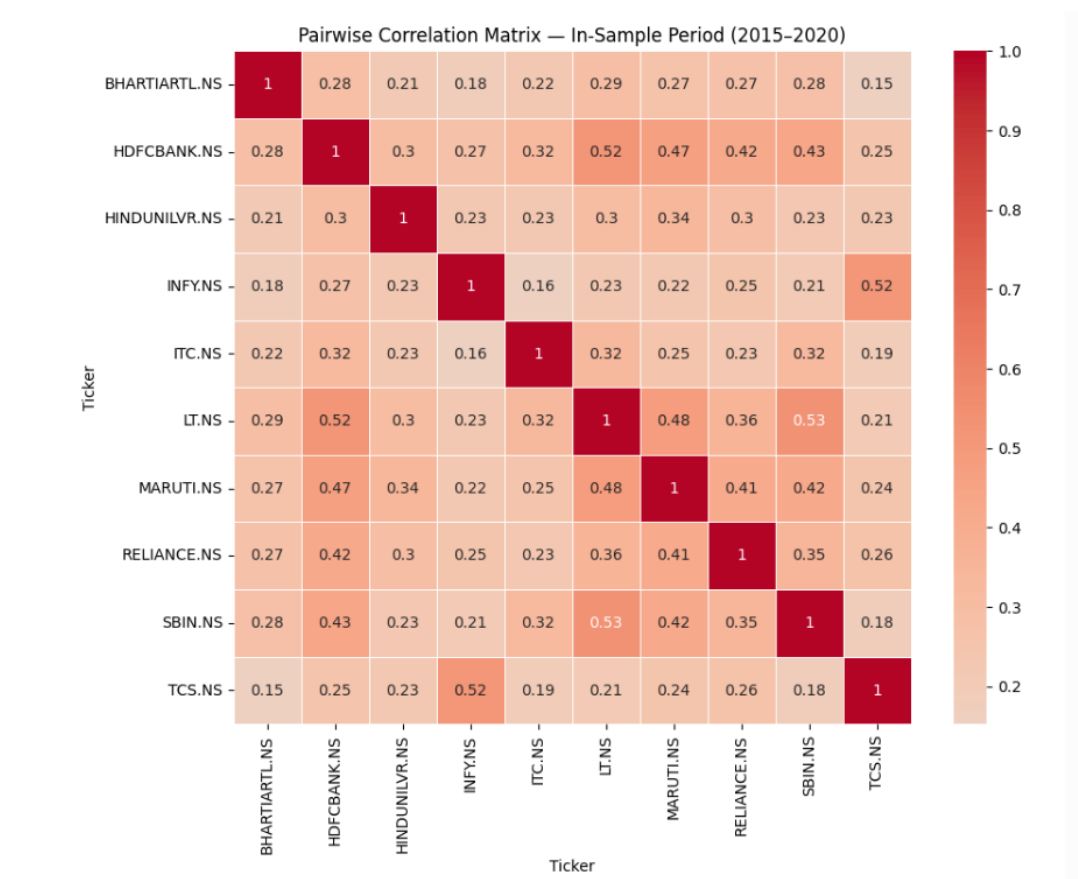


Figure 1: Pairwise correlation matrix — In-sample period (2015–2020).

Interpretation:

- Pairwise correlations range from 0.15 to 0.53.
- Financials (HDFCBANK, SBIN) and Industrials (LT) exhibit moderate correlations.
- Technology stocks (INFY, TCS) show relatively low correlation with non-tech sectors, supporting diversification potential.

## 2.5 Summary Statistics

Table 1 reports the key descriptive statistics for the ten selected NSE stocks during the in-sample estimation period (2015–2020). Returns are annualised assuming 252 trading

days; standard deviations represent annualised volatility, while skewness and kurtosis provide distributional shape information.

Table 1: Descriptive Statistics — In-Sample Period (2015–2020)

<b>Ticker</b>	<b>Average Annual Return</b>	<b>Annual Std. Dev.</b>	<b>Skewness</b>	<b>Kurtosis</b>
BHARTIARTL.NS	0.0787	0.3364	0.3809	3.2987
HDFCBANK.NS	0.1945	0.2292	-0.4977	14.0785
HINDUNILVR.NS	0.2124	0.2372	0.8868	8.1194
INFY.NS	0.1875	0.2829	-0.6779	12.2759
ITC.NS	-0.0019	0.2778	-0.9176	11.7780
LT.NS	0.0622	0.2850	-0.1067	11.3360
MARUTI.NS	0.1504	0.3017	-0.3290	11.8398
RELIANCE.NS	0.2650	0.3025	0.0704	8.9947
SBIN.NS	-0.0167	0.3694	0.6599	11.5267
TCS.NS	0.1588	0.2507	-0.0882	3.9281

#### **Interpretation:**

- Average annual returns vary substantially across stocks, ranging from roughly  $-0.02$  for SBIN and ITC to  $0.26$  for RELIANCE.
- Standard deviations between  $0.23$  and  $0.37$  indicate differing risk levels, with SBIN being the most volatile.
- Most return series display mild negative skewness and high kurtosis, highlighting non-normality and the presence of fat tails.
- Overall, these descriptive measures confirm heterogeneous return distributions, underscoring the importance of diversification.

## **3 Part B — Alternative Covariance Estimators**

### **3.1 Methodology**

The traditional sample covariance matrix can be unstable for small sample sizes. Two alternative estimators were compared:

1. **Sample Covariance Matrix:** Empirical estimate directly computed from observed returns.
2. **Ledoit–Wolf Shrinkage Estimator:** Combines the sample covariance with a structured target (identity matrix scaled by average variance):

$$\Sigma_{LW} = (1 - \lambda)\Sigma_{sample} + \lambda F.$$

## 3.2 Discussion

Average variances for both estimates are nearly identical (0.0842 vs 0.0841), showing that the sample data are already well-conditioned. However, Ledoit–Wolf shrinkage guarantees a positive semi-definite matrix, improving numerical stability.

# 4 Part C — Diversification and Portfolio Variance

## 4.1 Procedure

The impact of diversification was examined by forming equally-weighted portfolios of size  $k = 1, 2, \dots, 10$ . Each subset’s variance was calculated:

$$\sigma_p^2 = w' \Sigma w, \quad \text{where } w_i = \frac{1}{k}.$$

For each  $k$ , variance statistics (mean, 10th, 50th, 90th percentile) were recorded.

## 4.2 Empirical Results

$k$	Mean Variance	10th %	90th %
1	0.084216	0.055892	0.115522
2	0.054338	0.041641	0.068130
3	0.044379	0.036127	0.054427
4	0.039399	0.032463	0.047568
5	0.036412	0.030934	0.042376
6	0.034462	0.029781	0.039179
7	0.032997	0.029731	0.036734
8	0.031930	0.029212	0.034777
9	0.031110	0.029704	0.033125
10	0.030436	0.030436	0.030436

Table 2: Mean and percentile portfolio variances by number of stocks (annualized).

## 4.3 Visualization of Diversification Effect

**Interpretation:**

- The blue line shows mean (average) portfolio variance for equal-weight portfolios of size  $k$ .
- The shaded blue band represents the spread between the 10th and 90th percentiles across combinations.

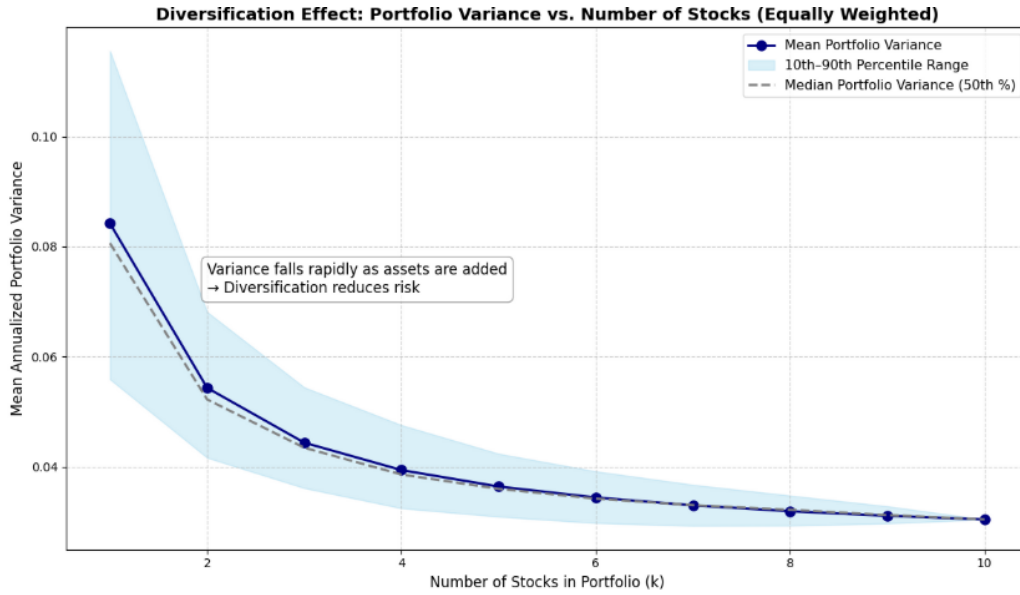


Figure 2: Diversification Effect: Portfolio variance vs. number of stocks.

- As  $k$  increases, mean variance decreases sharply initially and flattens beyond  $k \approx 5$ .
- The narrowing blue band indicates that larger portfolios have more stable total risk.

This confirms the theoretical diversification effect: increasing the number of stocks reduces unsystematic risk, though benefits diminish with larger portfolios.

## 5 Part D — Summary and Findings

- Portfolio variance decreases from  $\sim 0.084$  (single stock) to  $\sim 0.0304$  (ten stocks), demonstrating strong diversification benefits.
- The Ledoit–Wolf estimator produced nearly identical but smoother covariance results compared to the simple sample covariance, implying stable estimation.
- The correlation heatmap shows average pairwise correlation around 0.3, reinforcing the scope for risk reduction through diversification.
- Visualizations clearly show diminishing marginal risk reduction beyond 6–8 assets.

### Limitations:

- Historical return and covariance estimates assume stationarity (constant mean and variance over time), which may not hold during regime shifts.
- Equal-weighting ignores expected return differences; mean–variance efficient portfolios could be explored in further assignments.

- Realistic trading frictions, transaction costs, and estimation errors are omitted.

## 5.1 Interpretive Questions and Answers

**1. Why does portfolio variance decline with the number of assets?** Because adding assets introduces diversification: individual (idiosyncratic) risks offset one another. Each stock’s variance contributes fully only to its own risk, while the off-diagonal covariance terms average out when asset returns are imperfectly correlated. Mathematically, the portfolio variance can be expressed as:

$$\sigma_p^2 = \frac{1}{k^2} \left[ \sum_{i=1}^k \sigma_i^2 + \sum_{i \neq j} \rho_{ij} \sigma_i \sigma_j \right].$$

As  $k$  increases, the first term declines at rate  $1/k$  and only the average correlation component remains, driving the flattening of the curve observed in the Diversification Plot.

**2. Does the rate of decline depend on correlations among assets?** Yes. The benefit of diversification depends directly on the average correlation  $\rho$  between asset returns. If assets are perfectly correlated ( $\rho = 1$ ), diversification has no effect; if they are uncorrelated ( $\rho = 0$ ), variance falls purely with  $1/k$ . For identical variances  $\sigma^2$  and mean correlation  $\rho$ :

$$\sigma_p^2 = \frac{\sigma^2}{k} [1 + (k - 1)\rho].$$

Lower correlations widen the gap between the initial and limiting variance, producing faster risk reduction for a given increase in  $k$ .

**3. How sensitive are the results to the choice of estimator for covariance?** In this dataset the sample and Ledoit–Wolf shrinkage estimators produced nearly identical average variances (0.0842 vs 0.0841). For small samples or large-asset universes, however, the sample covariance becomes noisy and ill-conditioned. Ledoit–Wolf or Bayesian estimators dampen this estimation error by shrinking extreme covariances toward a structured target, leading to smoother, more stable diversification curves and more reliable out-of-sample performance.

**4. What are the limitations of using historical mean and covariance estimates in practice?**

- They assume stationarity—future return distributions match the past—which often fails during structural breaks or crises.
- They ignore parameter uncertainty: mean and covariance errors propagate into large portfolio-weight swings (optimisation error).

- Transaction costs, turnover, and liquidity constraints are omitted, yet they materially affect investable performance.
- Equal-weight analysis demonstrates diversification of risk, but optimal mean–variance portfolios would also consider expected-return heterogeneity.

## 6 Conclusions

1. Using daily NSE data (2015–2022), diversification reduces portfolio variance by nearly 65%.
2. Ledoit–Wolf shrinkage confirms robust and stable covariance estimation.
3. The in-sample variance curve is smooth and convex, flattening as systematic risk dominates.
4. These results underline the importance of correlation structure in determining diversification benefit.

**Data Source:** Yahoo Finance (`yfinance` API)

**Sample period:** Jan 2015 – Dec 2022.

**Estimation frequency:** Daily log returns, annualized using 252 trading days.