UCLA
Department of Statistics
& Data Science

Investment Strategies & Market Analysis for Real Estate in Los Angeles County

Anum Damani & Sean Mulherin

STATS 417

Summer 2024

INTRODUCTION

- Publicly Traded Real Estate Investment Trusts (REITs)
  - Companies that deal with income-producing real estate [1]
- Topics Covered
  - REIT Investment Portfolio
    - Diversification
    - Optimization
    - CAPM Model for Valuation
  - Interrupted Time Series & Change Point Analysis
  - Hedonic Pricing of Homes in Los Angeles

PORTFOLIO DIVERISICATION

- O: Walmart, Costco, Kroger, etc.
- PSA: public storage
- AMT: cell/broadcast towers
- NLY: mortgages
- VNO: office buildings, retail stores in NYC
- DHC: health care properties
- GTY: gas stations
- INN: hospitality, lodging, resorts

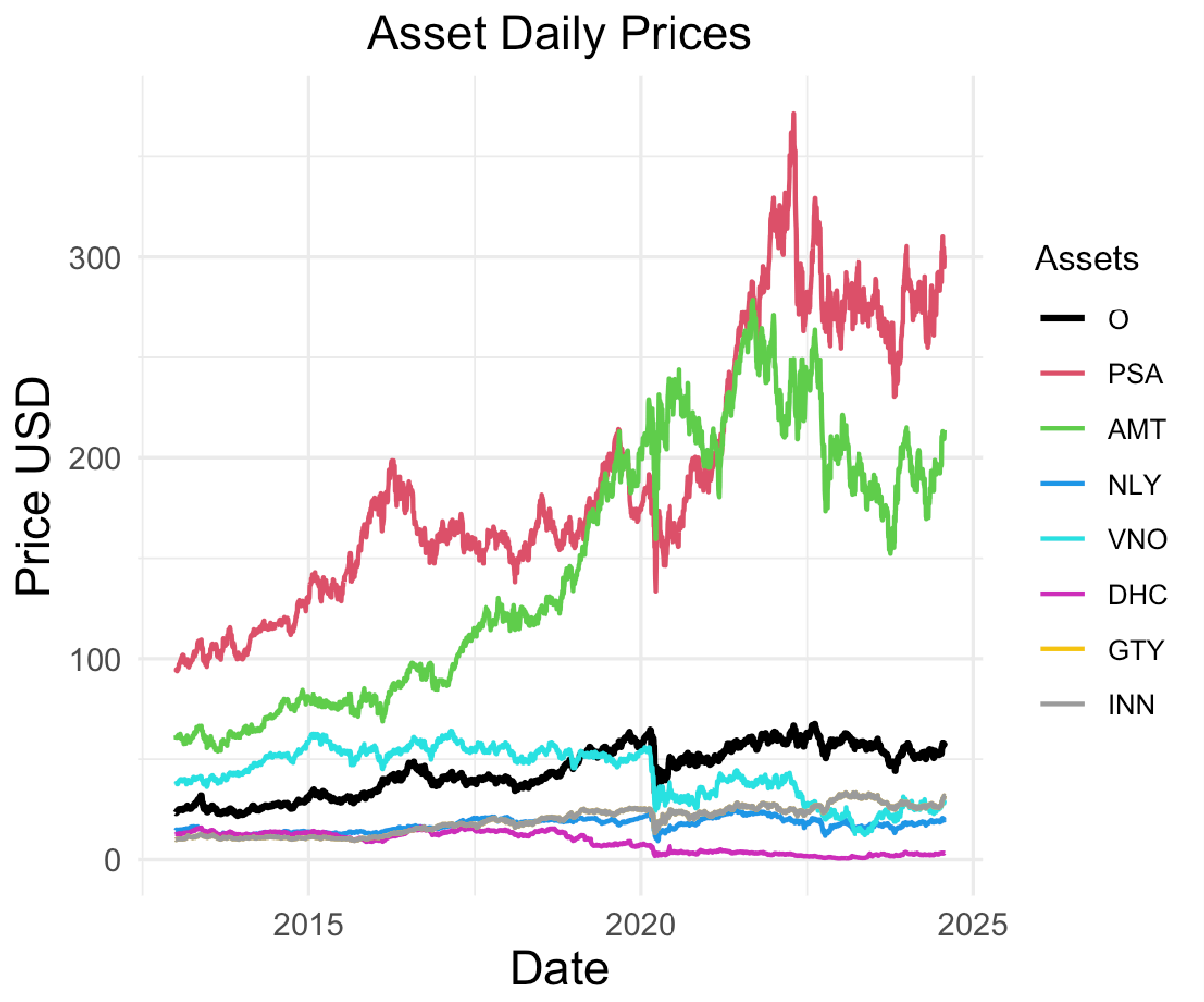


Figure 1: Asset Daily Prices

PORTFOLIO DIVERSIFICATION

- Most stocks are positively correlated with each other, which validates our initial expectation since these curated stocks are related to real estate.

- DHC (health care) is negatively correlated with GTY (gas). Also, O (convenience stores) is negatively correlated with GTY (gas).
    - These negative correlations create suspicion and indicate volatility.

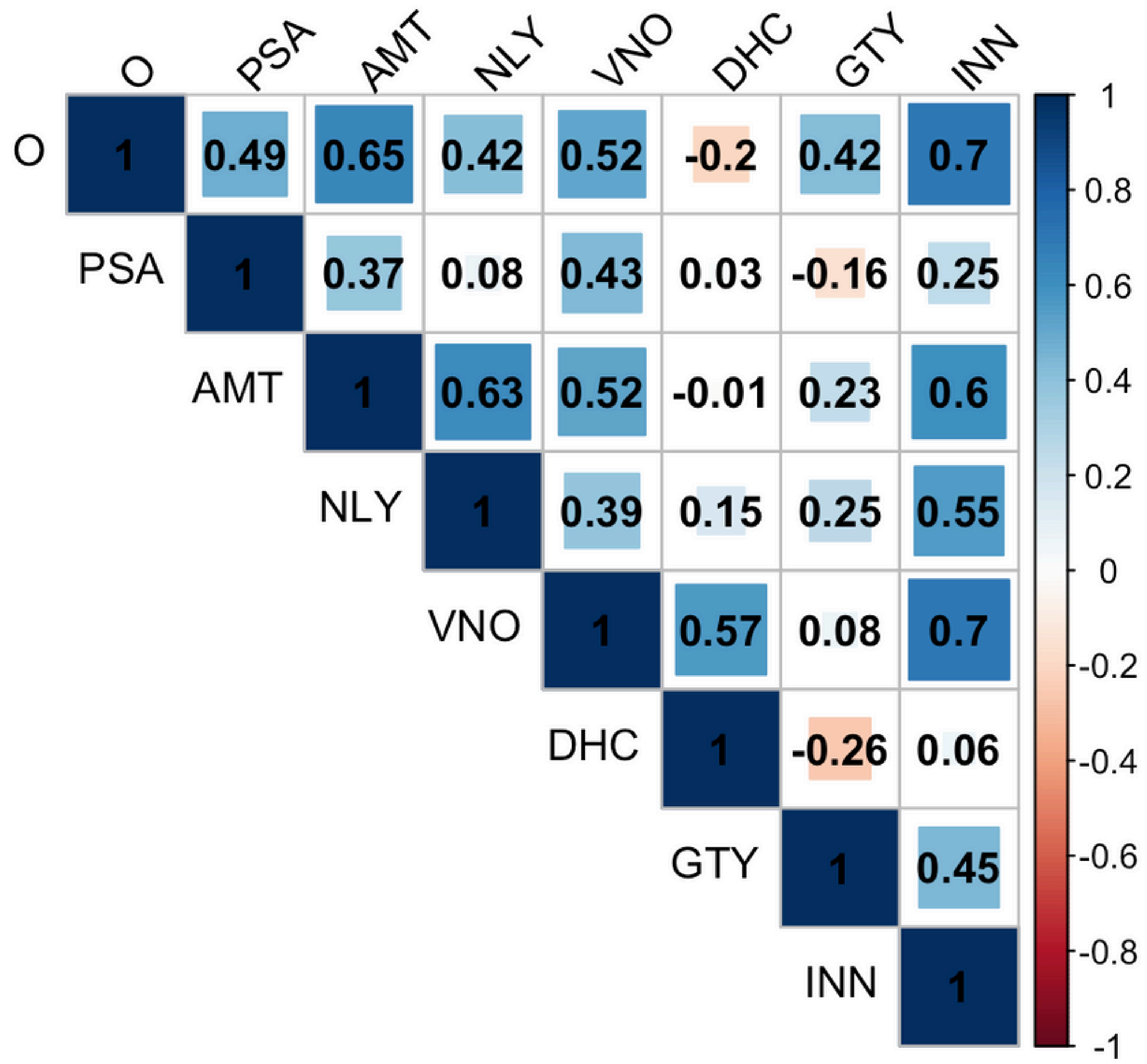


Figure 2: Correlation Matrix

PORTFOLIO OPTIMIZATION

- The simulated portfolio efficiency frontier allows us to visualize the optimal portfolio, indicated by the point on the red, dashed line.

- The optimal portfolio (red) is associated with the maximum Sharpe Ratio.

- Optimal Portfolio (red)
  - Sharpe Ratio = 0.556
  - Expected Return = 11.9%
  - Standard Deviation = 13.7%

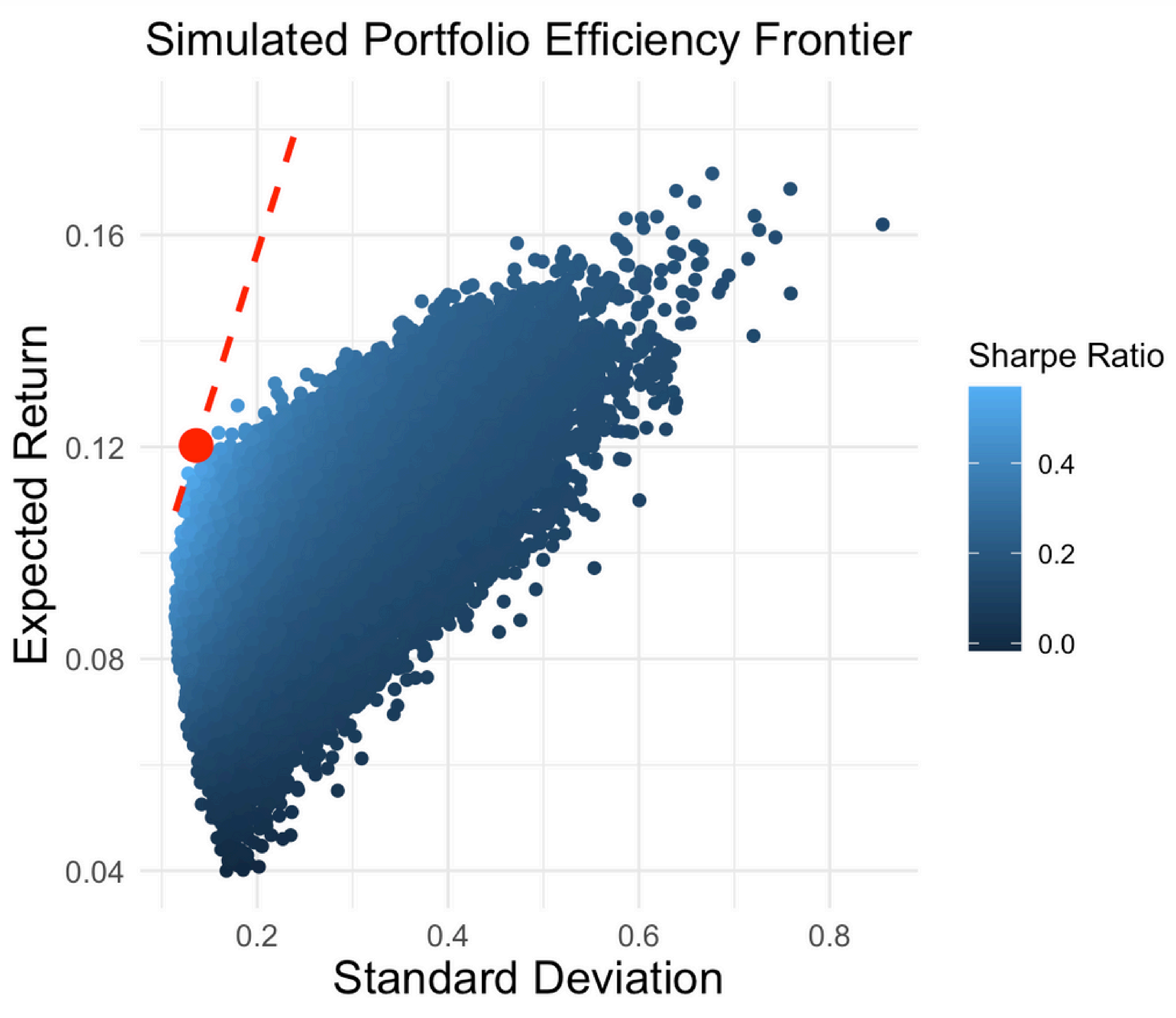


Figure 3: Simulated Portfolio Efficiency Frontier

PORTFOLIO OPTIMIZATION

- The optimal portfolio presented in the previous slide is comprised of the following asset weights.

- All asset weights sum to 1 with no negative values allowed (i.e. no shorting).

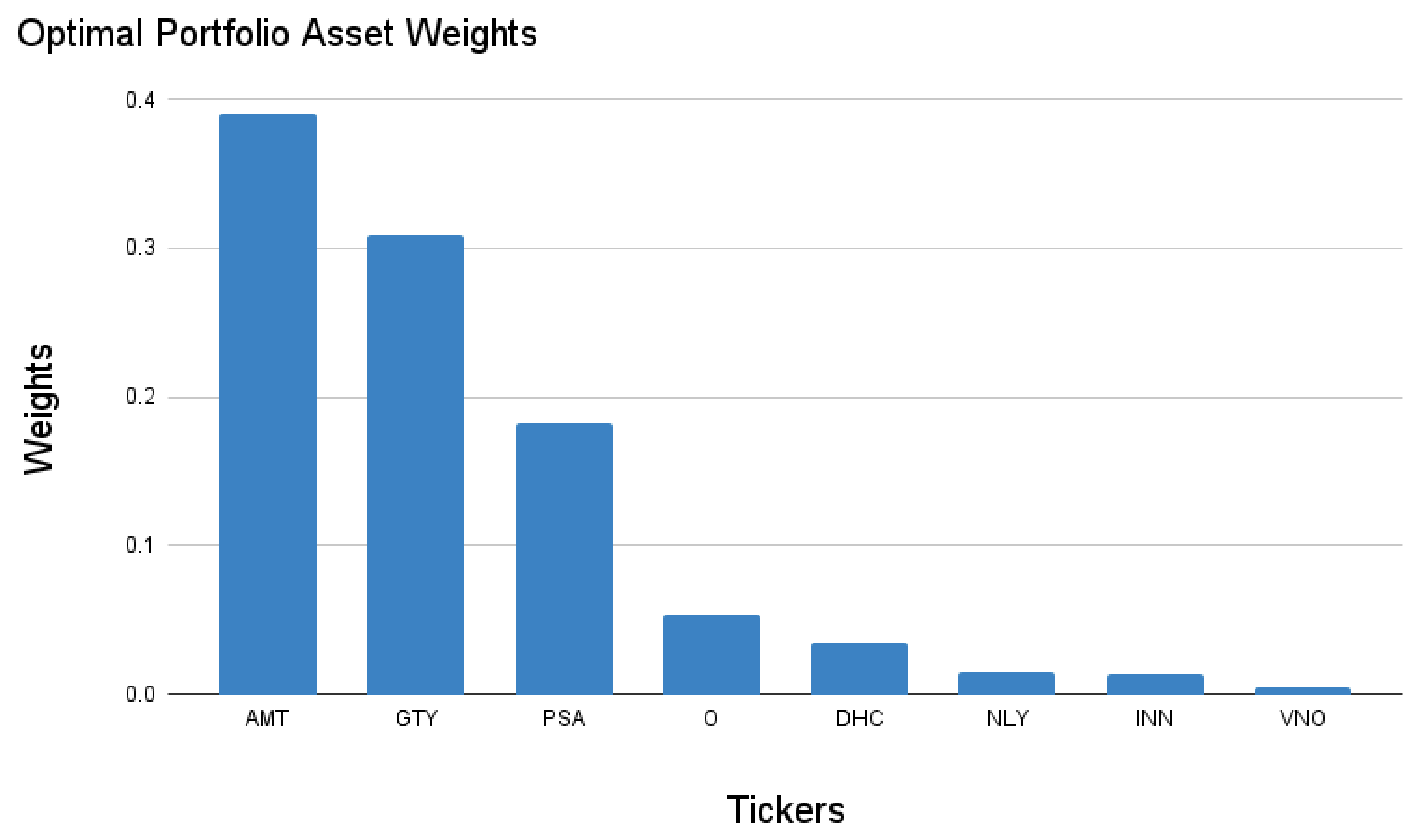


Figure 4: Optimal Weights

CAPM: PORTFOLIO VALUATION

- Our portfolio (red) is undervalued since it is above the security market line.

- In other words, we believe the expected returns are higher than the market's valuation.

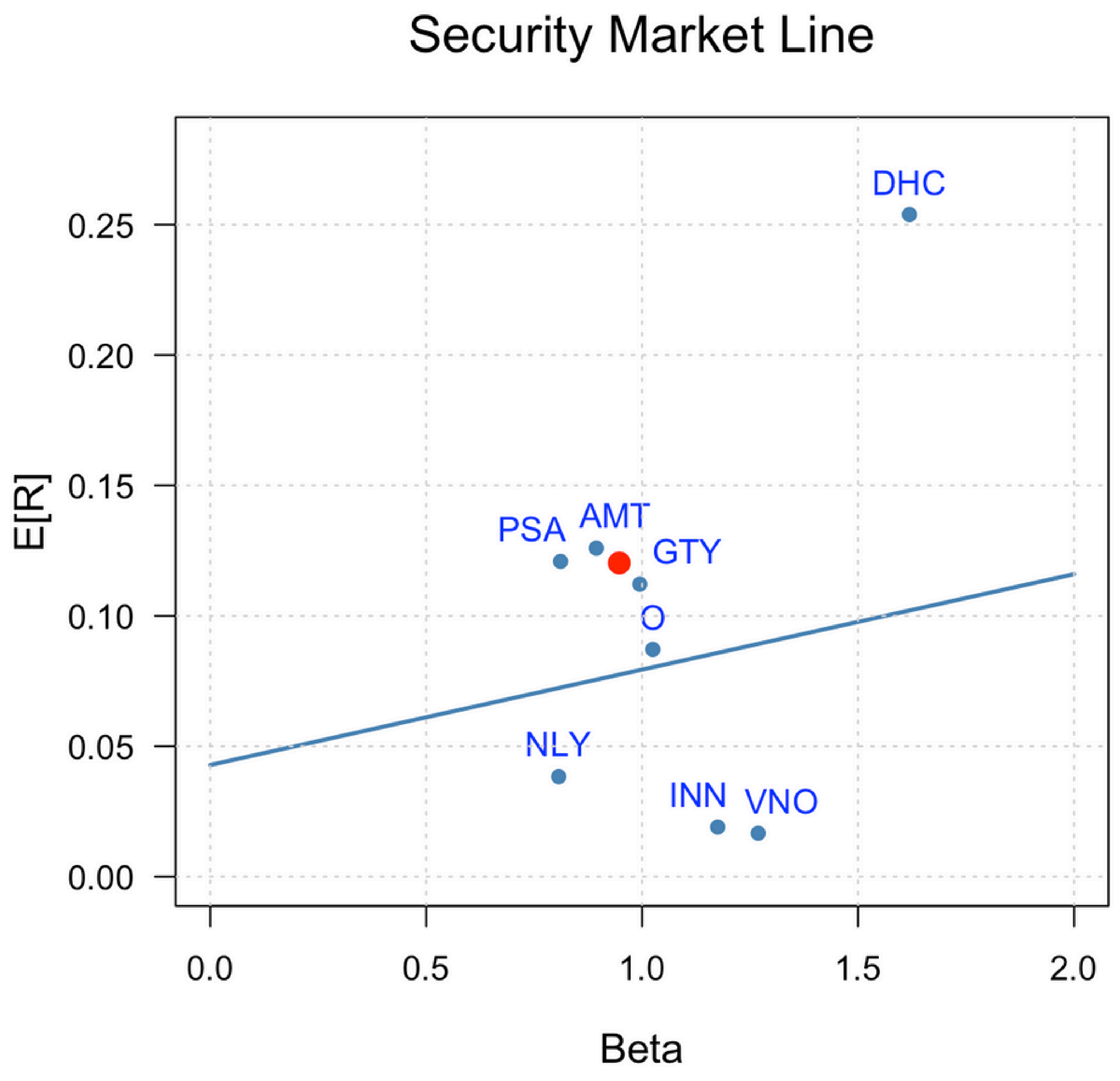


Figure 5: CAPM

LOS ANGELES COUNTY HOME PRICES

- Home prices for Beverly Hills continued to rise after the COVID-19 pandemic began.
- Notably, housing prices for Santa Monica, Silver Lake, South LA, and the area around USC began to plateau after 2020.

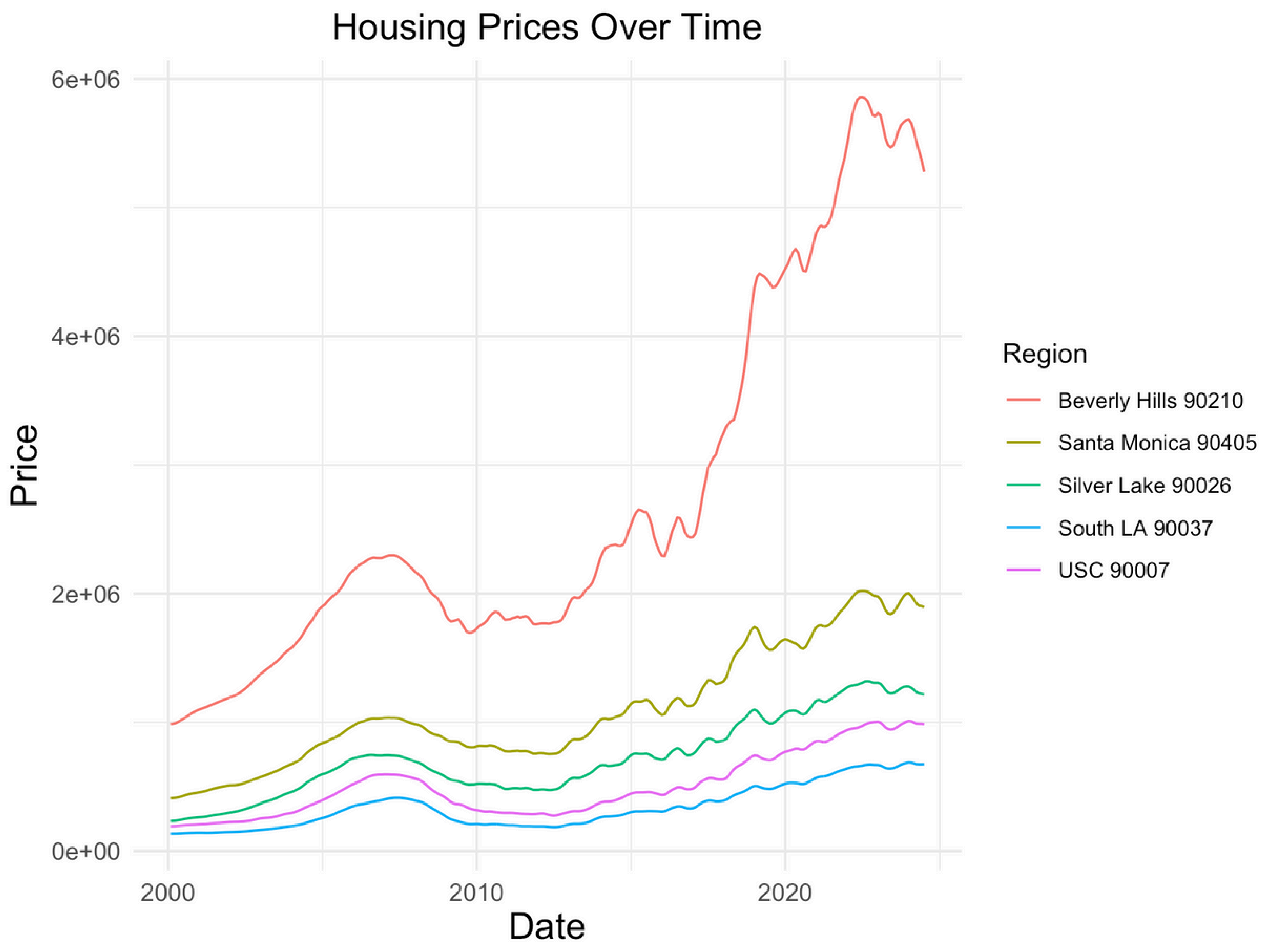


Figure 6: Neighborhood Home Prices

INTERRUPTED TIME SERIES

- Neighborhoods “Beverly Hills” and “South LA” are chosen to represent high and low economic status, respectively.

- We looked at two interruptions:
  - Great Financial Crisis (GFC)
    - September 1, 2007 - March 1, 2009

  - COVID-19 Pandemic
    - April 1, 2020 - April 1, 2022

INTERRUPTED TIME SERIES

Beverly Hills home prices <u>decreased</u> during the GFC and <u>increased</u> during the COVID-19 pandemic.

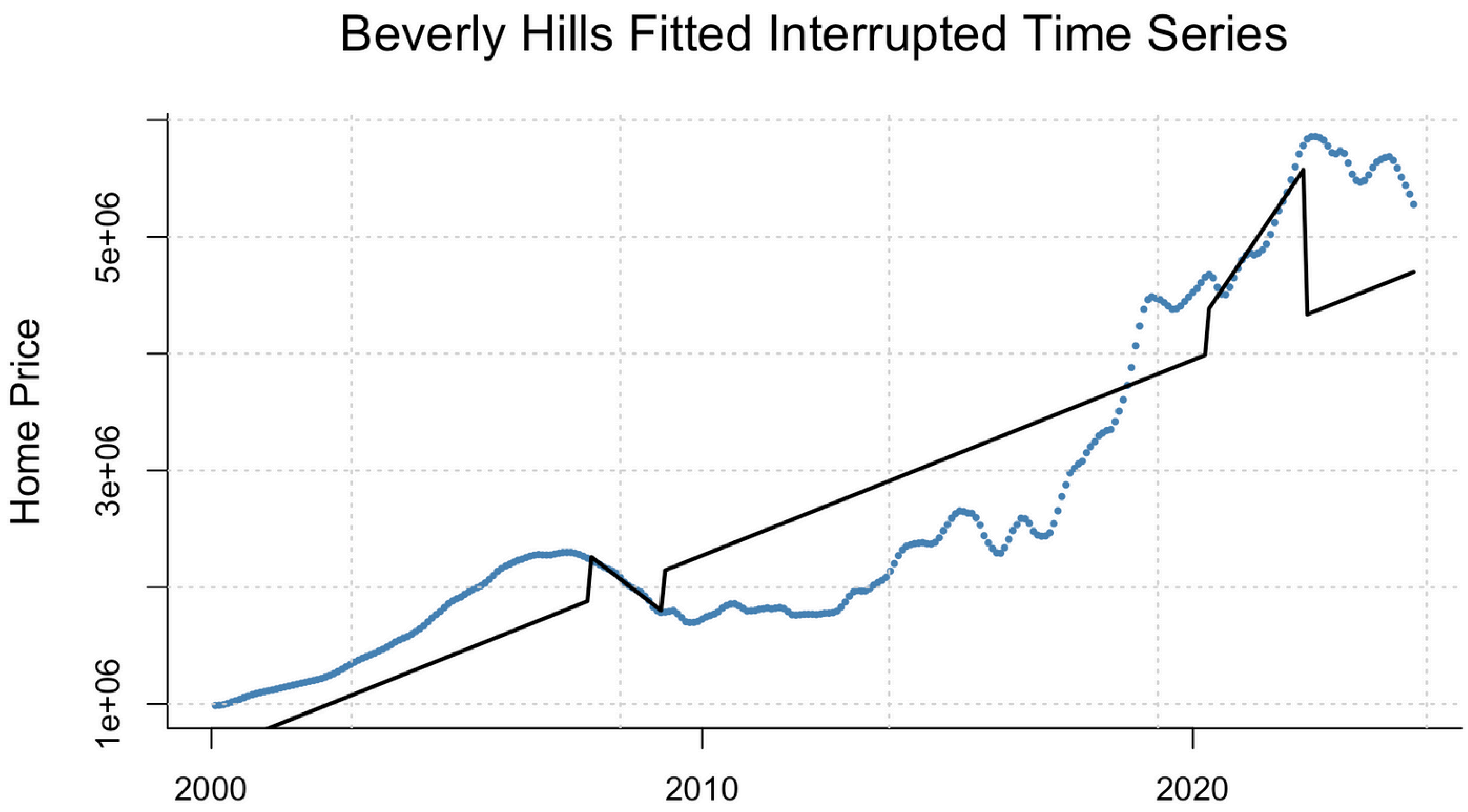


Figure 7: Interrupted Time Series - Beverly HIlls

INTERRUPTED TIME SERIES

South LA home prices <u>decreased</u> during the GFC and <u>increased</u> during the COVID-19 pandemic.

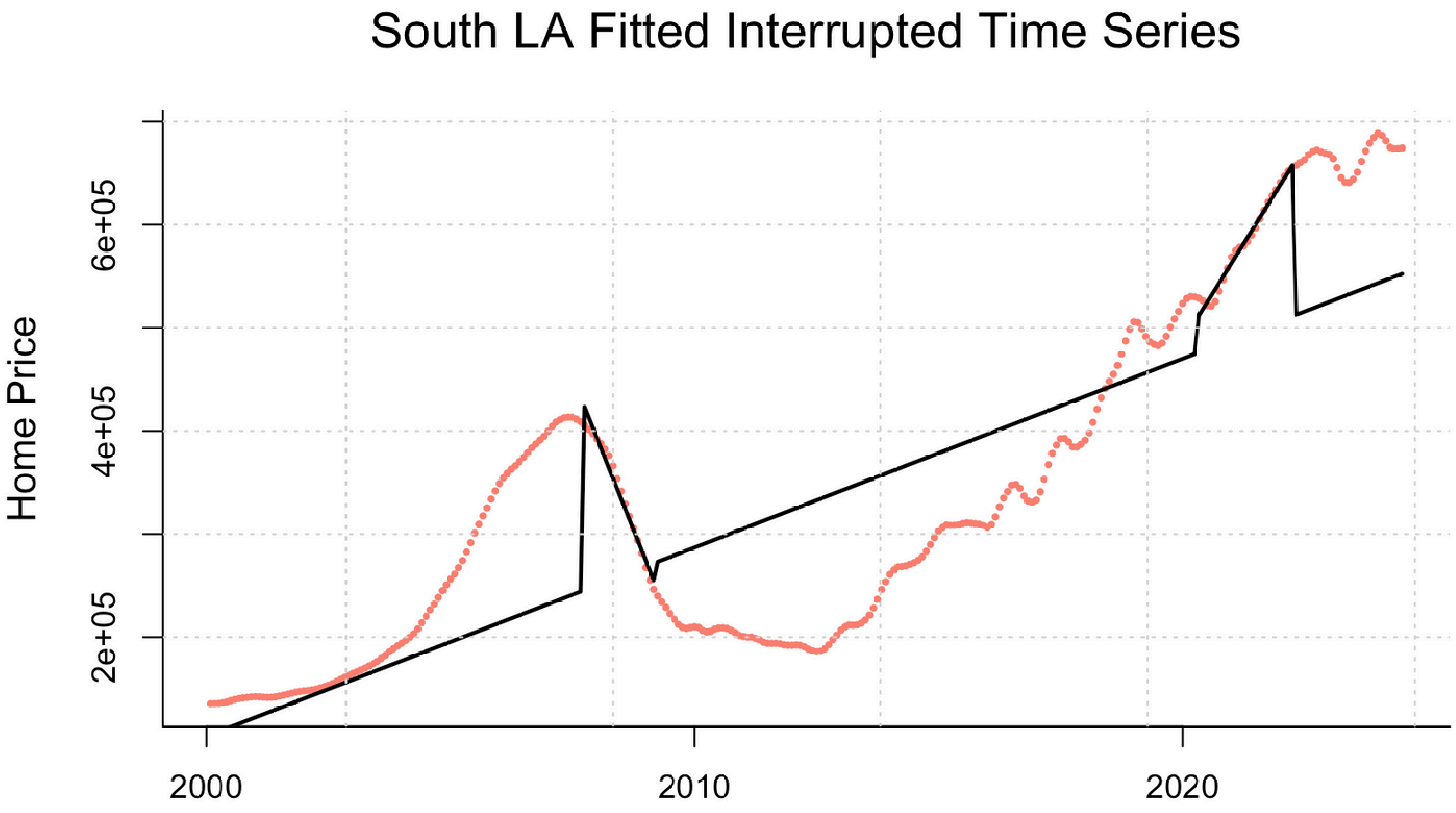


Figure 8: Interrupted Time Series - South LA

CHANGE POINT ANALYSIS

- Change Point Analysis was performed on the Interrupted Time Series Models (from previous slide).
- Historically, Beverly Hills has been more resilient to economic turmoil than South LA.

	Dependent Variable: Home Prices	
	Beverly Hills	South LA
(Intercept)	$-4,428,022.1^{***}$	$-445,372.8^{***}$
Date	458.6^{***}	50.2^{***}
GFC Indicator	$402,748.7$	$188,372.7^{***}$
COVID Indicator	$349,255.1$	$31,055.8$
GFC Indicator:Time Since GFC Start	$-1,340.5$	-375.2^{***}
COVID Indicator:Time Since COVID Start	$1,234.8^{**}$	158.1^{*}
Observations	294	294
R^2	0.813	0.724
Adjusted R^2	0.810	0.719
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Figure 9: Fitted Interrupted Time Series Modelsl

HEDONIC PRICING

- Data
  - We used the “housePrice” dataset from the “liver” R library.
- Data Cleaning
  - For variables with less than 100 missing values, we used the mode for imputing missing values. Otherwise, imputation was performed in a randomized manner.
- Method
  - We fit 80 Simple Linear Regression models and discovered that the three features that yield the highest R-squared values are: OverallQual, Neighborhood, and GrLivArea.
  - We filtered the dataset for five neighborhoods of interest.
    - We used their cumulative home prices to map them to neighborhoods in LA County: Beverly Hills, Santa Monica, Silver Lake, Santa Clarita, and South LA.

HEDONIC PRICING

- Final hedonic pricing model: Final_Hedonic_Model <- lm(SalePrice ~ OverallQual + Neighborhood + GrLivArea, data = Filtered_housePrices)
- This final model yields a R-squared of 84% and all predictors are statistically significant.
- The baseline neighborhood factor is ‘South LA’.

	Dependent variable:
	Home Price
(Intercept)	−149,913.4***
OverallQual	23,764.6***
NeighborhoodBeverly Hills	96,274.5***
NeighborhoodSanta Monica	69,482.7***
NeighborhoodSilver Lake	55,921.7***
NeighborhoodSanta Clarita	33,481.7***
GrLivArea	90.6***
Observations	330
R^2	0.837
Adjusted R^2	0.834
Residual Std. Error	43,529.320 (df = 323)
F Statistic	275.993*** (df = 6; 323)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 10: Hedonic Pricing Model

HEDONIC PRICING

- OverallQual: For every 1 unit increase in OverallQual, sale price increases by ~$23,750.
- Neighborhood: For example, sale prices for homes in Beverly Hills, on average, are ~$96,000 higher than the base (South LA).
- GrLivArea: For every additional square foot of GrLivArea, sale price increases by ~$90, on average.

	Dependent variable:
	Home Price
(Intercept)	−149,913.4***
OverallQual	23,764.6***
NeighborhoodBeverly Hills	96,274.5***
NeighborhoodSanta Monica	69,482.7***
NeighborhoodSilver Lake	55,921.7***
NeighborhoodSanta Clarita	33,481.7***
GrLivArea	90.6***
Observations	330
R^2	0.837
Adjusted R^2	0.834
Residual Std. Error	43,529.320 (df = 323)
F Statistic	275.993*** (df = 6; 323)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 10: Hedonic Pricing Model

REFERENCES & ACKNOWLEDGEMENTS

- [1] Reit.com

- Lectures and Homework from STAT 417 by Dr. Christopher Barr
- Data Providers:
  - Yahoo Finance
  - Zillow
  - “liver” R Package

Statistical Finance Final Codebook

Sean Mulherin

2024-07-10

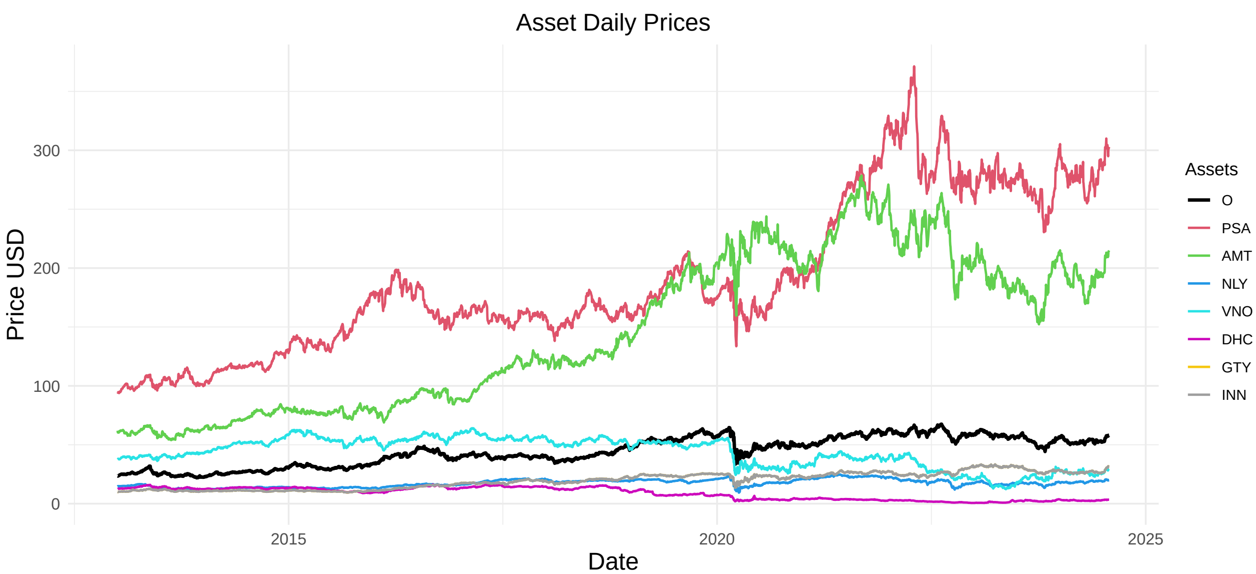
The objective of this study is to provide investment strategies for real estate investors based in Los Angeles (LA) County. To accomplish this, we employ modern statistical techniques to enhance the performance of financial investments both globally, through the investment of real estate investment trusts (REITs), and locally, through the investment of personal residential property. The topics explored through the REIT investment portfolio are diversification, optimization, and valuation using the CAPM model. The interrupted time series, change point analysis, and hedonic pricing analysis was performed within the context of home sale prices in Los Angeles County. The goal of this analysis is to better equip Los Angeles residents with information on how to succeed in their real estate investment pursuits.

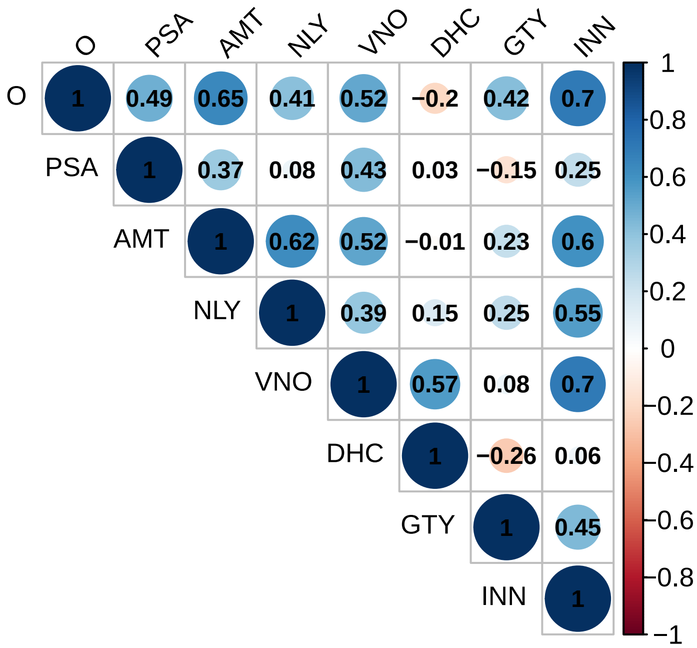
Diversification of Real Estate Portfolio

- O - walgreens, dollar tree, etc
- PSA - public storage
- AMT - cell/broadcast towers
- NLY - mortgages
- VNO - office buildings, retail stores in NYC
- DHC - health care properties
- GTY - gas stations
- INN - hospitality/lodging/resorts
- VNQ - Vanguard Real Estate ETF, acting as the REIT market

```
start_date = '2013-01-01'

tickers <- c( 'O', 'PSA', 'AMT', 'NLY', 'VNO', 'DHC', 'GTY', 'INN' )

df <- lapply( tickers,
              function(x) getSymbols( x, src = "yahoo", auto.assign = FALSE,
                                      from = start_date ) )

names( df ) <- tickers

data <- data.frame(
  O = df$O$O.Adjusted,
  PSA = df$PSA$PSA.Adjusted,
  AMT = df$AMT$AMT.Adjusted,
  NLY = df$NLY$NLY.Adjusted,
  VNO = df$VNO$VNO.Adjusted,
  DHC = df$DHC$DHC.Adjusted,
  GTY = df$GTY$GTY.Adjusted,
  INN = df$INN$INN.Adjusted
```

```
)

market_return <- getSymbols( "VNQ", src= "yahoo", auto.assign= FALSE,
                              from=start_date )

market_return_daily <- lapply( market_return$VNQ.Adjusted, dailyReturn ) |>
                       as.data.frame()

market_return_annual <- lapply( market_return$VNQ.Adjusted, annualReturn ) |>
                        as.data.frame()

market_return_avg <- mean( market_return_annual$yearly.returns )

data <- xts( data, order.by = as.Date( rownames( data ) ) )

annual_returns <- lapply( data, annualReturn)  |> as.data.frame()

names( annual_returns ) <- tickers

daily_returns <- lapply( data, dailyReturn )  |> as.data.frame()

names( daily_returns ) <- tickers
```



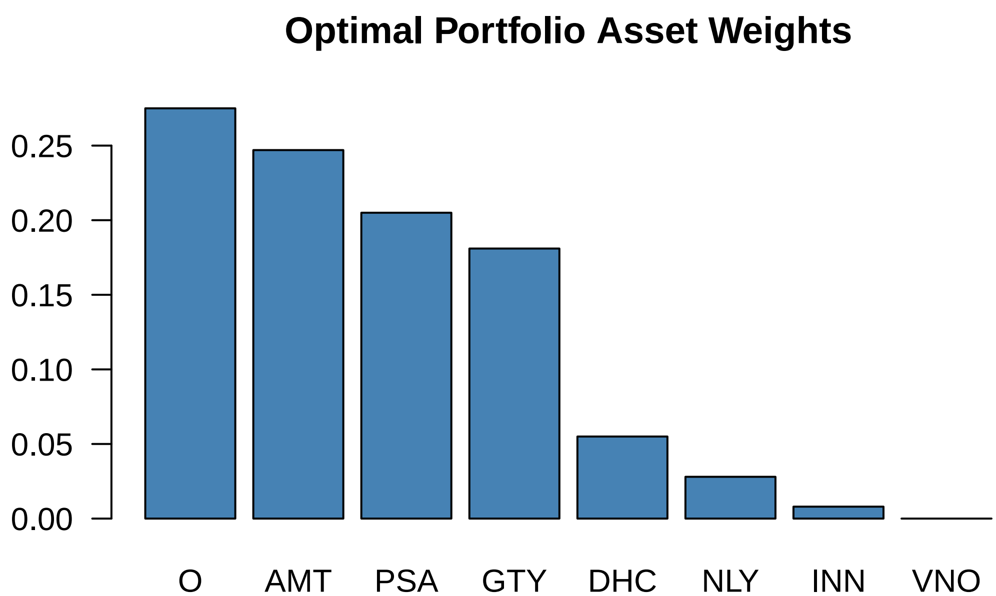


Portfolio Optimization and Valuation

- Efficiency Frontier
- Expected Return/Std

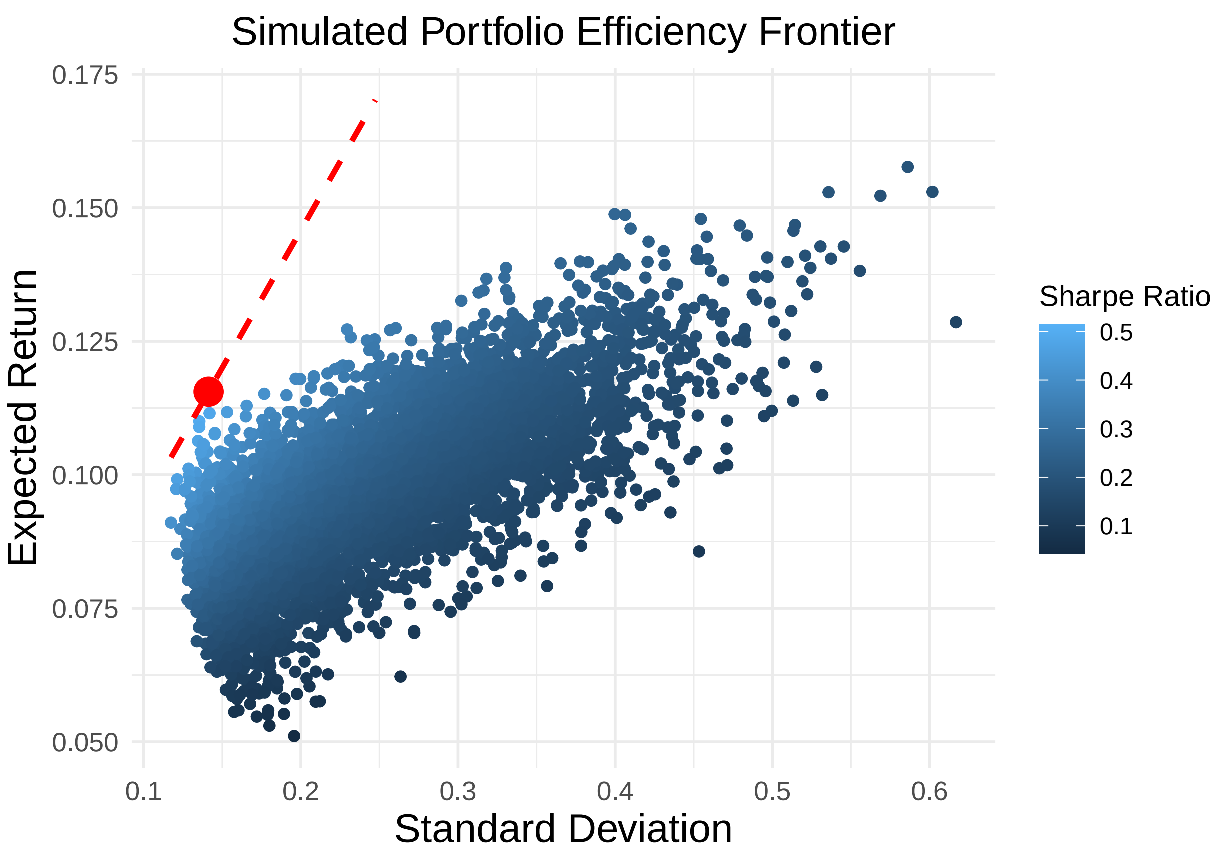
```r
n_assets <- ncol( annual_returns )

n_sims <- 5000

weights <- matrix( NA, nrow = n_sims, ncol = n_assets )

p_std <- array( dim = n_sims )

p_expReturn <- array( dim = n_sims )

for( i in 1:n_sims ) {
  weights[i, ] <- runif( n_assets )

  weights[i, ] <- weights[i, ] / sum( weights[i, ] )

  p_std[i] <- ( t( weights[i, ] ) %*% cov( annual_returns ) %*% weights[i, ] ) |> sqrt()

  p_expReturn[i] <- weights[i, ] %*% colMeans( annual_returns )
}

portfolio_data <- data.frame( p_expReturn, p_std )

risk_free_rate_10yr <- 0.0428

sharpe_ratios <- ( p_expReturn - risk_free_rate_10yr ) / p_std
```

```r
max_sharpe_ratio <- max( sharpe_ratios )

optimal_return <- p_expReturn[ which.max( sharpe_ratios ) ]

optimal_std <- p_std[ which.max( sharpe_ratios ) ]

optimal_weights <- weights[ which.max( sharpe_ratios ), ]

optimal_weights <- round( optimal_weights, 3 )

optimal_portfolio <- data.frame( Tickers = tickers,
                                 Weights = optimal_weights )
```




```
## [1] "Optimal (max) sharpe ratio = 0.515"
## [1] "Optimal expected return = 0.116"
## [1] "Optimal standard deviation = 0.141"
```

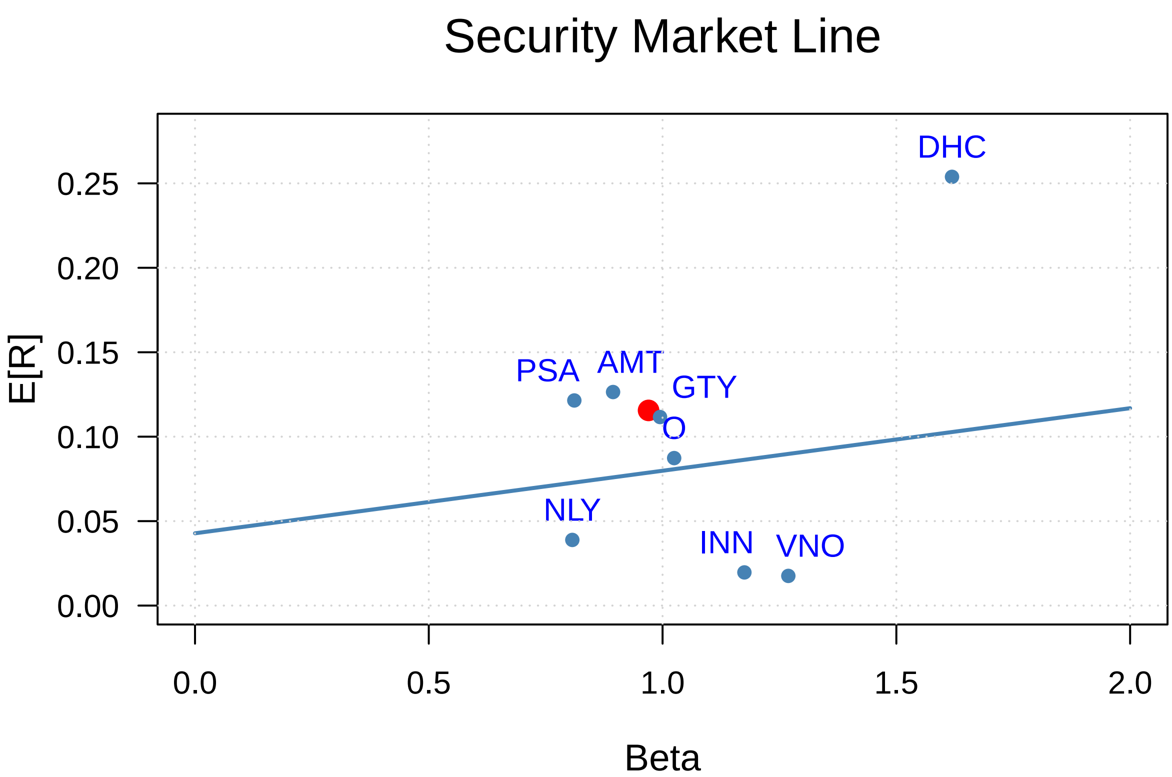
```r
cml_x <- seq( min( p_std ), mean( p_std ), 0.01 )

cml_y <- risk_free_rate_10yr + cml_x * ( optimal_return - risk_free_rate_10yr ) / optimal_std

cml <- data.frame( x = cml_x,
                   y = cml_y )
```



Portfolio Valuation

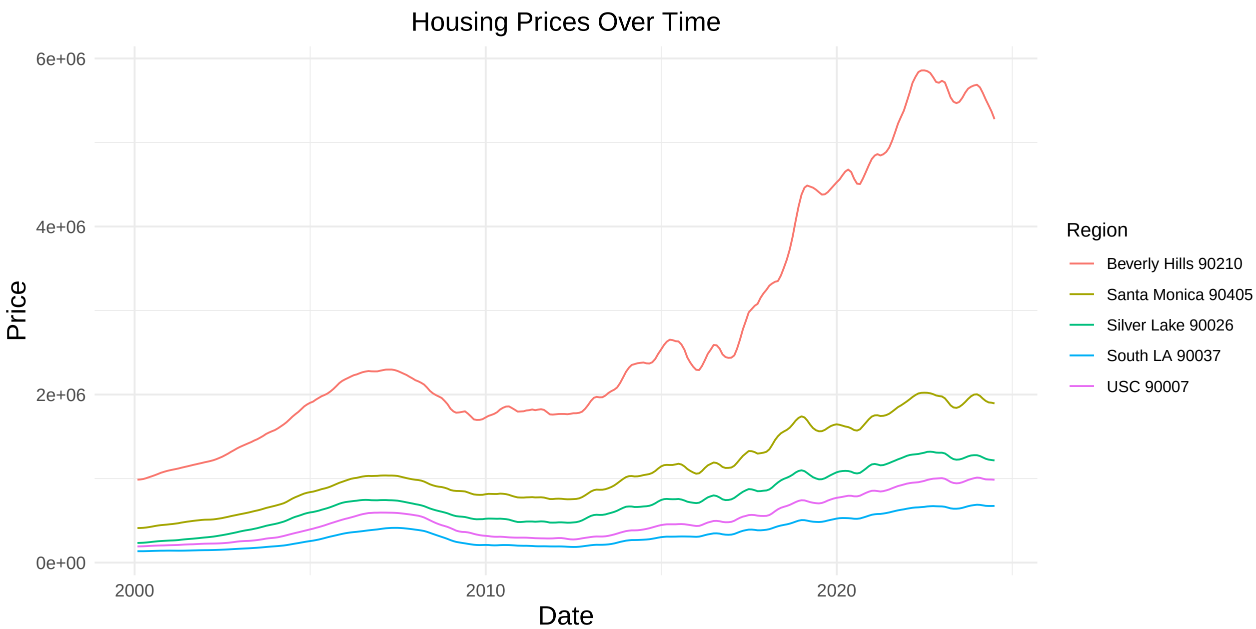
- Capital Asset Pricing Model
- Security Market Line

```
daily_returns <- daily_returns[ , 1:ncol( daily_returns ) - 1 ]

beta_i <- cov( daily_returns, market_return_daily ) / var( market_return_daily$daily.returns )

returns_i <- apply( annual_returns, 2, mean )

portfolio_beta <- sum( beta_i * optimal_weights )

beta_x <- seq( 0, 2, 0.01 )

SML <- risk_free_rate_10yr + beta_x * ( market_return_avg - risk_free_rate_10yr )
```

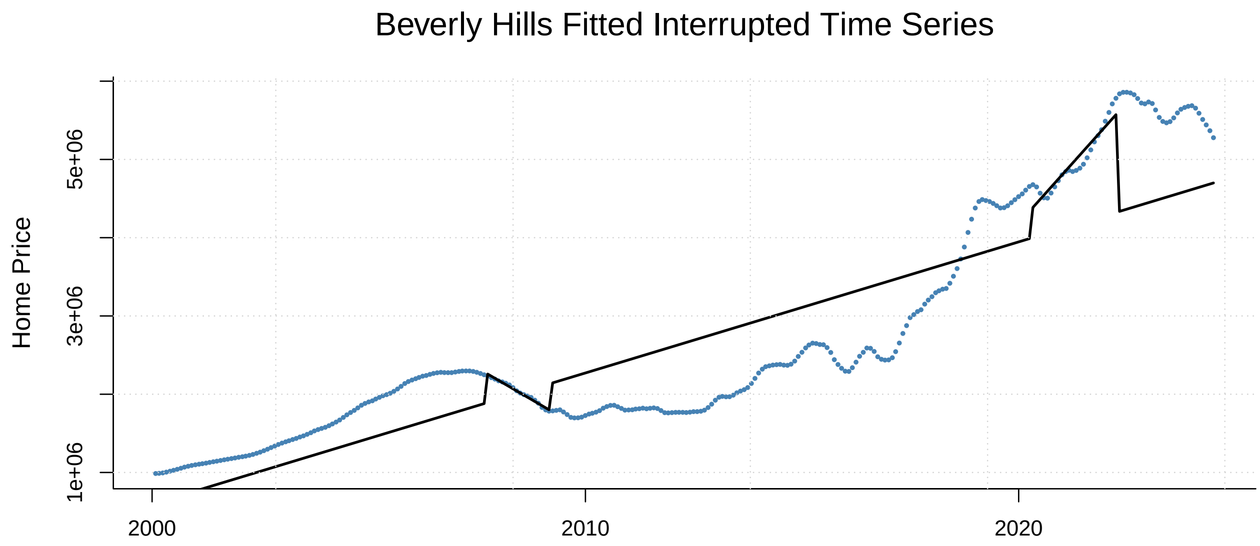


Our portfolio (red) is undervalued since it is above the security market line. In other words, we believe the expected returns are higher than the market's valuation.

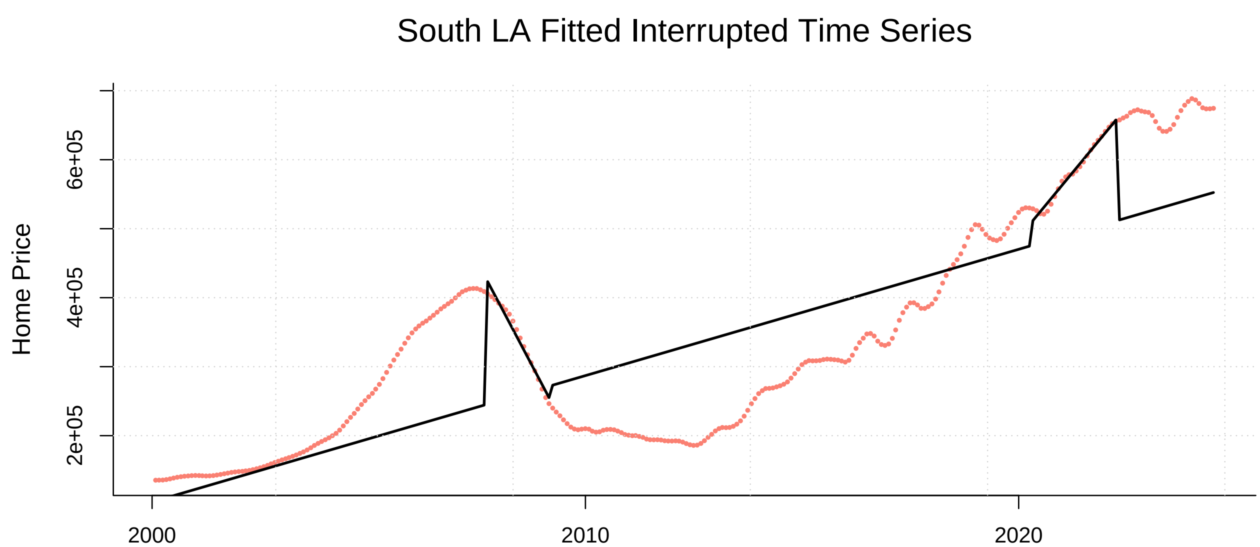
Interrupted Time Series



```r
dates <- colnames( zillow_df )[ 2:n ]

dates <- as.Date( gsub( "X", "", dates ), format = "%Y.%m.%d" )

housing_crisis_start = as.Date( "2007-09-01" )

housing_crisis_end = as.Date( "2009-03-01" )

covid_start = as.Date( "2020-04-01" )

covid_end = as.Date( "2022-04-01" )

Housing_Crisis_I <- 1 * ( dates > housing_crisis_start & dates < housing_crisis_end )

COVID_I = 1 * ( dates > covid_start & dates < covid_end )

BeverlyHills <- zillow_df[ zillow_df$Region == 'Beverly Hills 90210', 2:n ] |>
                as.numeric()

South_LA <- zillow_df[ zillow_df$Region == 'South LA 90037', 2:n ] |>
  as.numeric()

interrupted_ts_data <- data.frame(
  'Date' = dates,
  'BeverlyHills' = BeverlyHills,
  'South_LA' = South_LA,
  'GFC_I' = Housing_Crisis_I,
  'COVID_I' = COVID_I,
  'Time_Since_GFC' = dates - housing_crisis_start,
  'Time_Since_COVID' = dates - covid_start
)

BeverlyHills_Model <- lm(
  BeverlyHills ~ Date +
  GFC_I * Time_Since_GFC +
  COVID_I * Time_Since_COVID,
  data = interrupted_ts_data
)

South_LA_Model <- lm(
  South_LA ~ Date +
  GFC_I * Time_Since_GFC +
  COVID_I * Time_Since_COVID,
  data = interrupted_ts_data
)

Pred_BeverlyHills <- predict( BeverlyHills_Model )

Pred_South_LA <- predict( South_LA_Model )
```




```
##                               Estimate   Std. Error   t value     Pr(>|t|)
## (Intercept)              -4428021.8366 237264.59080 -18.662801 1.685306e-51
## Date                          458.5671     15.36083  29.853023 3.766648e-90
## GFC_I                      402748.6944 309539.41766   1.301122 1.942565e-01
## COVID_I                    349255.1471 270076.16152   1.293173 1.969876e-01
## GFC_I:Time_Since_GFC        -1340.4975    931.18213  -1.439565 1.510764e-01
## COVID_I:Time_Since_COVID     1234.8399    605.59179   2.039063 4.235654e-02
```




```
##                               Estimate   Std. Error    t value     Pr(>|t|)
## (Intercept)              -445372.77360 33280.392026 -13.3824377 4.330521e-32
## Date                          50.12947     2.154617  23.2660703 4.178220e-68
## GFC_I                     188372.70172 43418.165063   4.3385689 1.986547e-05
## COVID_I                    31055.77053 37882.772569   0.8197861 4.130160e-01
## GFC_I:Time_Since_GFC        -375.15898   130.614122  -2.8722697 4.378187e-03
```

```
## COVID_I:Time_Since_COVID     158.04461    84.944543    1.8605623 6.382511e-02
```

Hedonic Pricing

```r
Hedonic_Model <- lm( SalePrice ~ ., data = housePrice )

summary( Hedonic_Model )$r.squared
```

```
## [1] 0.9363984
```

```r
salePrice <- housePrice$SalePrice

features <- housePrice[, -which( names( housePrice ) == "SalePrice" ) ]

num_features <- ncol( features )

r_sqr <- numeric( ncol( features ) )

for( i in 1:num_features ){

  fit <- lm( SalePrice ~ features[ , i], data = housePrice )

  r_sqr[i] <- summary( fit )$r.squared
}
```

```r
Final_Hedonic_Model <- lm(
  SalePrice ~ OverallQual +
    Neighborhood + GrLivArea,
  data = Filtered_House_Prices
)

summary( Final_Hedonic_Model )$coefficients
```

```
##                                Estimate   Std. Error   t value     Pr(>|t|)
## (Intercept)                -149913.44030 15688.624638 -9.555550 3.230005e-19
## OverallQual                  23764.63798  2955.189723  8.041662 1.695565e-14
## NeighborhoodBeverly Hills    96274.51698 12760.907863  7.544488 4.655780e-13
## NeighborhoodSanta Monica     69482.66682 12515.089440  5.551911 5.902588e-08
## NeighborhoodSilver Lake      55921.73691 10699.630701  5.226511 3.106574e-07
## NeighborhoodSanta Clarita    33481.66952 11808.411791  2.835408 4.865246e-03
## GrLivArea                       90.61822     6.596423 13.737480 3.881988e-34
```