Risk and Portfolio Management Spring 2011

PCA, Dynamic PCA

& Applications to Risk-Management

Basic investment equation

$E_t =$ equity in a trading account at time t (liquidation value)

$R_{it} =$ return on stock i from time t to time $t+\Delta t$ (includes dividend income)

$Q_{it} =$ dollars invested in stock i at time t

$r =$ interest rate

$$E_{t+\Delta t} = E_t + E_t r\Delta t + \sum_{i=1}^{N} Q_{it}R_{it} - \left(\sum_{i=1}^{N} Q_{it}\right) r\Delta t \qquad \left(\text{before rebalancing, at time } t+\Delta t\right)$$

$$E_{t+\Delta t} = E_t + E_t r\Delta t + \sum_{i=1}^{N} Q_{it}R_{it} - \left(\sum_{i=1}^{N} Q_{it}\right) r\Delta t + \varepsilon \sum_{i=1}^{N} | Q_{i(t+\Delta t)} - Q_{it} | \qquad \left(\text{after rebalancing, at time } t+\Delta t\right)$$

$\varepsilon =$ transaction cost (as percentage of stock price)

Leverage

$$E_{t+\Delta t} = E_t + E_t r\Delta t + \sum_{i=1}^{N} Q_{it} R_{it} - \left(\sum_{i=1}^{N} Q_{it} \right) r\Delta t$$

$$\boxed{\text{Leverage} = \frac{\sum_{i=1}^{N} |Q_{it}|}{E_t}}$$

Ratio of (gross) investments to equity

$Q_{it} \geq 0$ ``Long-only position"

$Q_{it} \geq 0,\ \sum_{i=1}^{N} Q_{it} = E_t$ Leverage $= 1$, long only position

Reg-T: Leverage ≤ 2 (margin accounts for retail investors)

Day traders: Leverage ≤ 4

Professionals & institutions: Risk-based leverage

Portfolio Theory

Introduce dimensionless quantities and view returns as <u>random variables</u>

$$\theta_i = \frac{Q_i}{E_i} \qquad \text{Leverage} = \sum_{\mathrm{i}=1}^{N} |\theta_i|$$

Dimensionless ``portfolio weights''

$$\frac{\Delta\Pi}{\Pi} = \frac{E_{t+\Delta t} - E_t - E_t r\Delta t}{E_t} = \frac{\Delta E}{E} - r\Delta t$$

$$\tilde{R}_i = R_i - r\Delta t$$

All investments financed (at known IR)

$$\boxed{\frac{\Delta\Pi}{\Pi} = \sum_{i=1}^{N} \theta_i \tilde{R}_i}$$

$$E\left(\frac{\Delta\Pi}{\Pi}\right) = \sum_{i=1}^{N} \theta_i E\left(\tilde{R}_i\right); \qquad \sigma^2\left(\frac{\Delta\Pi}{\Pi}\right) = \sum_{ij=1}^{N} \theta_i\theta_j Cov\left(\tilde{R}_i, \tilde{R}_j\right) = \sum_{ij=1}^{N} \theta_i\theta_j \sigma_i\sigma_j\rho_{ij}$$

Sharpe Ratio

$$s = s\left(\theta_1,\ldots,\theta_N\right) = \frac{E\left(\dfrac{\Delta\Pi}{\Pi}\right)}{\sigma\left(\dfrac{\Delta\Pi}{\Pi}\right)} = \frac{\sum_{i=1}^{N}\theta_i E\left(\tilde{R}_i\right)}{\sqrt{\sum_{i=1}^{N}\theta_i\theta_j\sigma_i\sigma_j\rho_{ij}}}$$

Sharpe ratio is homogeneous of degree zero in the portfolio weights. Hence, it is independent of the leverage.

Investors who prefer large returns & low volatility may use the Sharpe Ratio as a metric to evaluate investment portfolios.

Quoting a hedge fund manager (in 2006): ``I don't eat volatility''.
(HF managers are compensated as % of returns, not Sharpe ratio ☺.)

Arbitrage Pricing Theory (Ross, 1971)

Assumptions:

- There are N stocks
- m <u>tradable assets</u> as factors (e.g. ETFs or baskets, or portfolios)
- Linear regression of stock returns on factors has <u>uncorrelated residuals</u> (uncorrelated with the factors and among each other)

$$\tilde{R}_i = \alpha_i + \sum_{k=1}^{m} \beta_{ik} \tilde{F}_k + \varepsilon_i$$

$\tilde{R}_i = R_i - r\Delta t =$ return on stock i, financed

$\tilde{F}_k = F_k - r\Delta t =$ return on factor k, financed

$\varepsilon_i =$ residual of the linear regression of stock ret. on factor returns

$\alpha_i =$ excess returns

Market-neutral portfolios

Consider the following investment portfolio:

$$\begin{cases} Q_i & \text{dollars in stock } i \\ -\sum_{i=1}^{N} Q_i \beta_{ik} & \text{dollars in asset } k \end{cases}$$

$$\frac{\Delta \Pi}{\Pi} = \sum_{i=1}^{N} Q_i \tilde{R}_i - \sum_{k=1}^{m} \left(\sum_{i=1}^{N} Q_i \beta_{ik} \right) \tilde{F}_k$$

$$= \frac{1}{\Pi} \sum_{i=1}^{N} \theta_i \left(\tilde{R}_i - \sum_{k=1}^{M} \beta_{ik} \tilde{F}_k \right), \quad \theta_{\mathrm{i}} = \frac{Q_i}{\Pi}$$

$$= \sum_{\mathrm{i}=1}^{N} \theta_i \left(\alpha_i + \varepsilon_i \right) = \sum_{\mathrm{i}=1}^{N} \theta_i \alpha_i + \sum_{\mathrm{i}=1}^{N} \theta_i \varepsilon_i$$

Market neutral portfolios are only exposed to residuals

Diversified Market-neutral portfolio

$$E\left(\frac{\Delta\Pi}{\Pi}\right)=\sum_{i=1}^{N}\theta_i\alpha_i \qquad \sigma^2\left(\frac{\Delta\Pi}{\Pi}\right)=\sum_{i=1}^{N}\theta_i^2\sigma^2(\varepsilon_i)$$

$$\boxed{\theta_i=\frac{sign(\alpha_i)}{N}}$$

Choose weights of opposite to sign of excess return & uniformly distributed

$$E\left(\frac{\Delta\Pi}{\Pi}\right)=\frac{1}{N}\sum_{i=1}^{N}|\alpha_i|, \qquad \sigma^2\left(\frac{\Delta\Pi}{\Pi}\right)=\frac{1}{N^2}\sum_{i=1}^{N}\sigma^2(\varepsilon_i)$$

$$\boxed{s=\frac{\frac{1}{N}\sum_{i=1}^{N}|\alpha_i|}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sigma^2(\varepsilon_i)}}\approx\sqrt{N}\cdot\frac{\int_{-\infty}^{\infty}|\alpha|\,dF(\alpha)}{\sqrt{\int_{0}^{\infty}\sigma^2 dG(\sigma^2)}}=\sqrt{N}\frac{\langle|\alpha|\rangle}{\sqrt{\langle\sigma^2\rangle}}}$$

No-arbitrage argument (S. Ross, 1971/1974)

If a significant fraction of the excess returns are non-zero, i.e. if $\langle|\alpha|\rangle > 0$ investing in the diversified MN portfolio would rise to an infinite Sharpe ratio as *N* tends to infinity (or, for large enough *N*, a huge Sharpe ratio).

This would represent a clear arbitrage opportunity which would be eventually eliminated by trading.

Conclusion, if the ``factor model'' holds true with tradable factors and uncorrelated residuals then excess returns should vanish and stocks returns should satisfy the normative model

$$\tilde{R}_i = \sum_{k=1}^{m} \beta_{ik} \tilde{F}_k + \varepsilon_i \qquad \text{with} \qquad E(\varepsilon_i) = 0$$

Excess returns are zero

$$E\left(\tilde{R}_i\right) = \sum_{k=1}^{m} \beta_{ik} E\left(\tilde{F}_k\right)$$

This can be seen as an ``equilibrium'' pricing relation of sorts.

Capital Asset Pricing Model (Lintner, Sharpe, 1965)

The capital asset-pricing model is a one-factor model to explain stock returns and stock prices.

$$\boxed{\tilde{R}_i = \beta_i \tilde{F}_1 + \varepsilon_i, \qquad \langle \varepsilon_i \rangle = 0,\ \langle \varepsilon_i \varepsilon_j \rangle = 0,\ \text{if } i \neq j}$$

$$\tilde{F}_1 = \ \text{return of the market portfolio (cap - weighted market portfolio)}$$

$$E\left(\tilde{R}_i\right) = \beta_i E\left(\tilde{F}_1\right), \qquad \beta_i = \frac{Cov(R_i, F_1)}{Var(F_M)} = \frac{\sigma_i \rho_i}{\sigma_M}$$

Remarks & comments

APT introduces a probability measure on the set of returns, which is not specified explicitly. The existence of such probability would imply stationarity of returns, which in practice does not hold.

Technicality: the diversified portfolio constructed here has finite leverage as N tends to infinity. This follows from the fact that there are finitely many factors and that the leverage on the ``stocks'' before hedging with the factors is 1.

APT, as presented here, makes no assumptions about dynamics, i.e. portfolios that vary in time, or about market-timing strategies, holding periods, etc.

The theory was made before the proliferation of ETFs, which are tradable factors.

ETFs provide an important set of factors (sectors, size, etc) that can be used to make APT ``concrete'' and applicable.

Testing APT with data from Jan 5, 2009 to Jan 29 2010

Using daily returns from Jan 5 2009 to Jan 29, 2010 for the components of the S&P 500 index, we explored the ``partition'' of the eigenvectors/eigenvalues into significant and noise components.

We did this by testing for $\alpha = 0$ and for the correlations of residuals.

One way to analyze the correlations of residuals is by doing a PCA again and analyzing the corresponding eigenvalues and DOS.

The assumption that residuals are uncorrelated allows comparison with theoretical eigenvalue distributions such as Marcenko-Pastur.

One Factor Model (CAPM)

1. Compute correlation matrix of S&P 500

2. Compute the first eigenportfolio

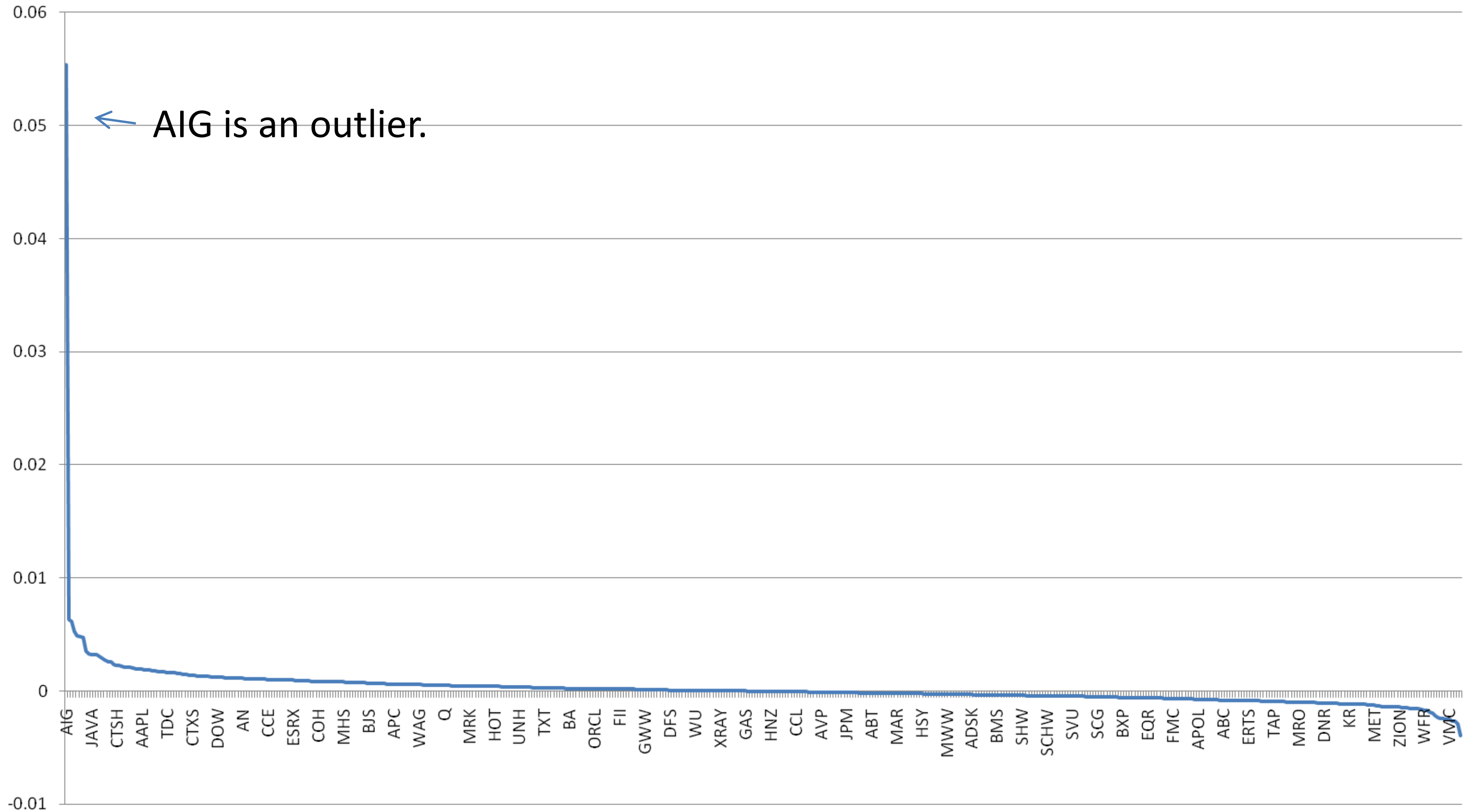
3. Compute residuals for all 500 stocks by regression

$$\tilde{R}_i = \alpha_i + \beta_i \tilde{F}_1 + \varepsilon_i$$

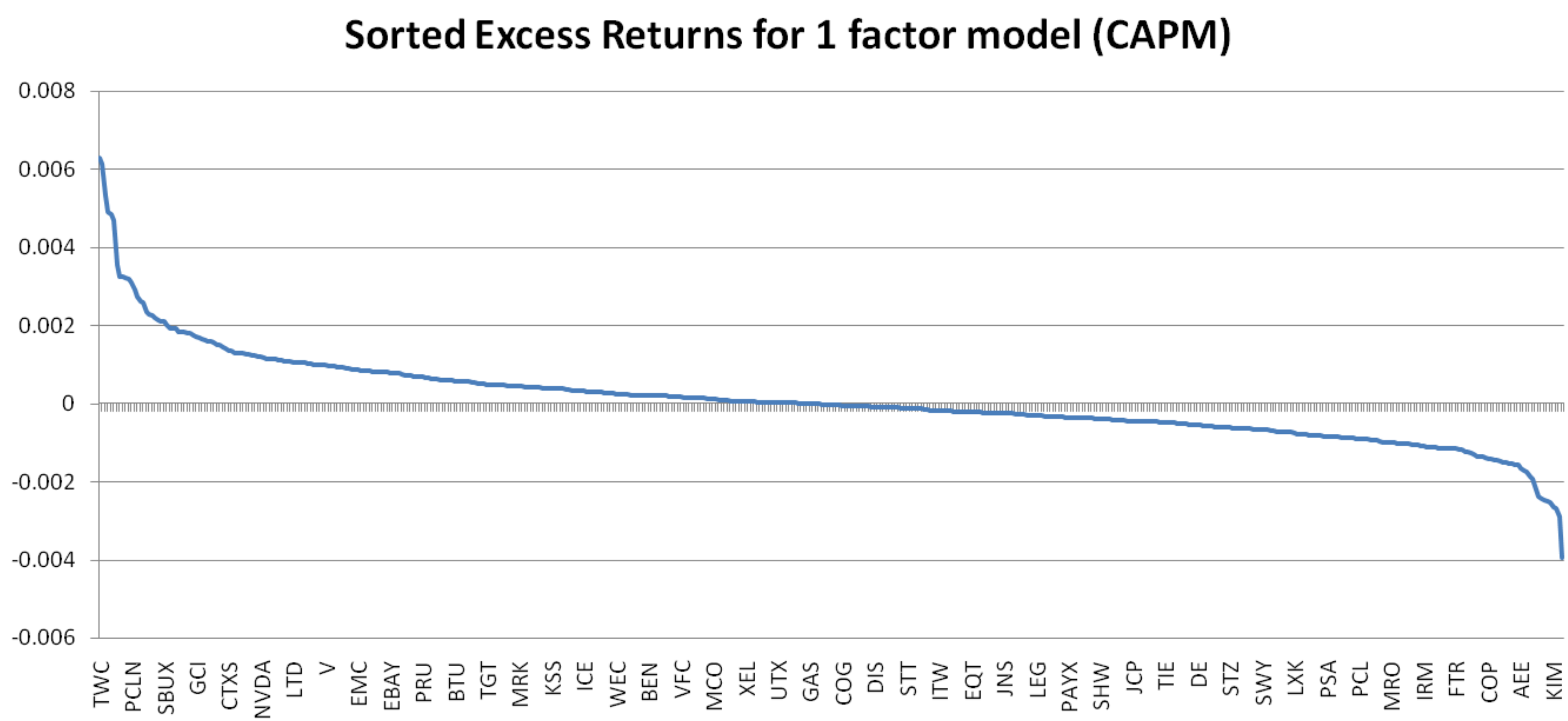
4. Analyze the vector of alphas

5. Analyze the correlation matrix of the residuals

Sorted Excess Returns, 1-factor

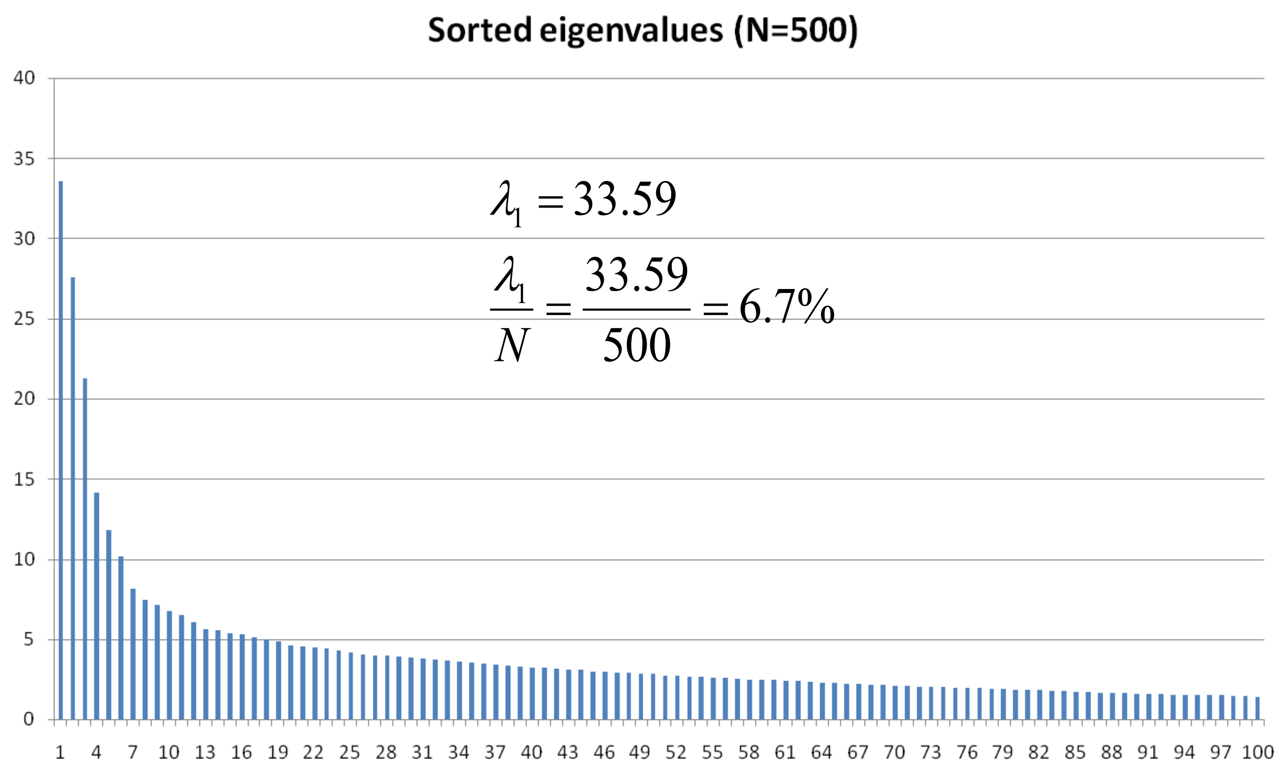


Sorted Excess Returns 1-factor, excluding AIG



Max=60 pbs, Min=-40 bps, average=1.9 bps, stdev=11bps

Eigenvalues of the correlation matrix of residuals (m=1)



$$\lambda_1 = 33.59$$

$$\frac{\lambda_1}{N} = \frac{33.59}{500} = 6.7\%$$

Recall that lambda_1 for the original correlation matrix was ~ 220, so the residuals matrix has ``smaller'' correlations.

Ratio lambda/N is a proxy for the average correlation.

First eigenvalue & average correlation

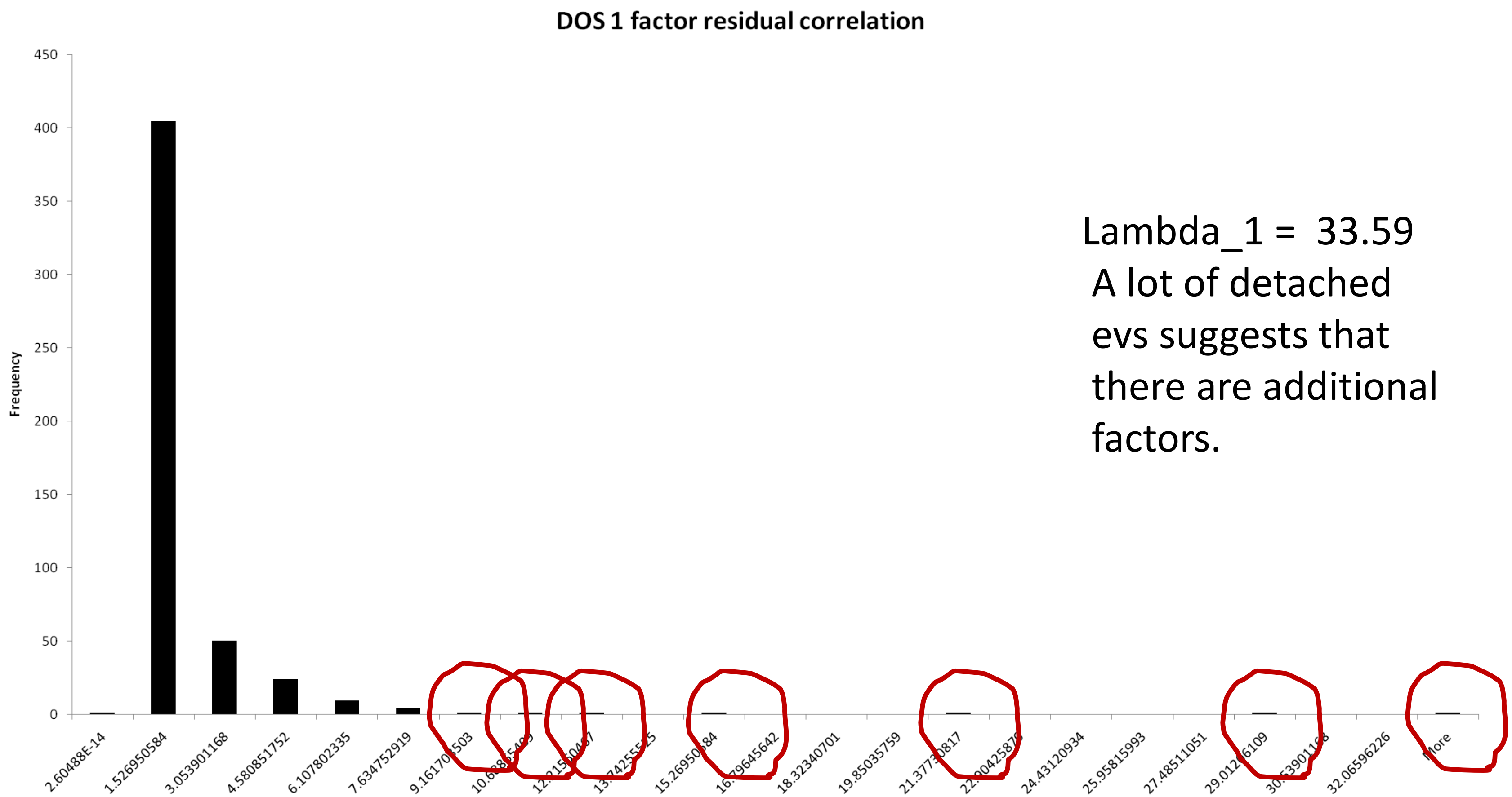
$$\begin{aligned}
\lambda_1 &= V^{(1)T} C V^{(1)} \\
&= \sum_{i=1}^{N} \left(V_i^{(1)}\right)^2 + \sum_{i \neq j} V_i^{(1)} V_j^{(1)} \rho_{ij} \\
&= 1 + \sum_{i \neq j} V_i^{(1)} V_j^{(1)} \rho_{ij} \\
&= 1 + \left(\sum_{i \neq j} V_i^{(1)} V_j^{(1)} \right) \cdot \frac{\sum_{i \neq j} V_i^{(1)} V_j^{(1)} \rho_{ij}}{\sum_{i \neq j} V_i^{(1)} V_j^{(1)}}
\end{aligned}$$

$$\frac{\lambda_1 - 1}{\sum_{i \neq j} V_i^{(1)} V_j^{(1)}} = \frac{\sum_{i \neq j} V_i^{(1)} V_j^{(1)} \rho_{ij}}{\sum_{i \neq j} V_i^{(1)} V_j^{(1)}} \qquad \therefore \quad V_i^{(1)} \approx \frac{1}{\sqrt{N}}, \quad \sum_{i \neq j} V_i^{(1)} V_j^{(1)} \approx \frac{N(N-1)}{N} = N - 1$$

$$\frac{\lambda_1 - 1}{N - 1} \approx \langle \rho \rangle$$

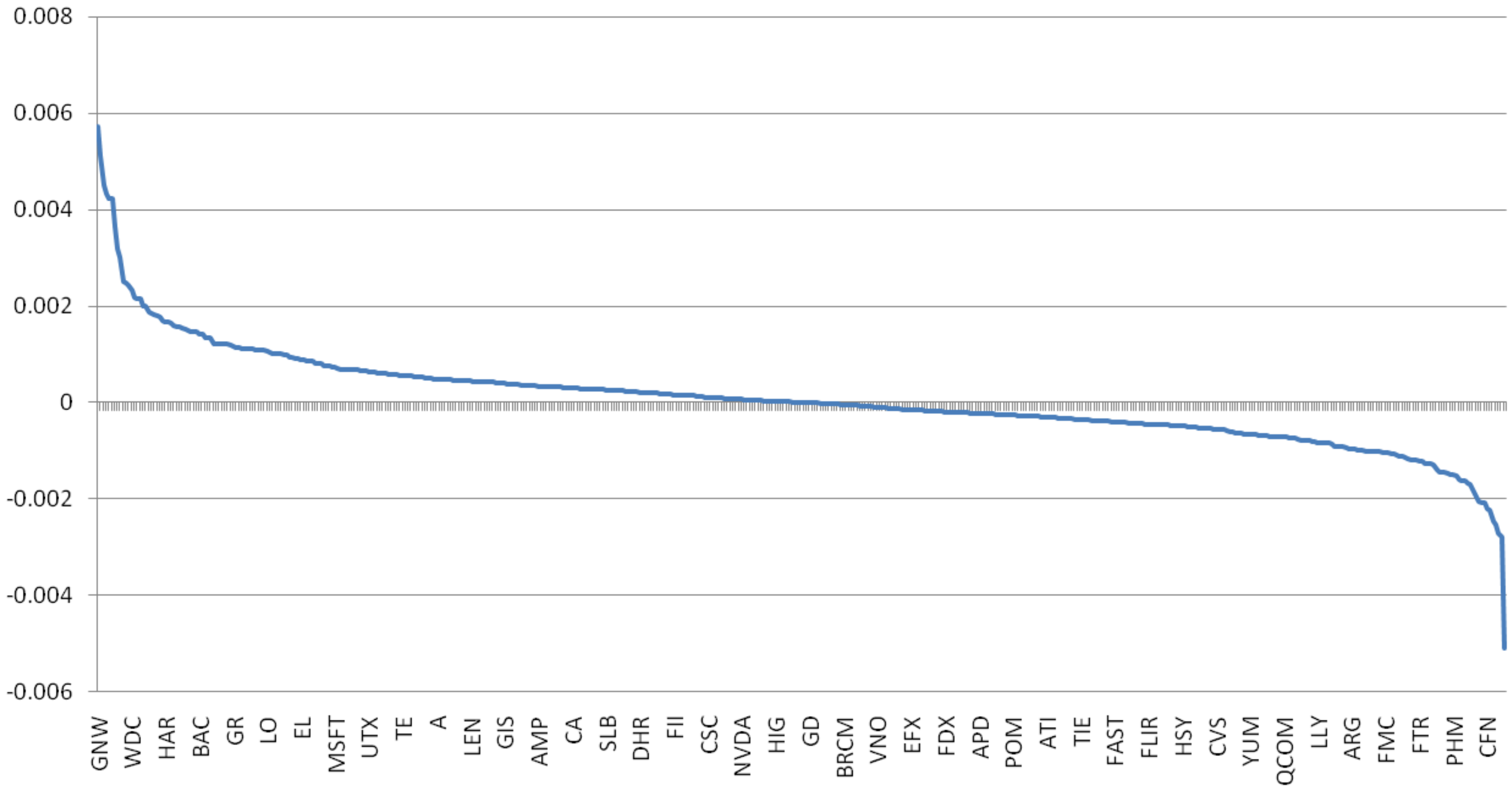
$$\boxed{\langle \rho \rangle \approx \frac{\lambda_1}{N}}$$

Density of States, or Histogram, of Eigenvalues

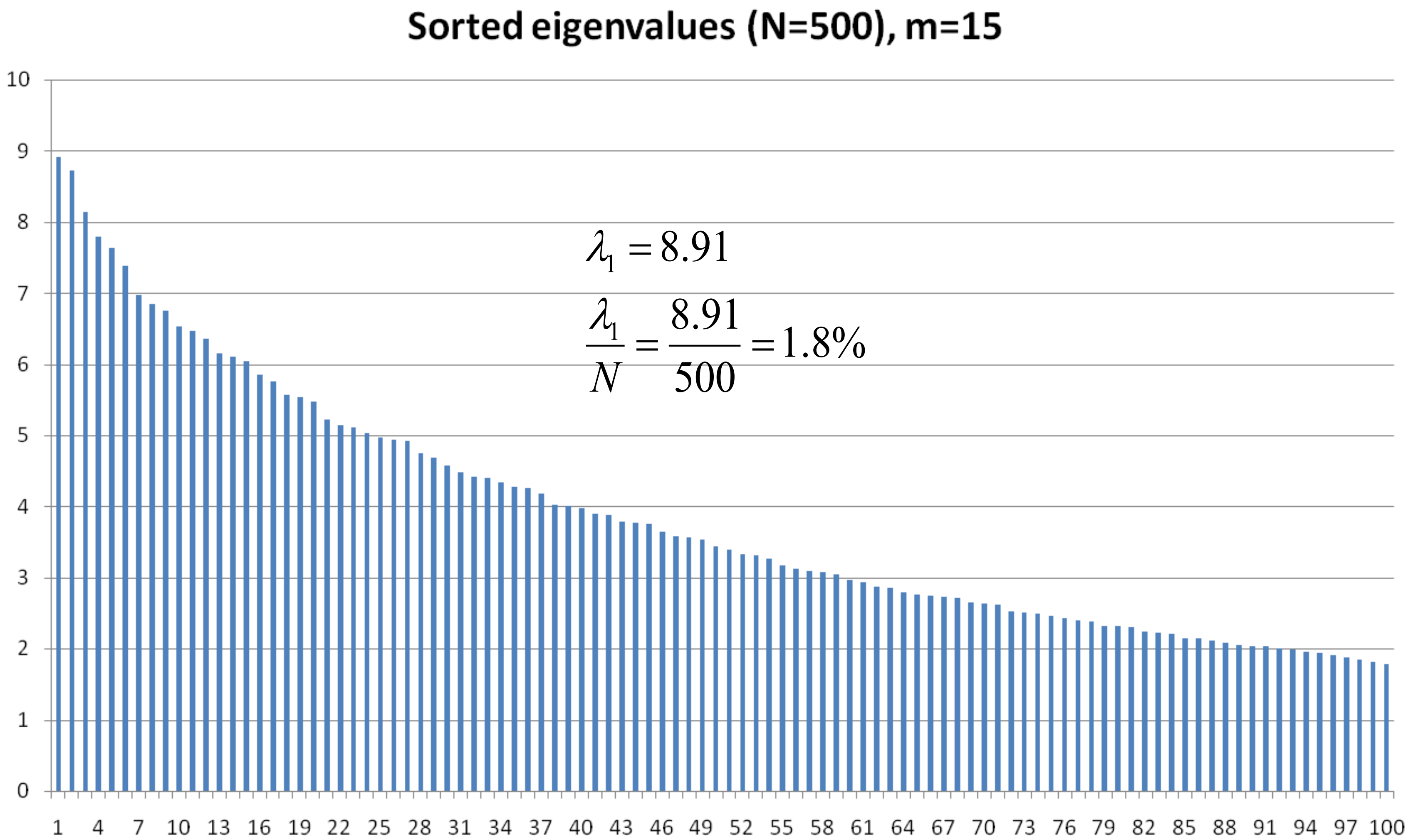


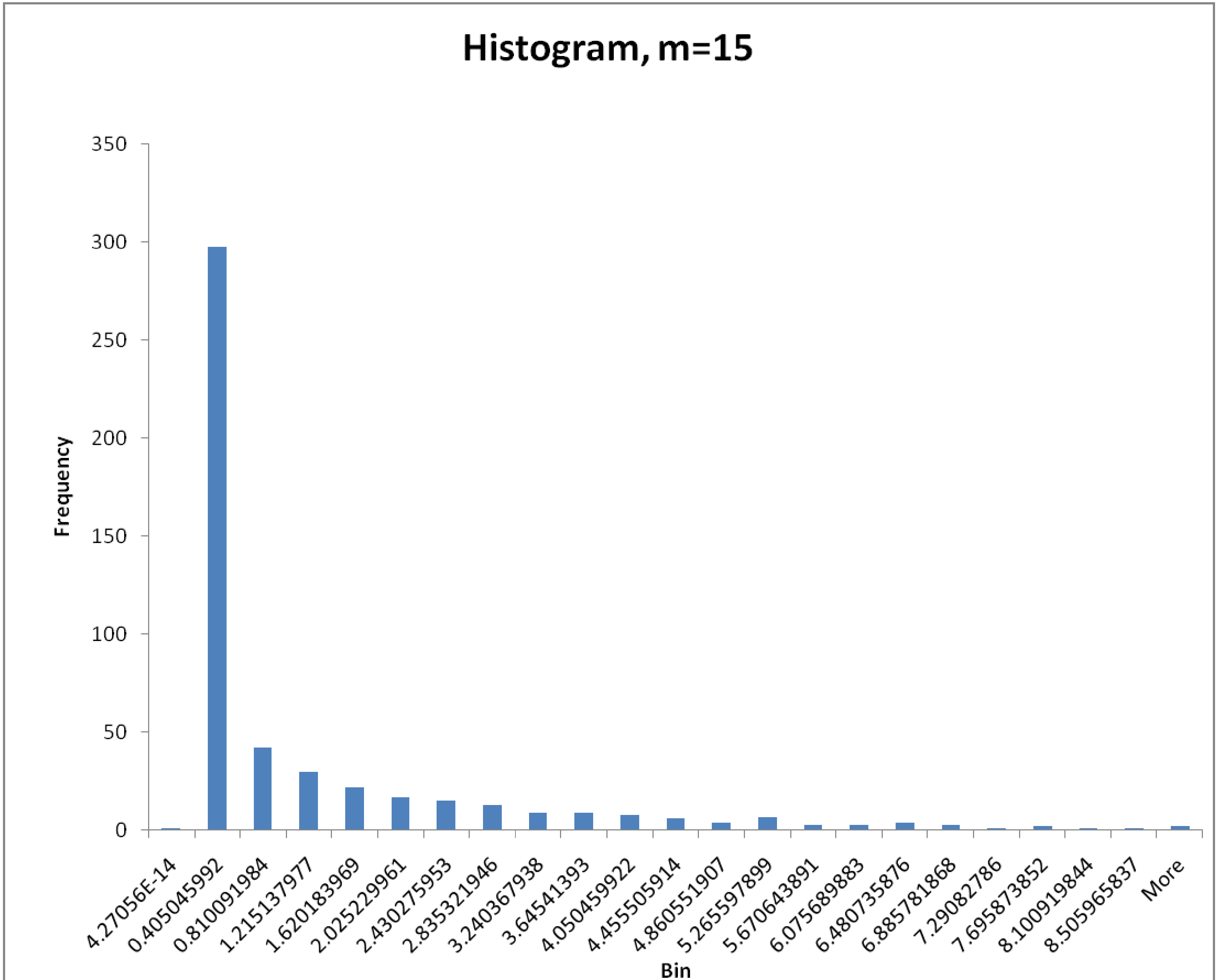
Lambda_1 = 33.59
A lot of detached evs suggests that there are additional factors.

Sorted excess returns, m=15 (without AIG)

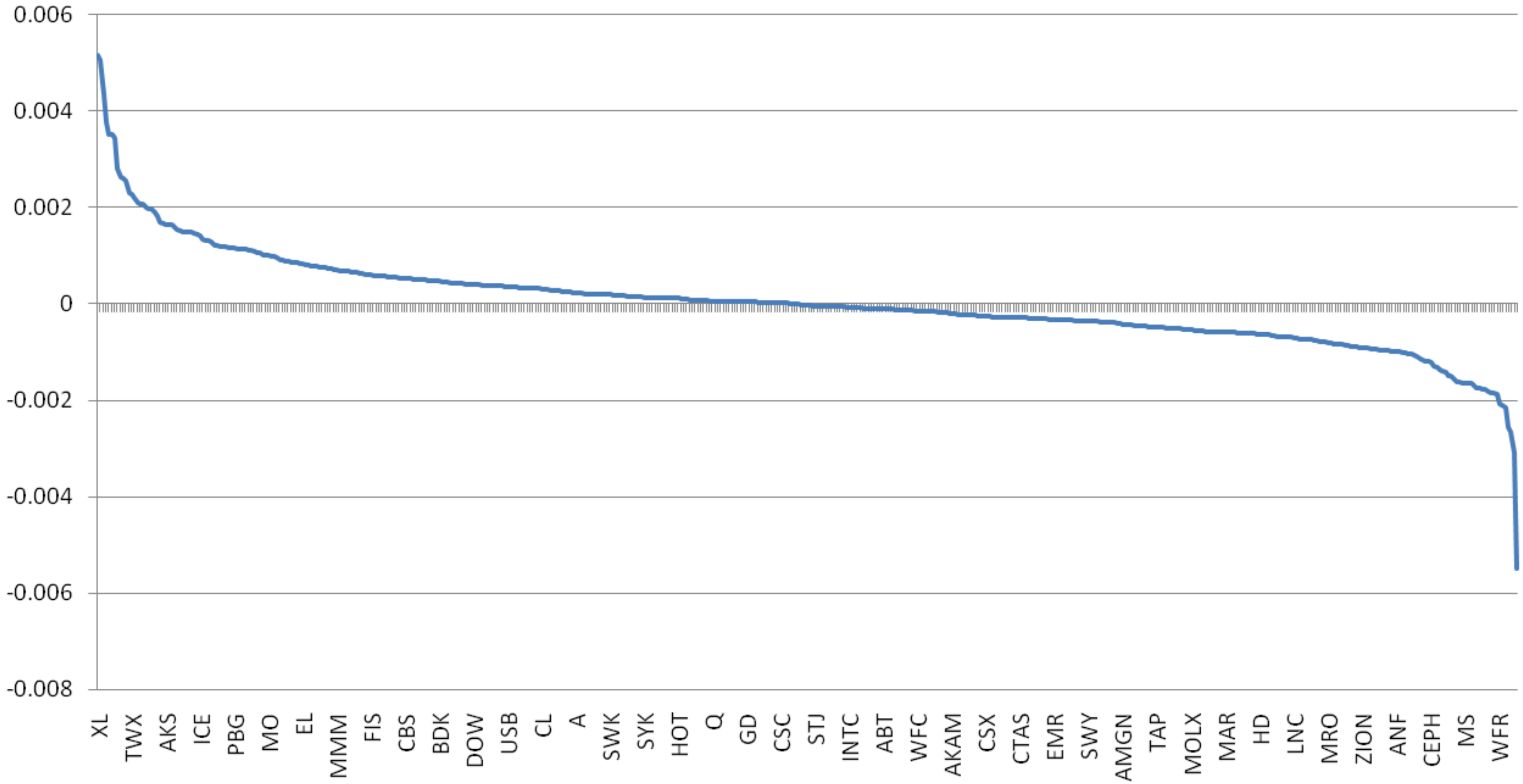


Top 100 eigenvalues of the correlation matrix of residuals (m=15)

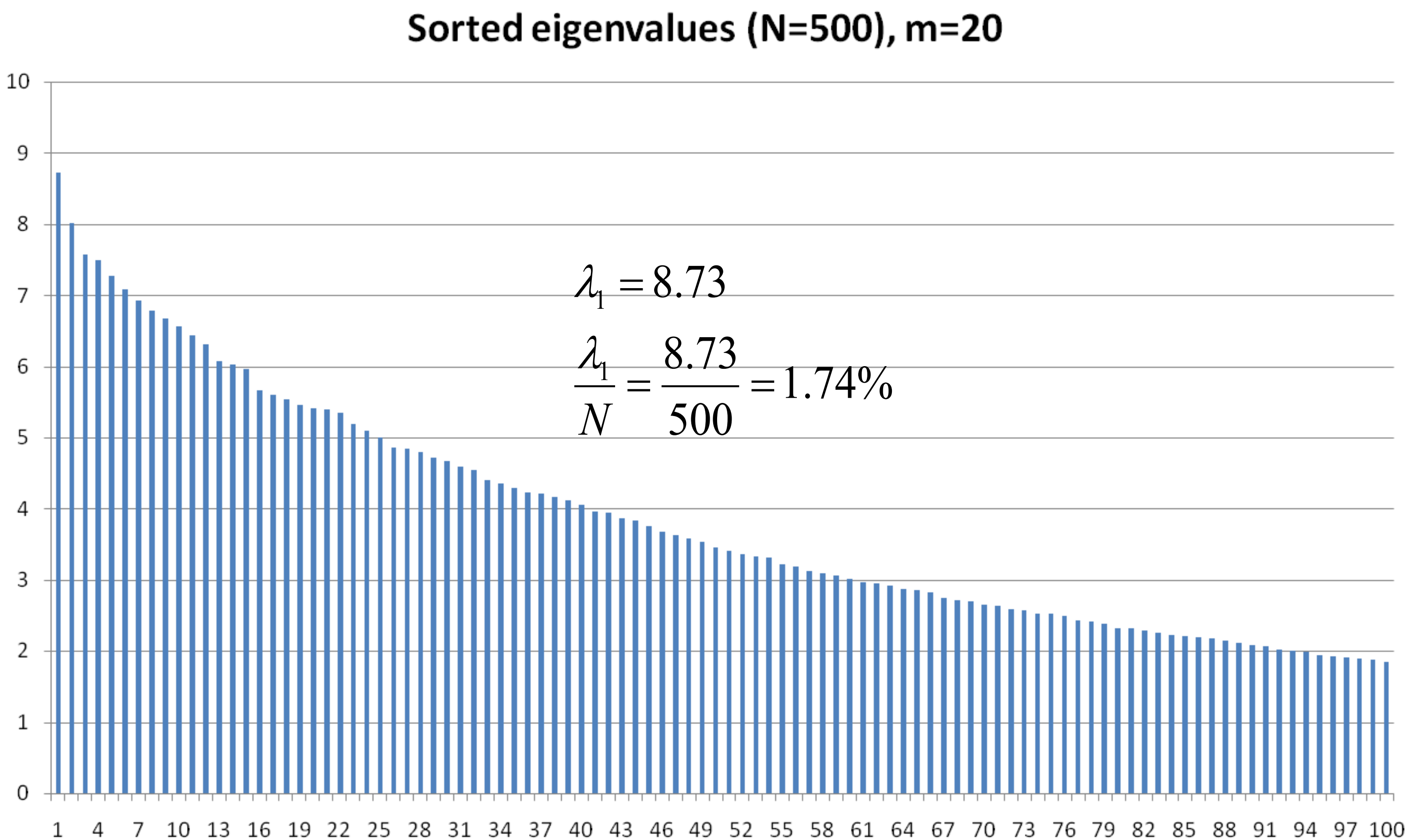



Histogram, m=15
Frequency
350
300
250
200
150
100
50
0
4.27056E-14
0.405045992
0.810091984
1.215137977
1.620183969
2.025229961
2.430275953
2.835321946
3.240367938
3.64541393
4.050459922
4.455505914
4.860551907
5.265597899
5.670643891
6.075689883
6.480735876
6.885781868
7.29082786
7.695873852
8.100919844
8.505965837
More
Bin

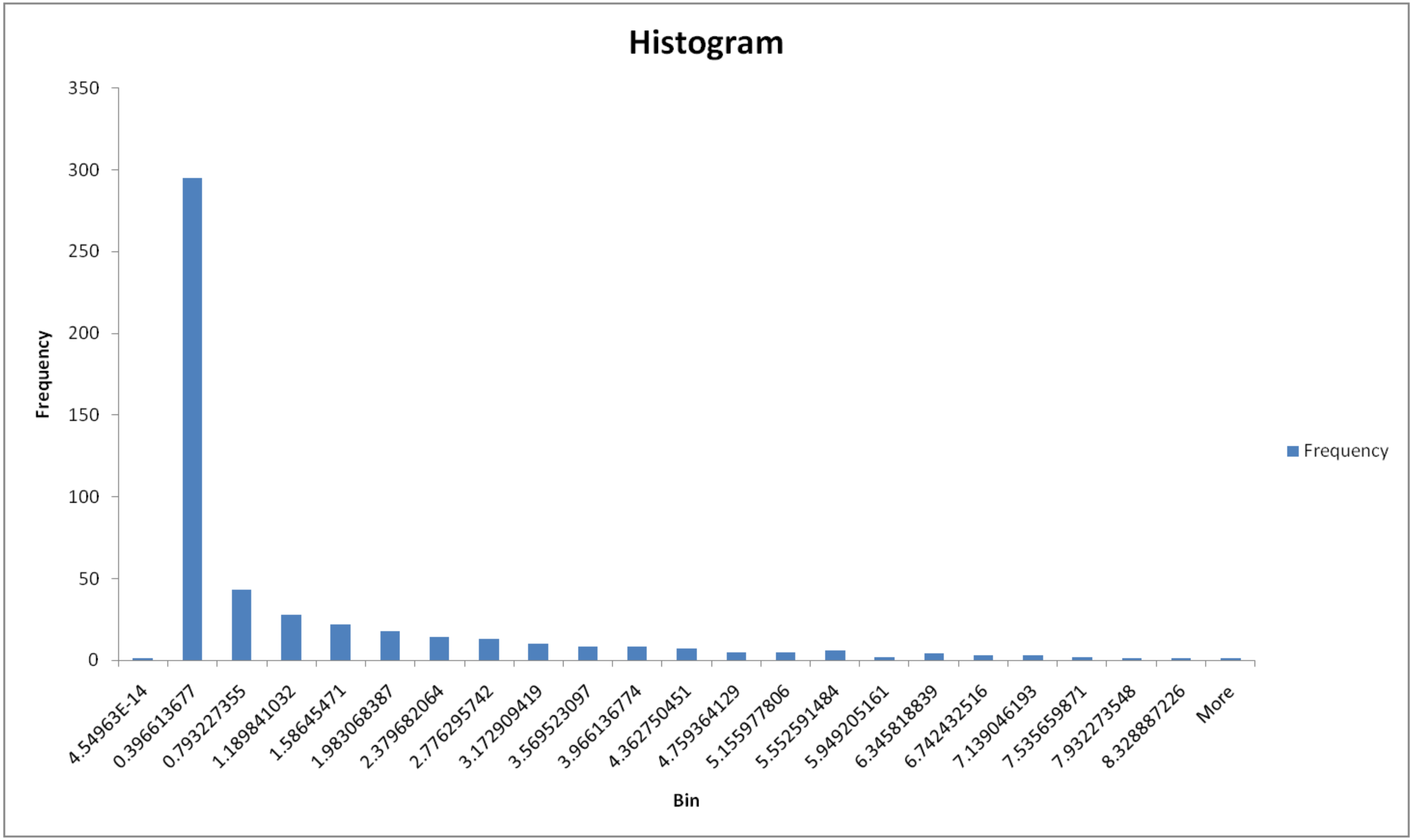
Sorted Excess Returns, m=20



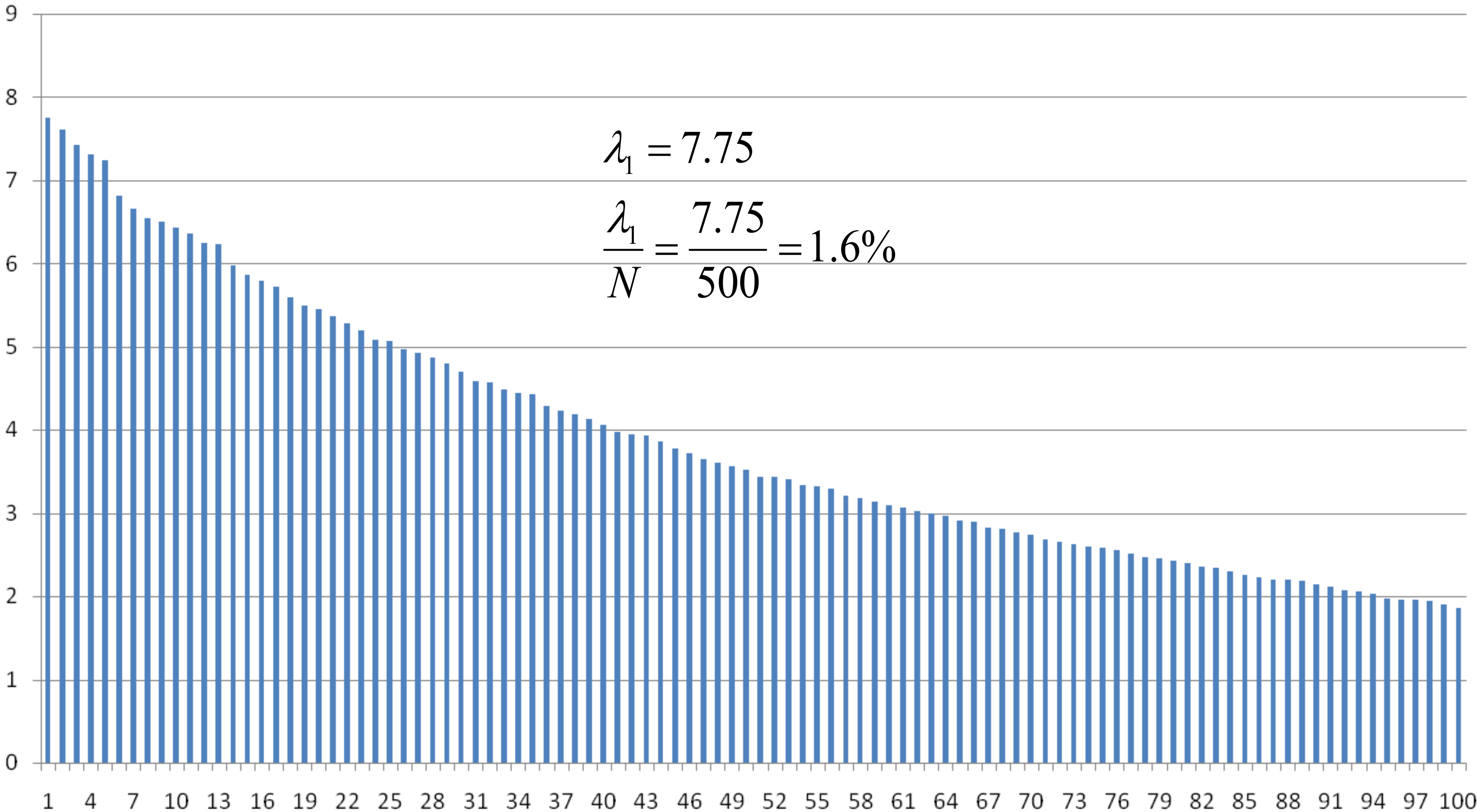
Sorted Eigenvalues, m=20



Density of States m=20



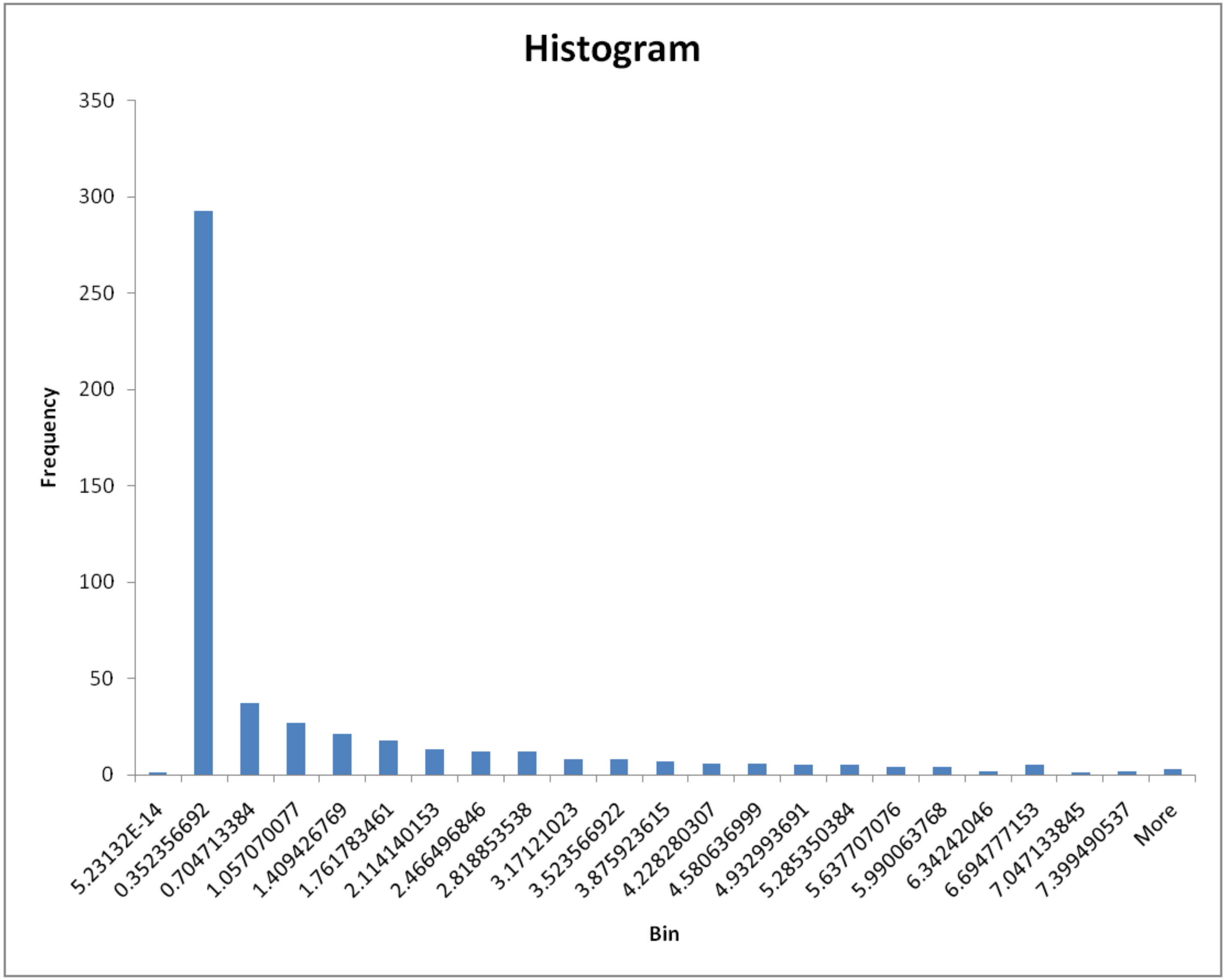
Sorted eigenvalues (N=500), m=30



$$\lambda_1 = 7.75$$

$$\frac{\lambda_1}{N} = \frac{7.75}{500} = 1.6\%$$

Density of States m=30



Excess returns (alpha) as a function of the number of eigenportfolios (m)

m	max	min	average	average abs	stdev
1	0.6283%	-0.3941%	0.0196%	0.0776%	0.1129%
15	0.5095%	-0.5096%	0.0065%	0.0687%	0.1004%
20	0.5095%	-0.5485%	0.0968%	0.0667%	0.0968%
30	0.5095%	-0.3957%	0.0049%	0.0664%	0.0960%

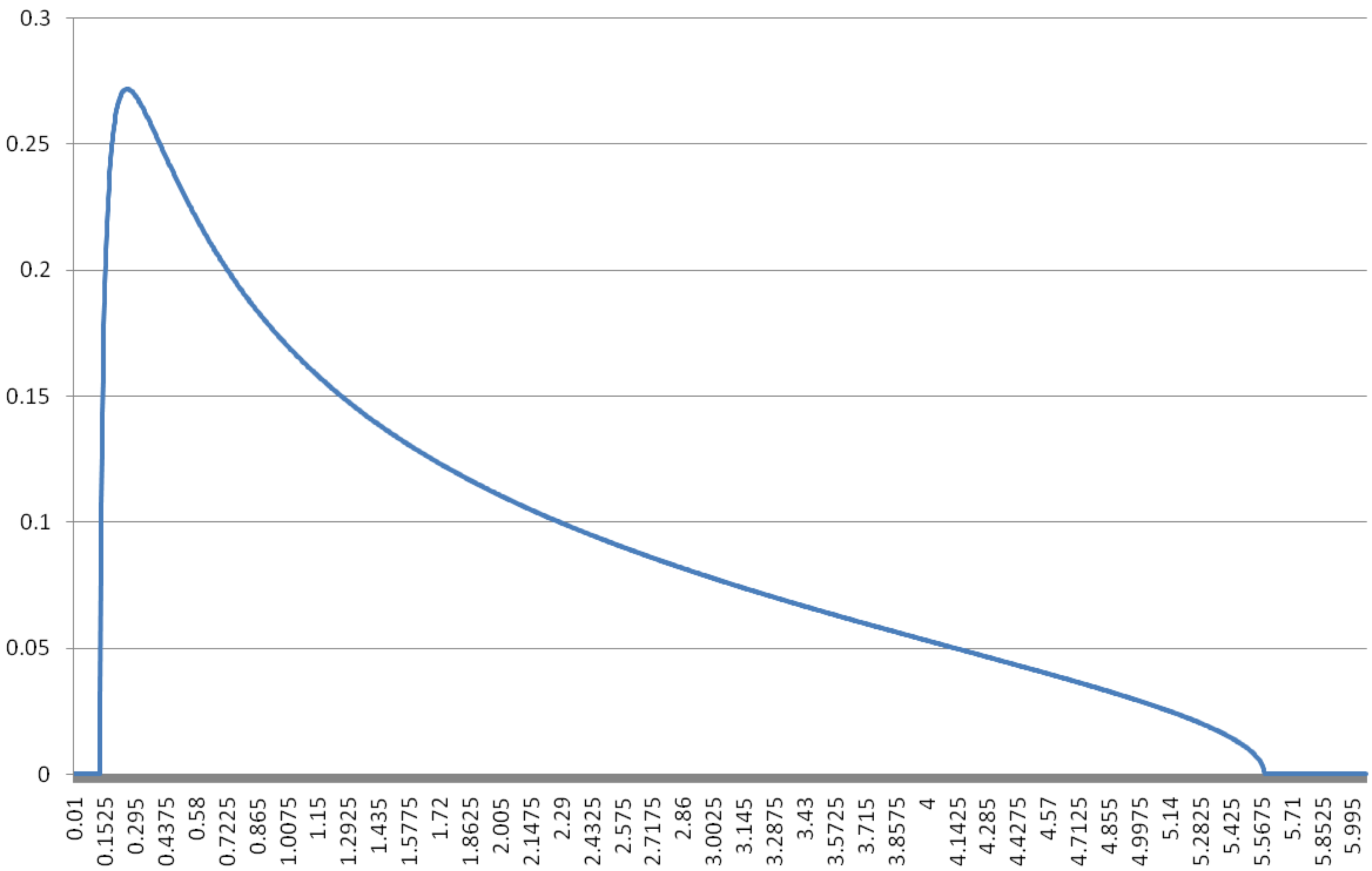
Marcenko-Pastur Distribution for the DOS of a Random Correlation Matrix

<u>Theorem</u>: Let *X* be a *T* by *N* matrix of standardized normal random variables and let *C=X'X*. Then, the DOS of C approaches the Marcenko Pastur distribution as *N,T* tend to infinity with the ratio *N/T* held constant.

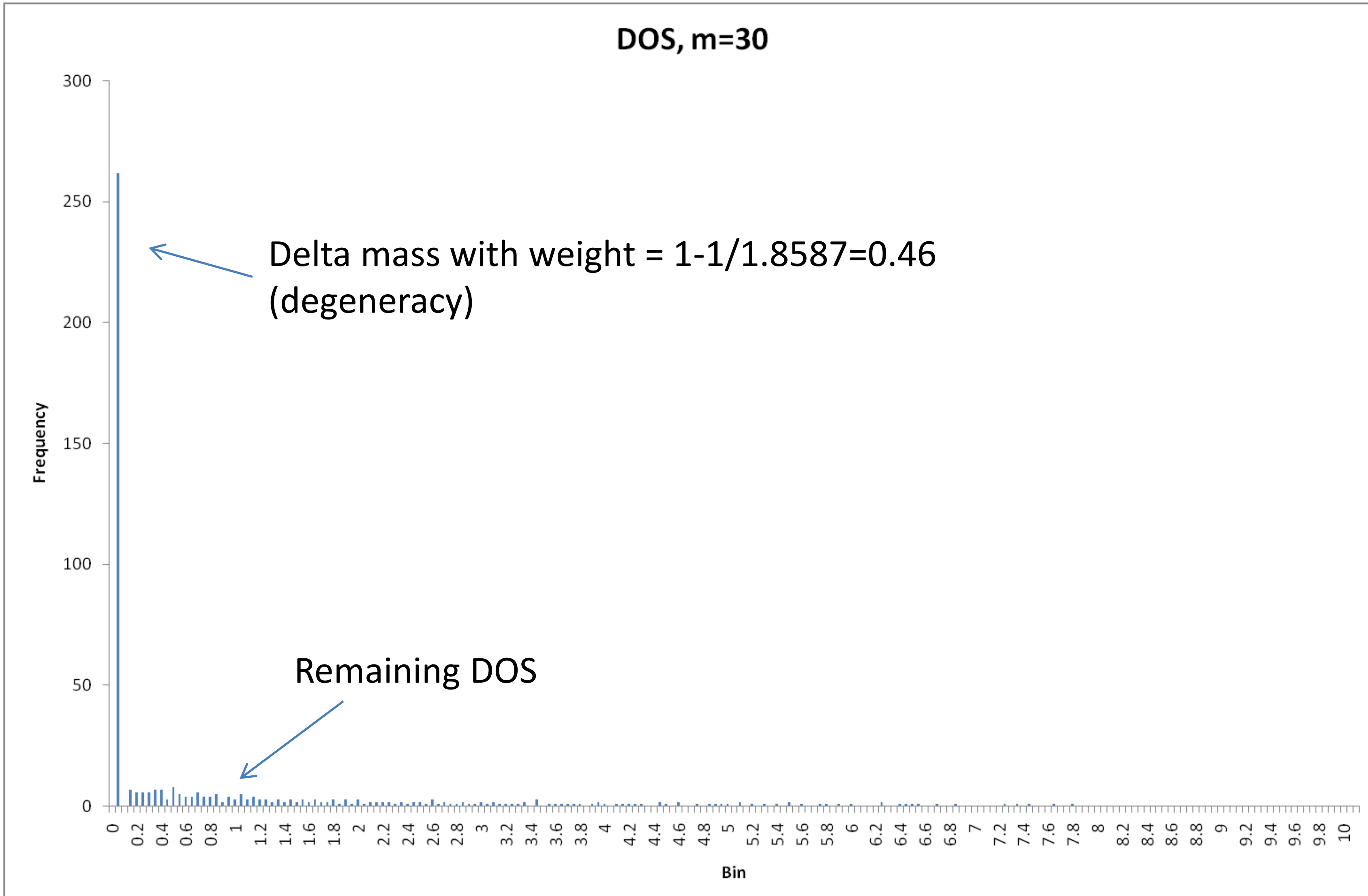
$$\gamma = \frac{N}{T} \qquad \lambda_+ = \left(1+\sqrt{\gamma}\right)^2 \qquad \lambda_- = \left(1-\sqrt{\gamma}\right)^2$$

$$MP(\lambda) = \left(1-\frac{1}{\gamma}\right)^+ \delta(\lambda) + \frac{1}{2\pi\gamma}\frac{\sqrt{(\lambda_+-\lambda)(\lambda-\lambda_-)}}{\lambda}$$

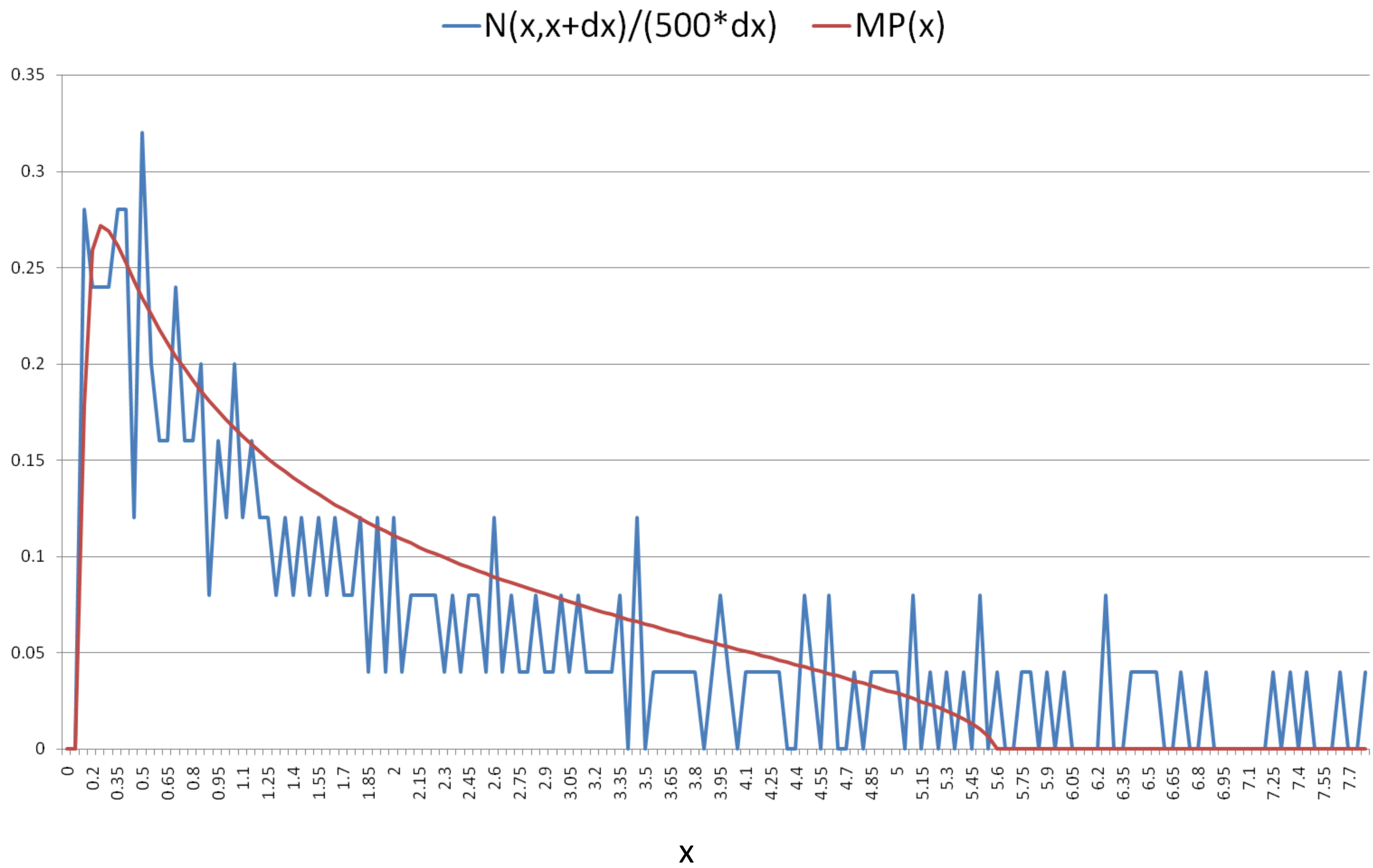
Marcenko-Pastur Distribution
gamma=500/269= 1.858736



Interpreting the DOS for the residuals in terms of Marcenko Pastur



Marcenko Pastur compared to data with m=30



Evaluating the use of ETFs as factors in APT

We found out how many eigenportfolios are needed approximately
to explain the systematic portion of stock returns using panel data for stock returns.

We obtained a matrix of random residuals if we choose m=15 or higher.

Since eigenportfolios are not tradable (except perhaps for the first one), this leaves us with the identification problem.

We perform an analysis of APT using <u>sector ETFs</u> as factors.

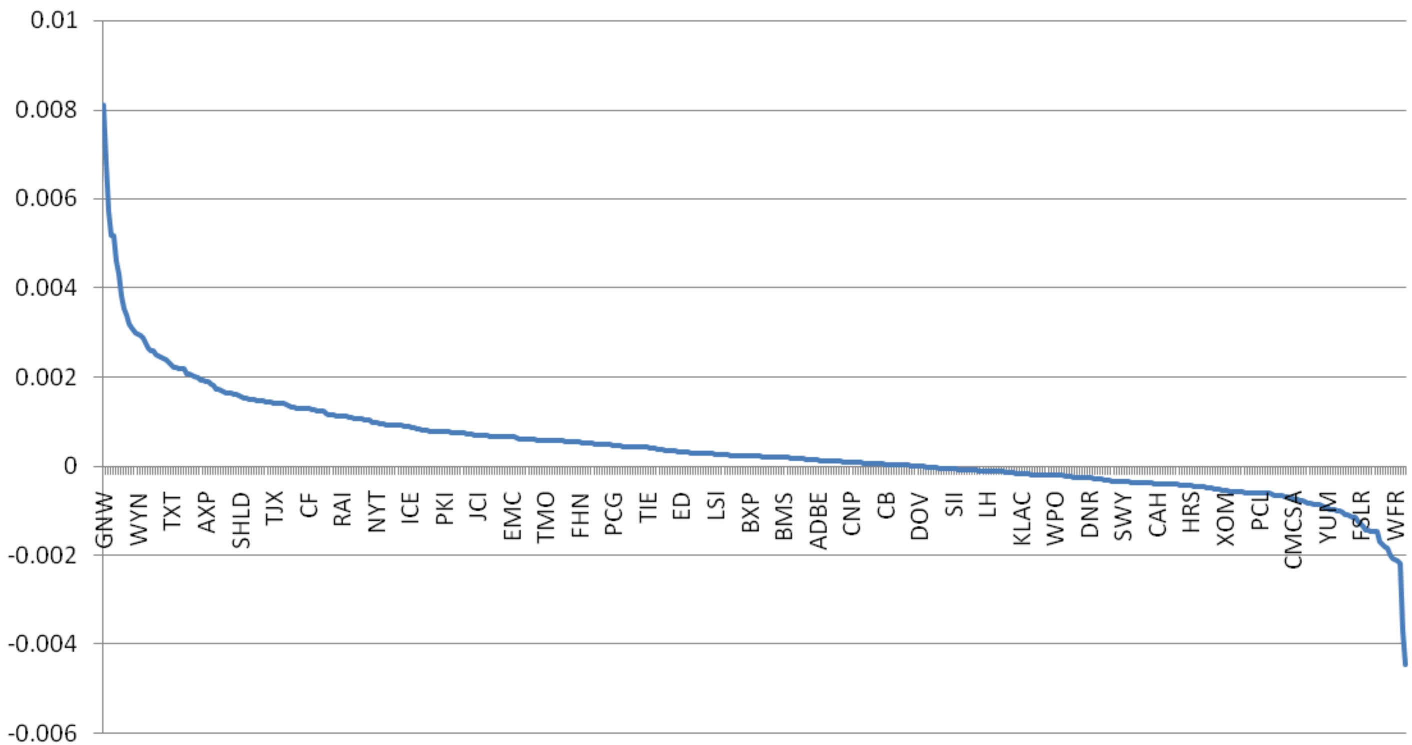
Three experiments:

* <u>Multiple regression</u> on 19 ETFs

* <u>Matching pursuit</u> on 19 ETFs

* Association of a <u>single ETF</u> to each stock

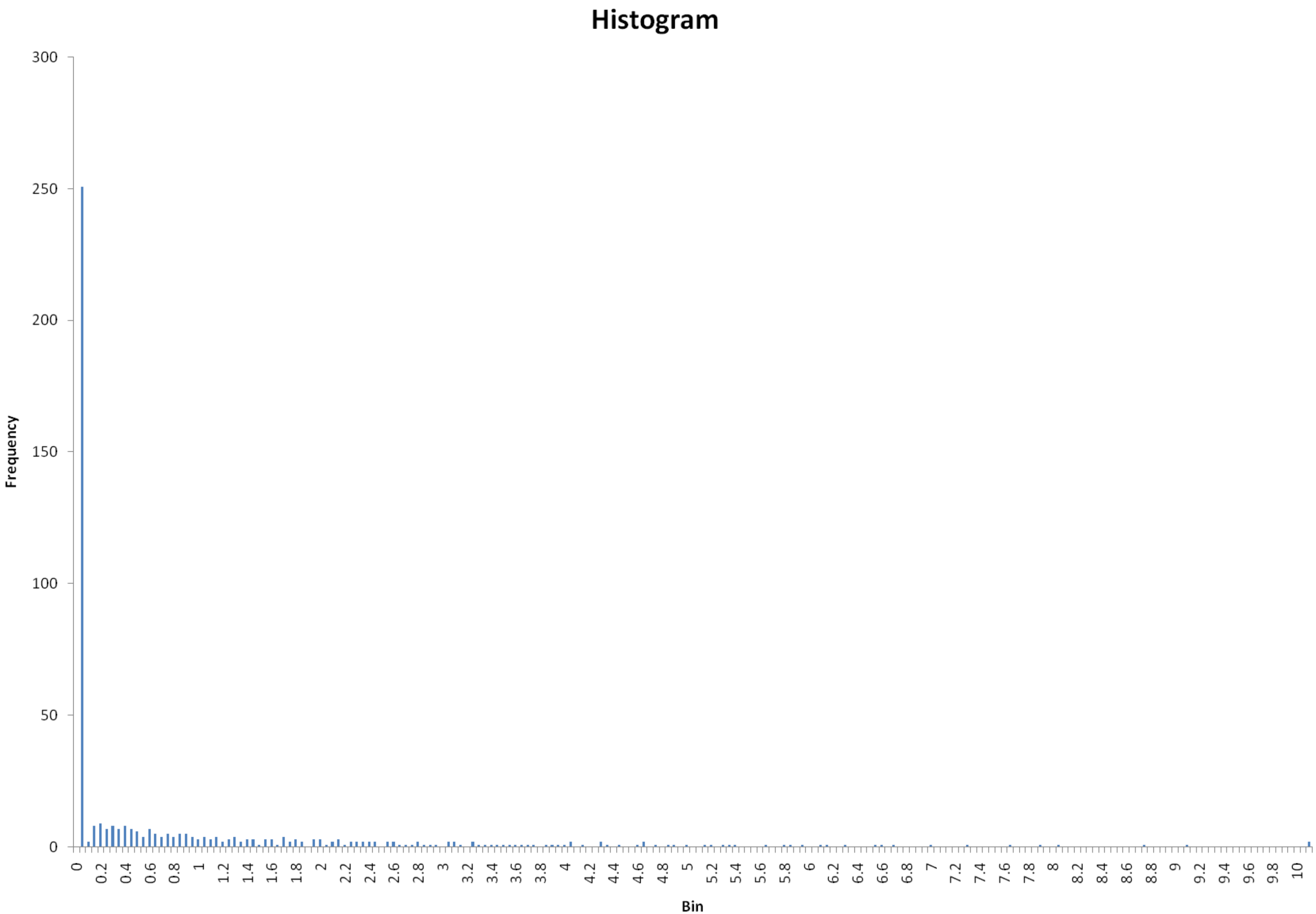
Sorted Excess Returns, Factors=19 ETFs Multiple Regression



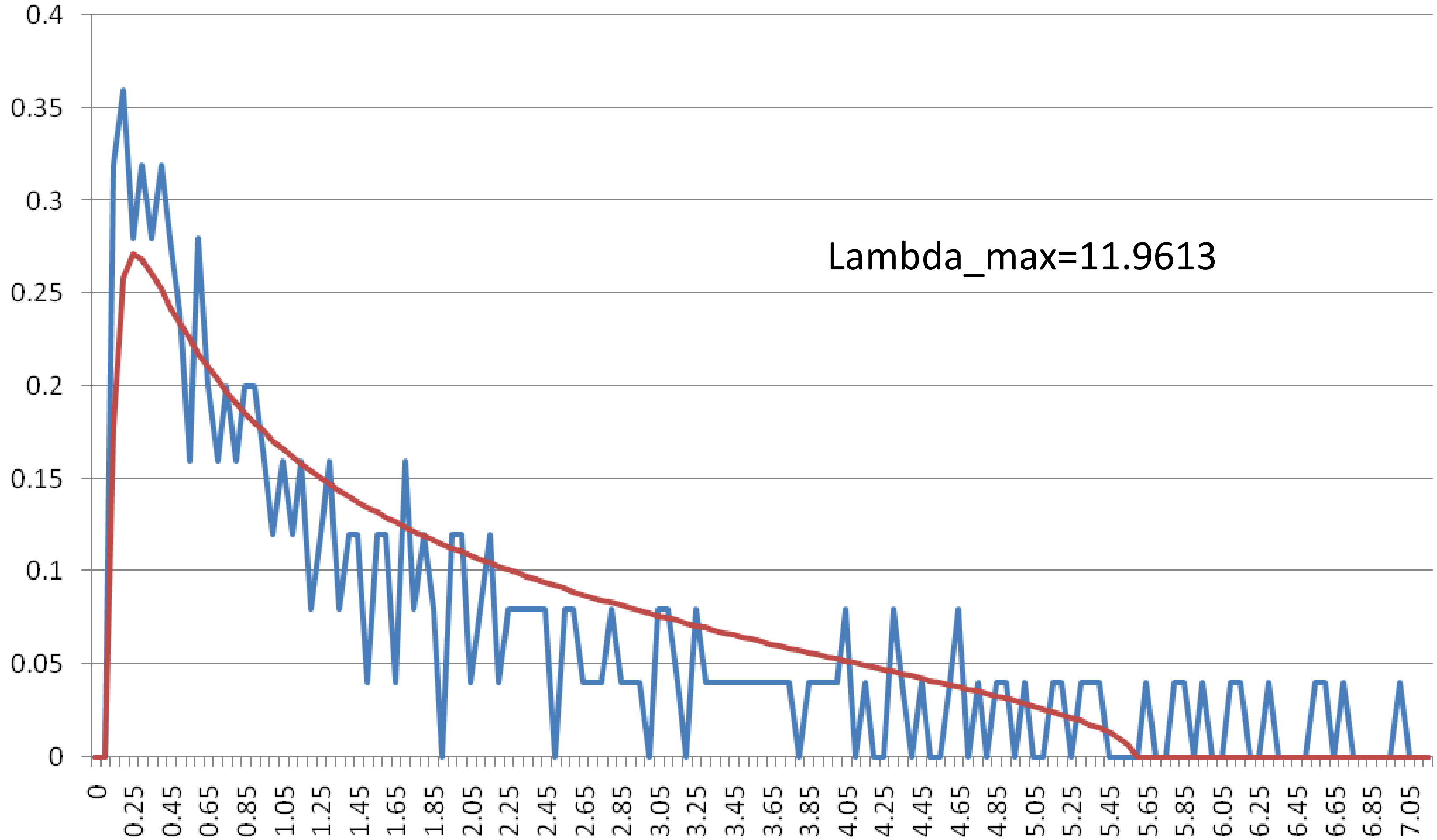
average	average abs	stdev
0.0390%	0.2755%	0.1167%

(AIG not included)

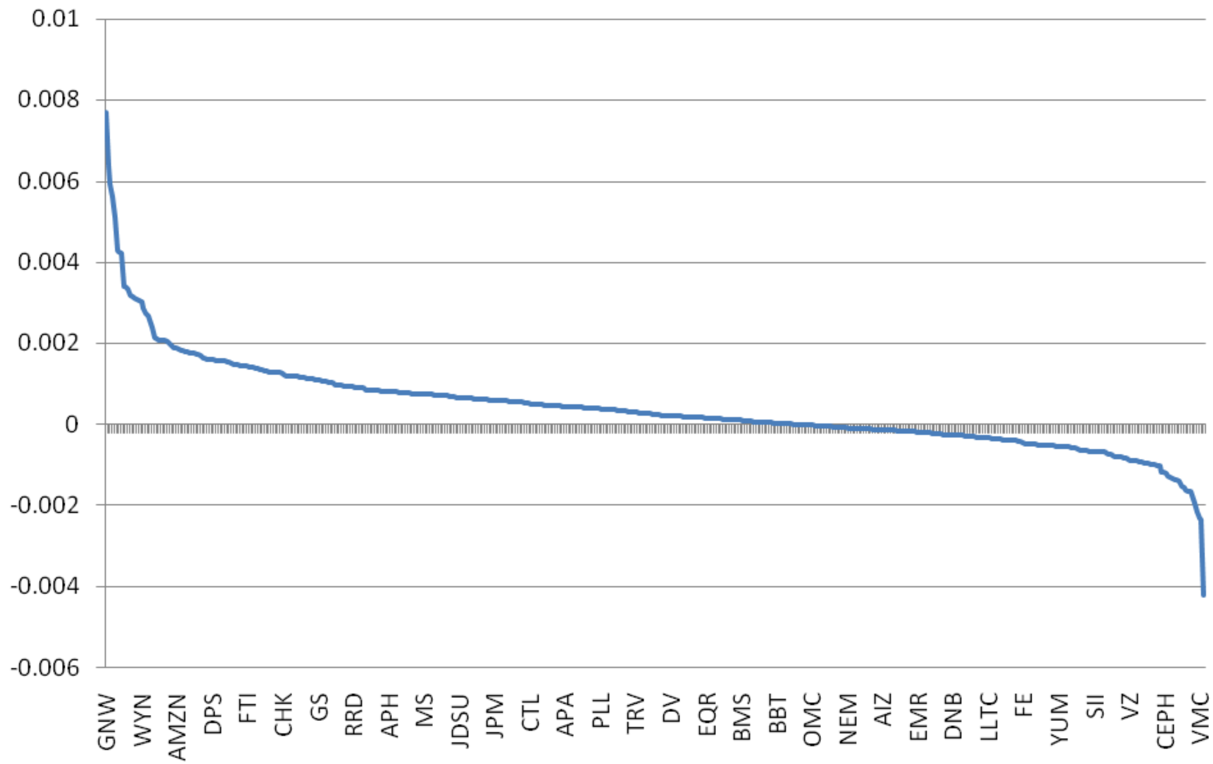
Density of States, residuals with 19 ETFs



After removing mass at zero (19 etfs, MR) (Red line=Marcenko-Pastur)

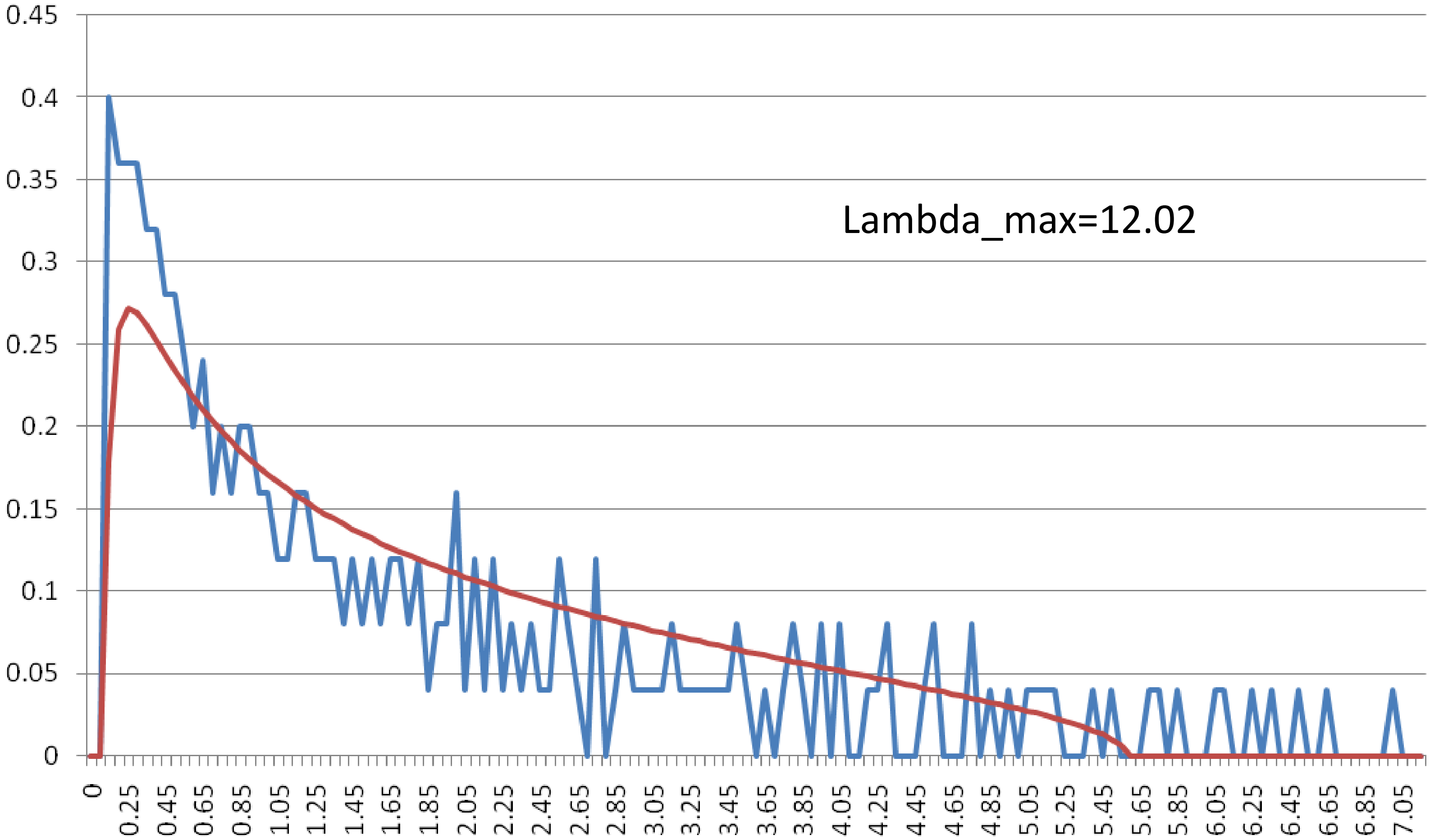


Excess Returns: Matching Pursuit, 19 ETFs (w/o AIG)

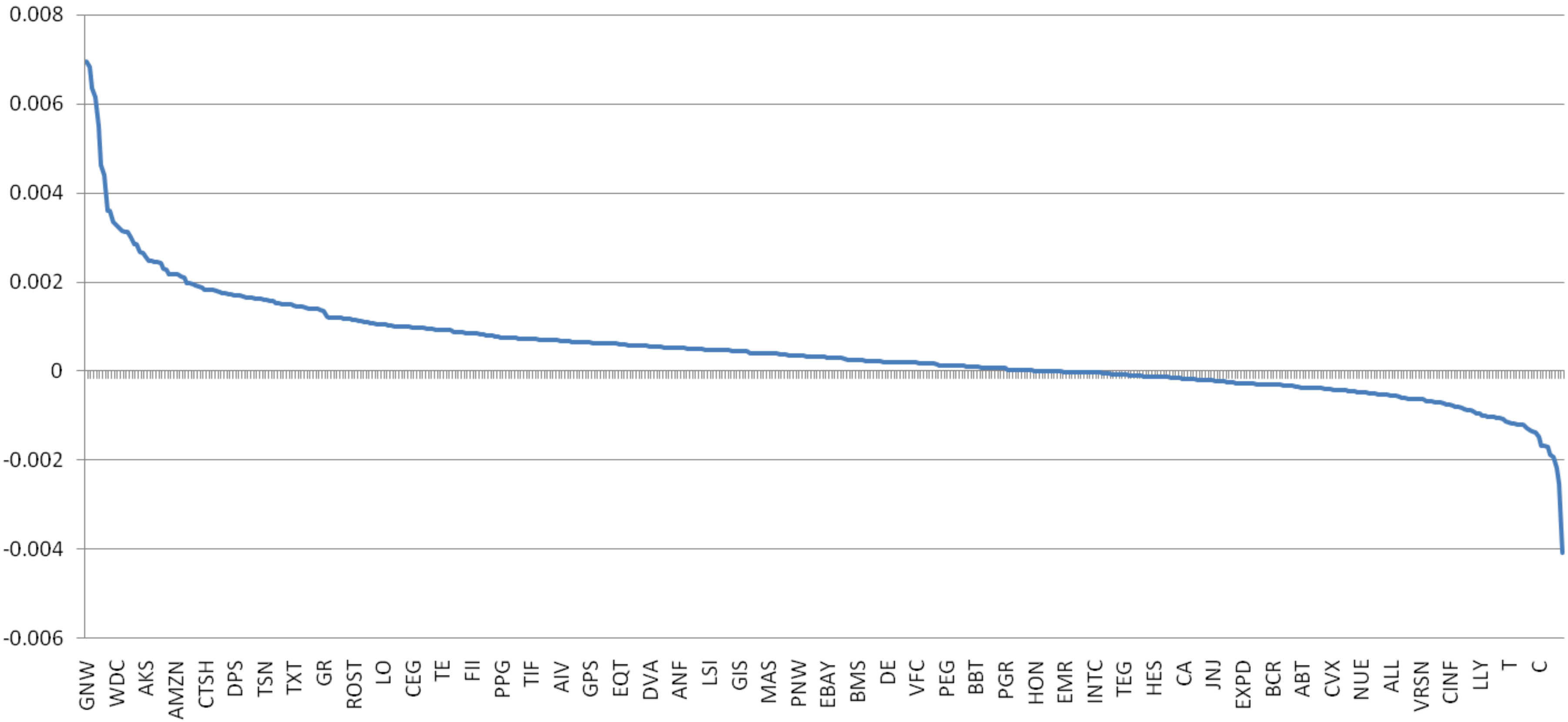


average	average abs	stdev
0.0405%	0.0799%	0.0904%

Noise Spectrum for Matching Pursuit Residuals

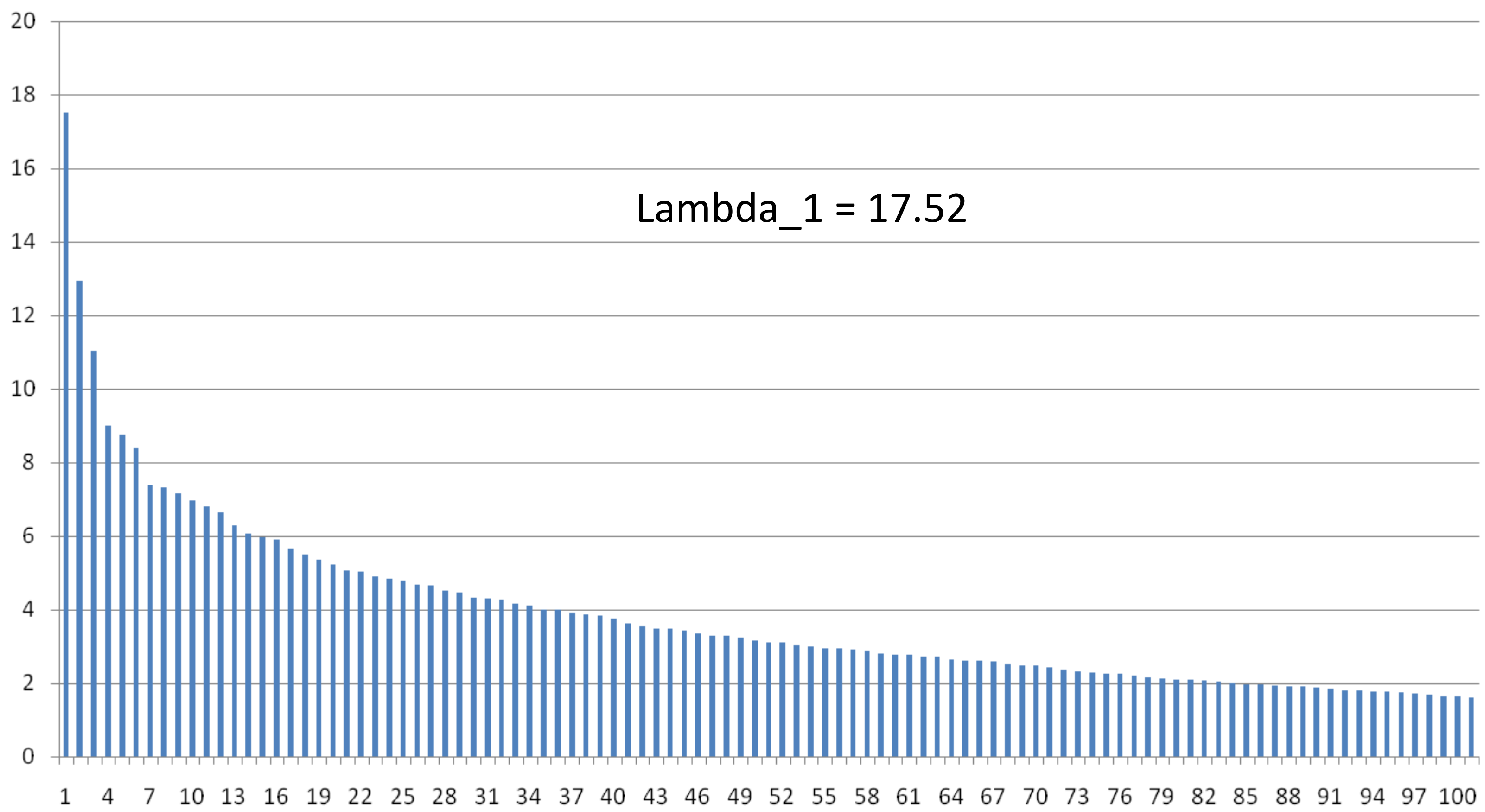


Excess Returns after projecting on the corresponding industry ETF for each stock

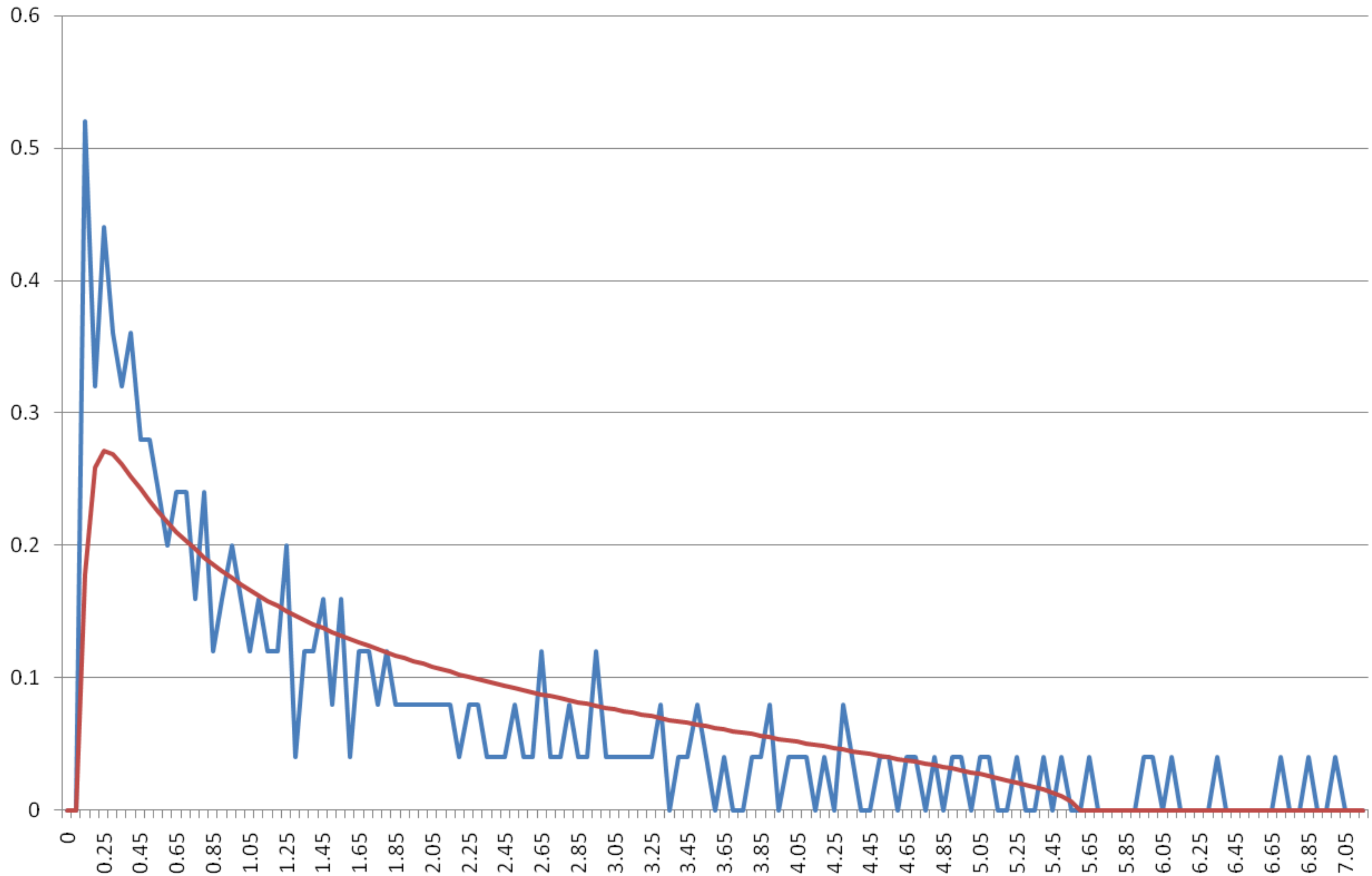


average	average abs	stdev
0.0461%	0.0823%	0.1153%

100 top eigenvalues for residuals after removing industry ETF



Density of States for Correlation matrix of residuals after removing the sector ETF for each stock



Dynamics

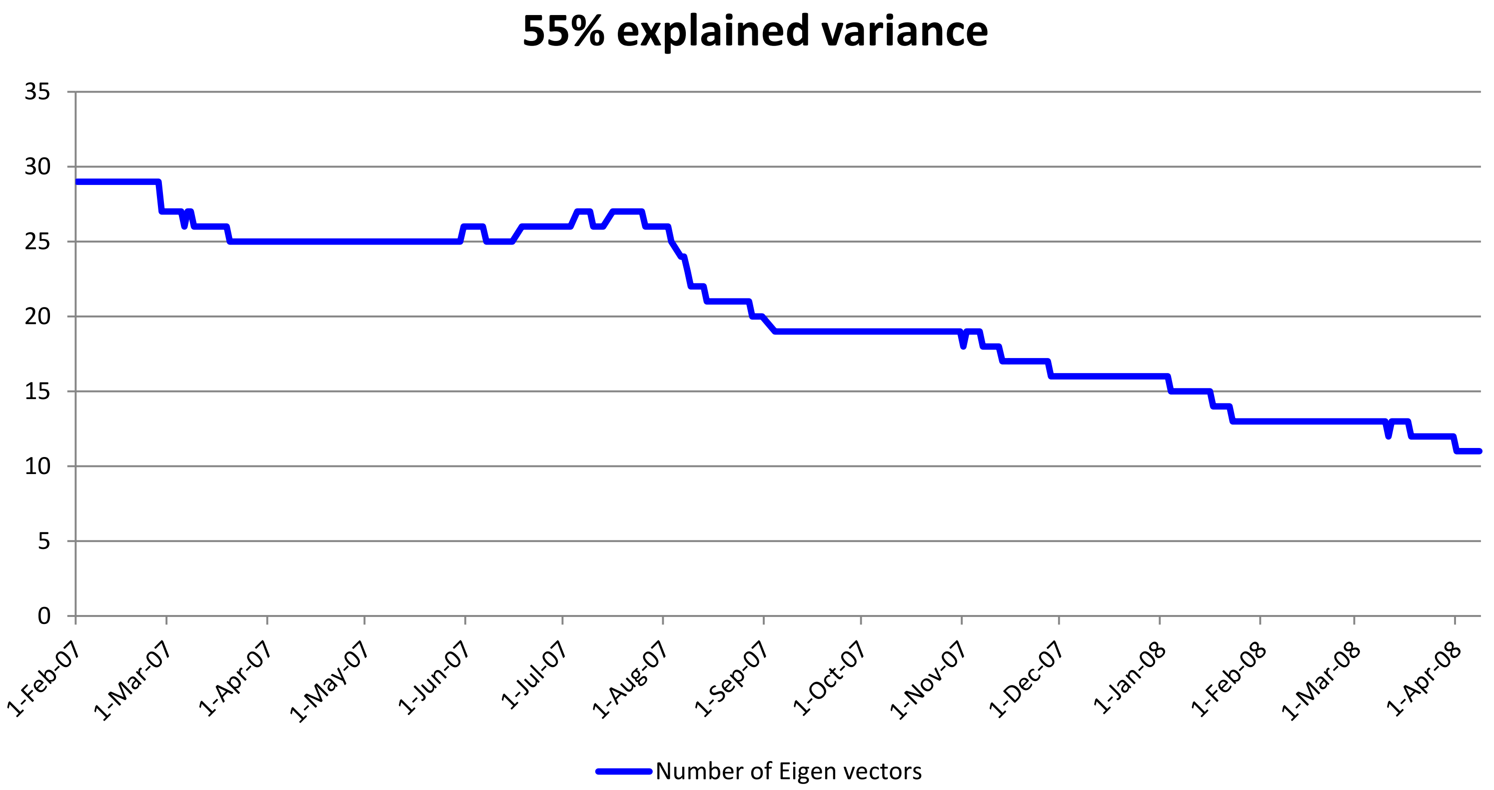
So far, the analysis that we made assumes a fixed window.

We should ask ourselves how these relationships change across time, i.e. if the factor count and the R-squared that we obtained are stable across time.

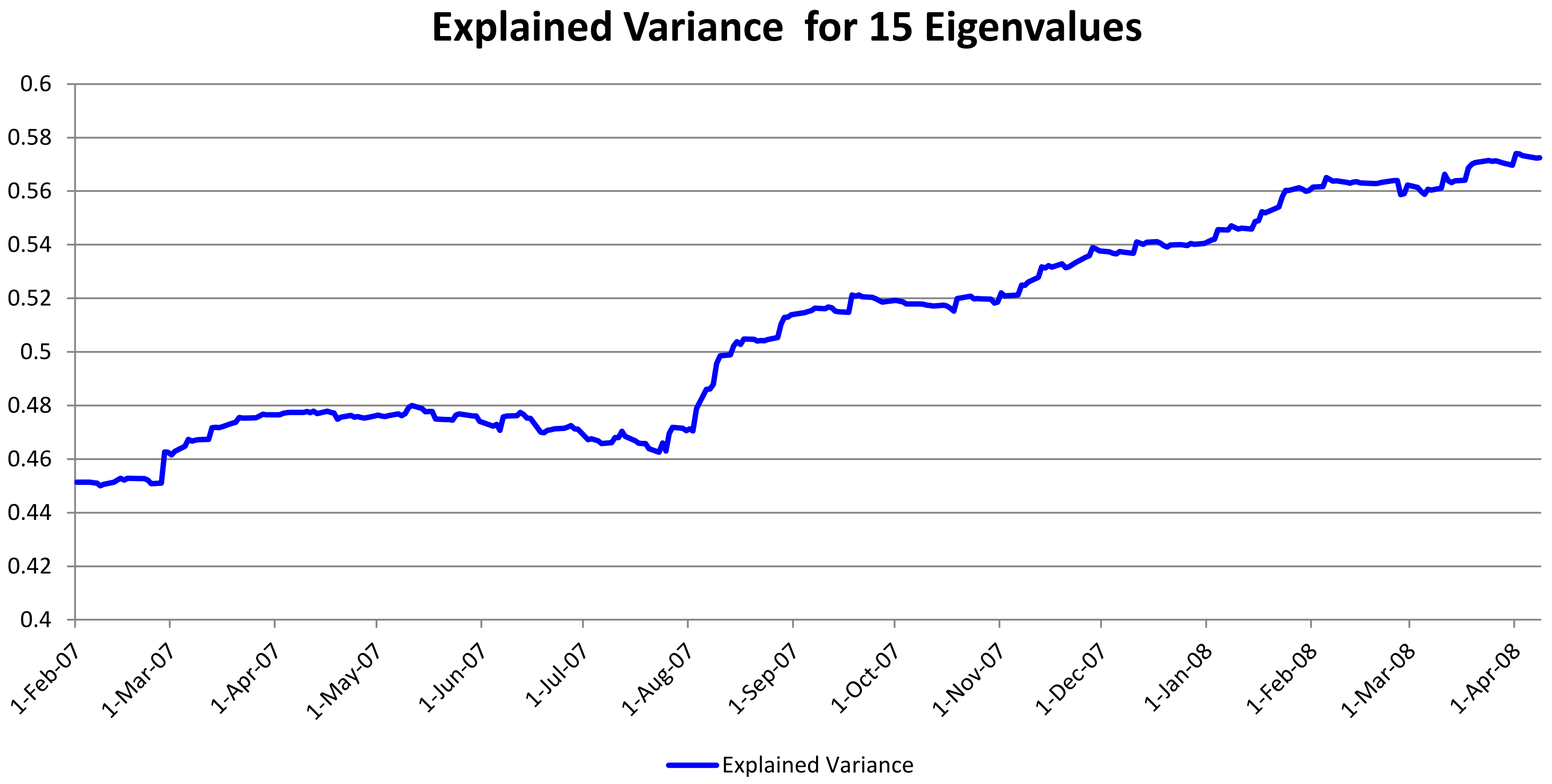
This is particularly important if the factor model is used for hedging or for relative valuation of stocks with respect to ETFs.

The following charts show some results of the PCA analysis viewed as time passes, using a moving window to calculate the eigenvalues and eigenvectors.

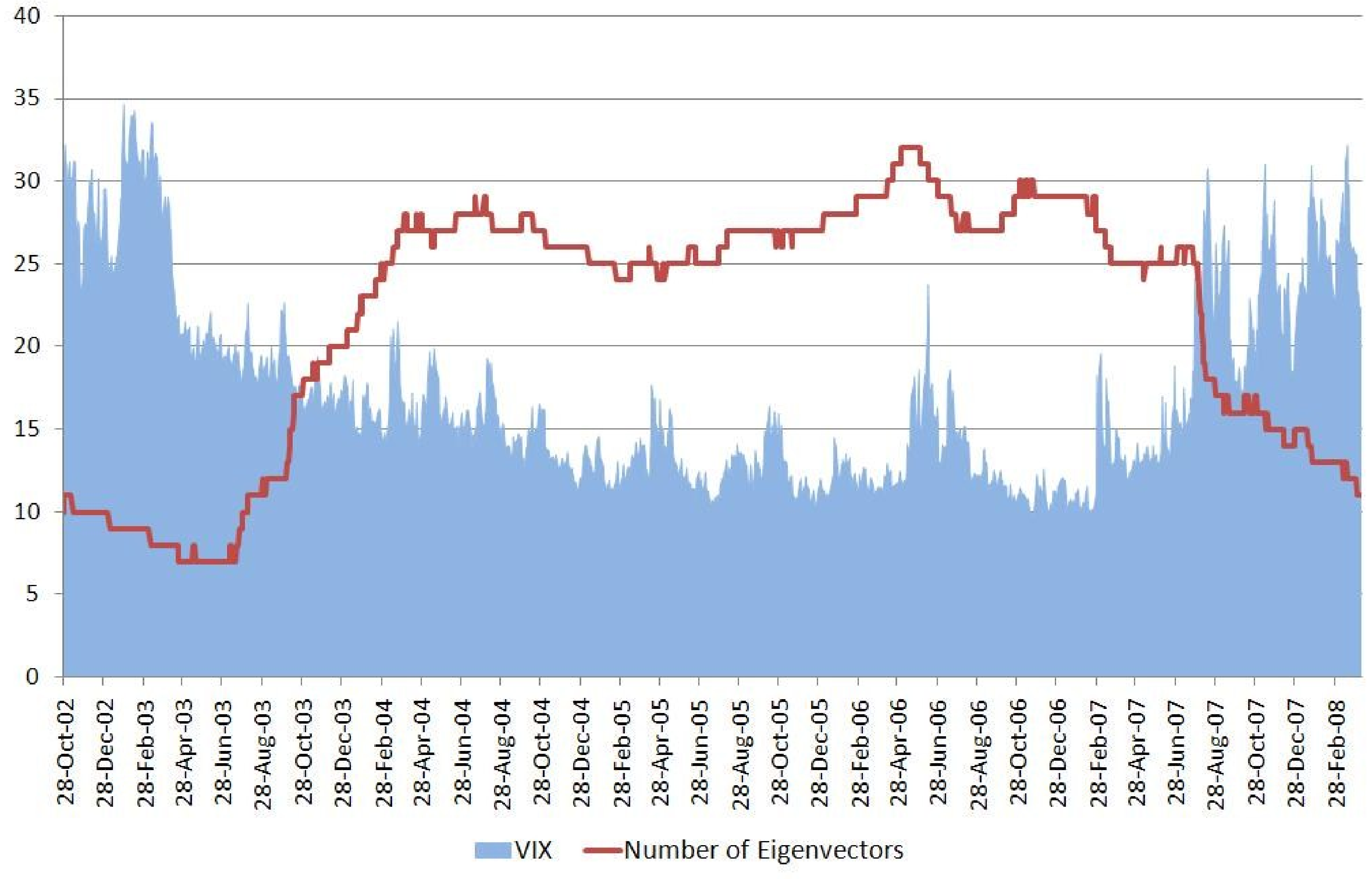
Number of significant eigenvectors at 55% level



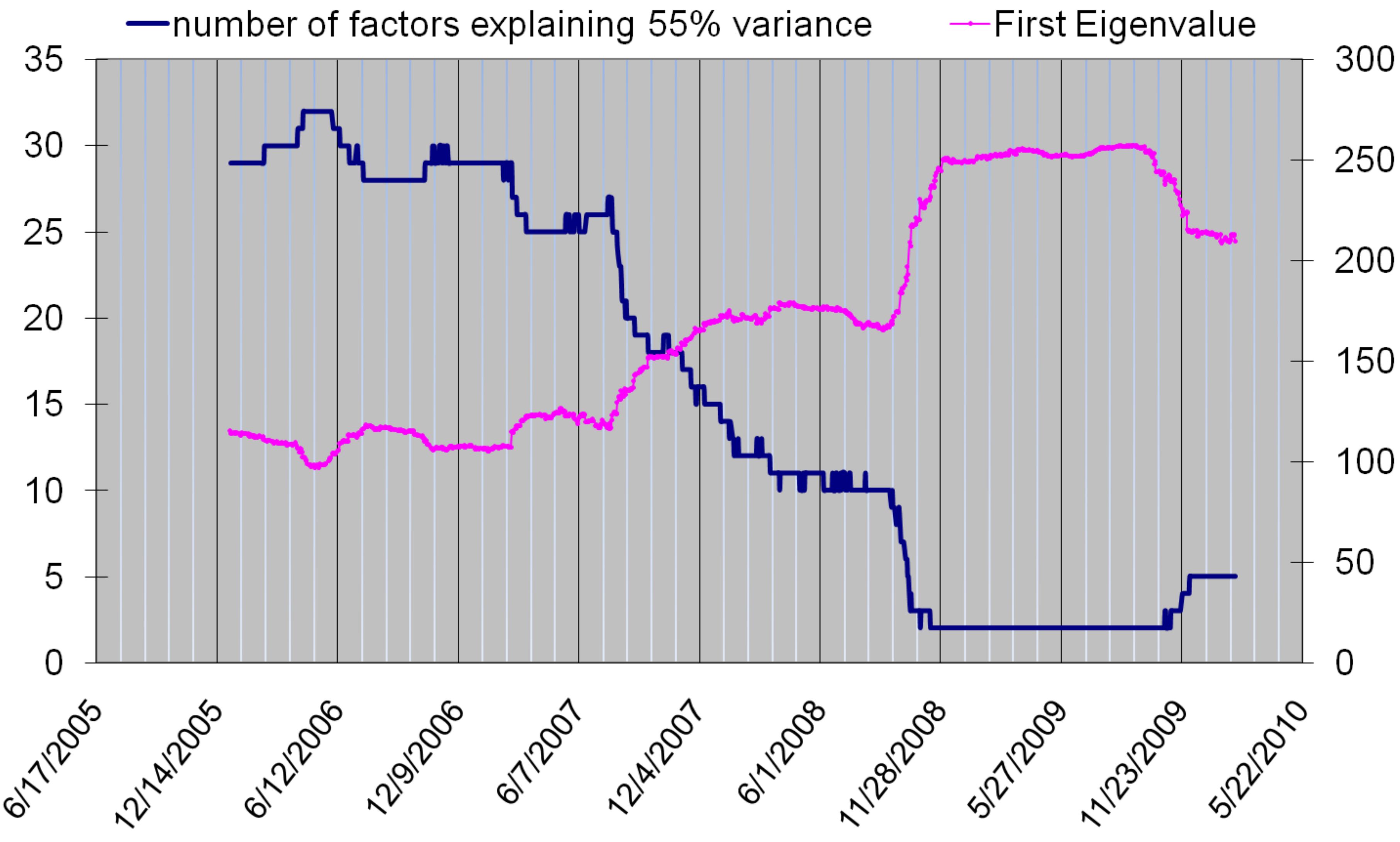
Fixed number of eigenvectors (factors)



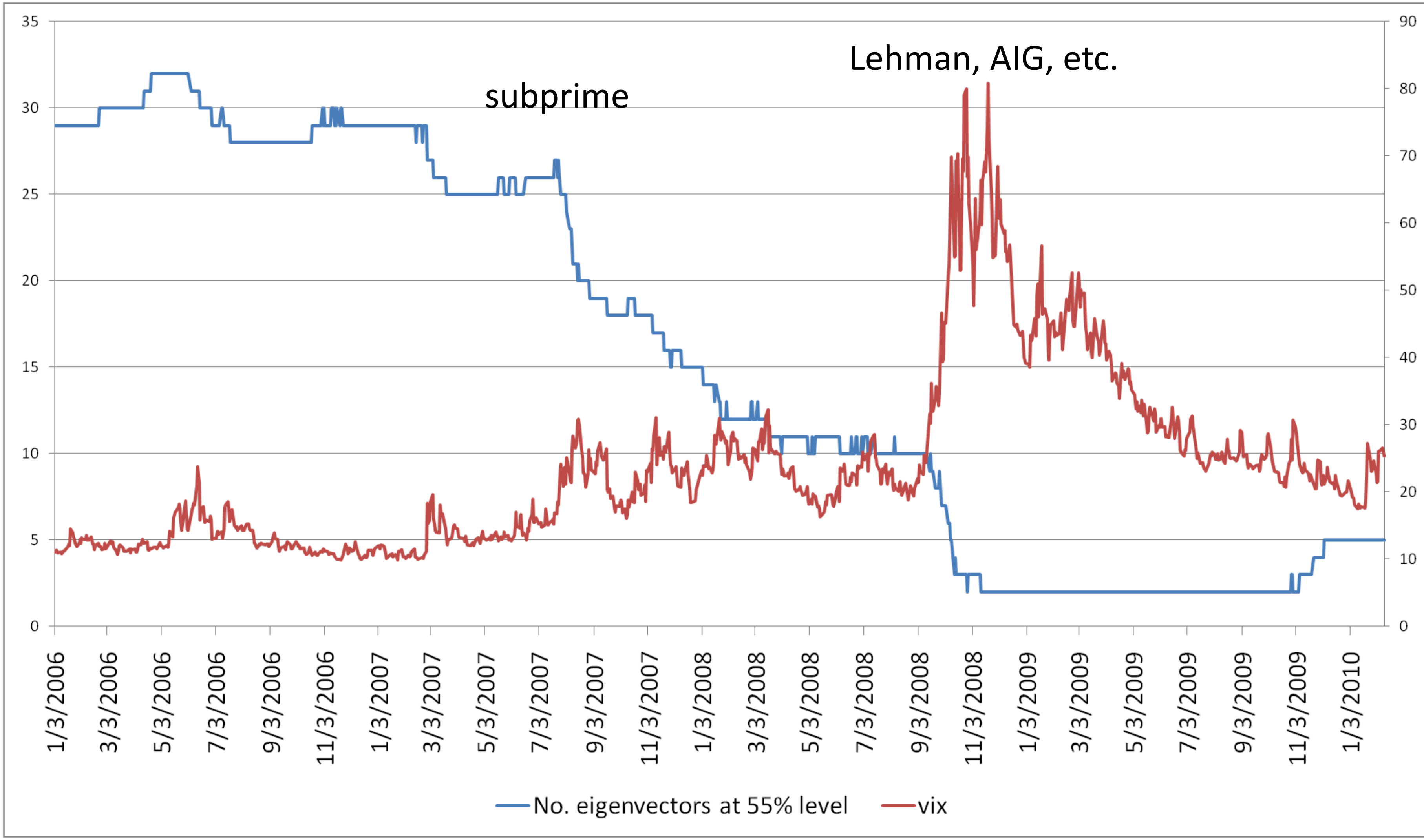
Number of factors explaining 55% of the variance versus VIX volatility index (2002-2008)



Number of explanatory factors vs. first eigenvalue of correlation matrix



Number of EVs versus VIX (1/2006-2/2010)



Application to Risk-management

Problem: given a portfolio of correlated assets, model the loss distribution over a given investment horizon

- Fat tails

- Correlations (systemic risk)

- Idiosyncratic risk

Rather than doing this one portfolio at a time, develop a set of factors and project the stock returns to compute betas

Model fat tails for the factors.

Risk-management system for equities

$$R_i = \sum_{k=1}^{m} \beta_{ik} F_k + \varepsilon_i$$

$F_k, k = 1,2...,m \sim$ Student - t with 3.5 degrees of freedom, independent

$\varepsilon_i = \sigma_i^r \nu_i, \quad \nu_i \sim$ "

$$\left(\sigma_i^r\right)^2 = Variance\left(R_i - \sum_{k=1}^{m} \beta_{ik} F_k \right)$$

The portfolio variability is given by the random variable

$$\boxed{d\Pi = \sum_{i=1}^{N} Q_i \left(\sum_{k=1}^{m} \beta_{ik} F_k + \varepsilon_i \right) = \sum_{k=1}^{m} \left(\sum_{i=1}^{N} Q_i \beta_{ik} \right) F_k + \sum_{i=1}^{N} Q_i \varepsilon_i}$$

$$Variance(d\Pi) = \sum_{k=1}^{m} \left(\sum_{i=1}^{N} Q_i \beta_{ik} \right)^2 \left(\sigma_{F_k}\right)^2 + \sum_{i=1}^{N} Q_i^{\,2} \left(\sigma_i^r\right)^2$$