Using MATLAB to Develop Asset Pricing Models

Robert Taylor

16 November 2006

Introduction

- Valuation Models and MATLAB
- Asset Pricing Models
- Estimation of the Models
- Seemingly-Unrelated Regression (SUR)
- Dealing with Missing Data
- MATLAB Example
- IPOs and Market Bubbles
- Discussion

Valuation Models and MATLAB

- Goal
  - Construct an asset-pricing model to explain asset returns
  - Examine possible asset-pricing anomalies shortly after IPOs
- Theory
  - Capital Asset Pricing Model
  - Fama & French Three-Factor Model
- Practice
  - Descriptive
  - Factor models great for risk management
  - Still need to predict factor returns to predict asset returns
  - Difficulties with missing data
  - Need roughly three years of data before a good estimate is available
- Solution
  - Use MATLAB to move the time horizon for analysis closer to IPO dates

Capital Asset Pricing Model (CAPM)

- CAPM of Sharpe (1964) and Lintner (1965)
  - Asset – Cash = Alpha + Beta x (Market – Cash)
  - Beta = cov(Asset, Market)/var(Market)
- Theory says Alpha = 0 and Beta ~ 1
  - Beta = 1 implies asset moves with market
  - Beta < 1 implies asset less volatile than market
  - Beta > 1 implies asset more volatile than market
- Efficient market hypothesis of Fama (1970)

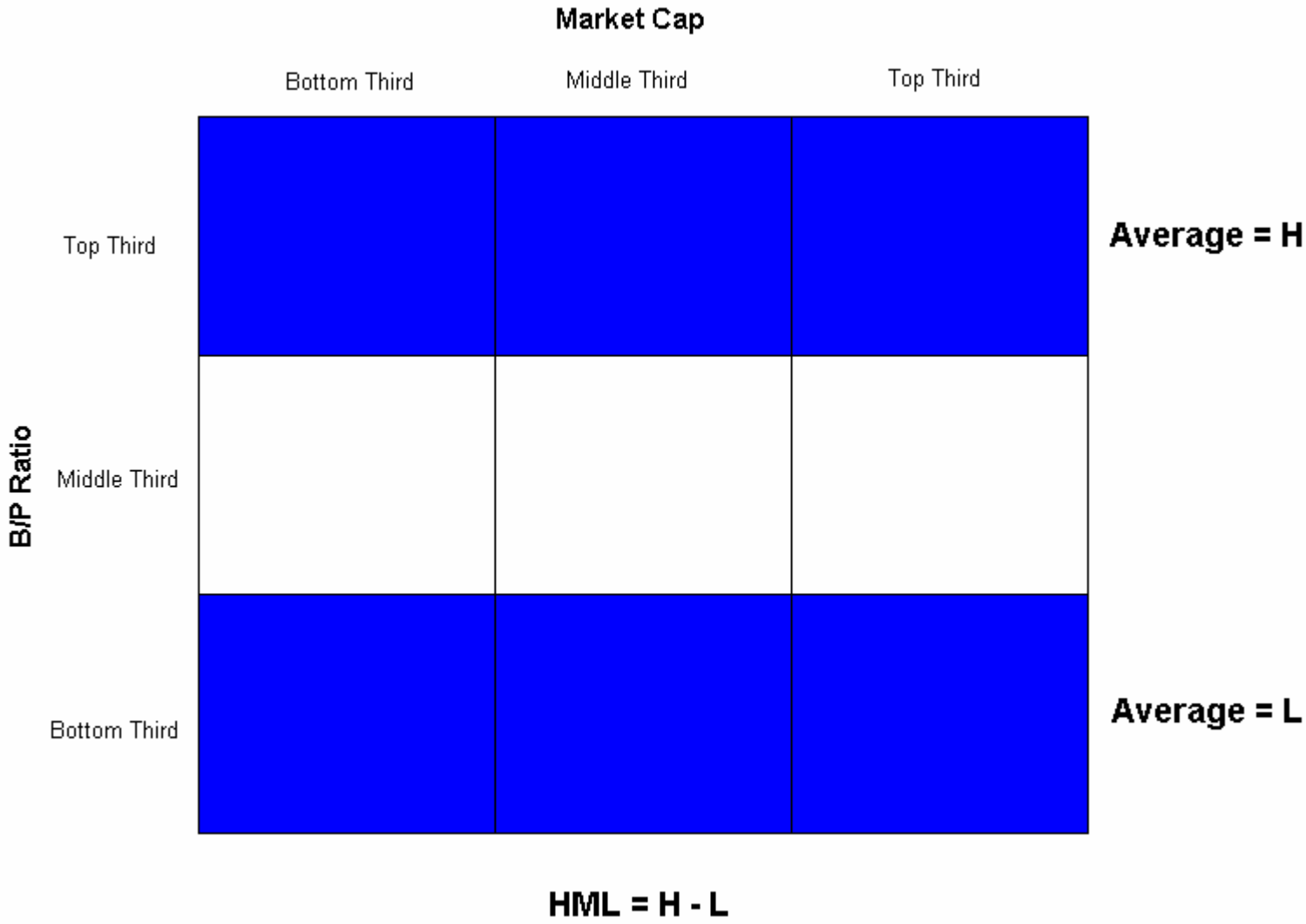
Cross-Sectional Return Models

- Arbitrage Pricing Theory (APT) of Ross (1976)
  - Linear combinations of factors explain asset returns
  - Does not say what those factors are or how many there should be
- Multiple risk and industry factor model of Rosenberg & McKibben (1973) and Rosenberg (1974)
  - Evolved into the BARRA US equity risk model with 68 factors
- Fama & French (1992, 1993, 1995, 1996) developed a parsimonious cross-sectional return model with only 3 factors

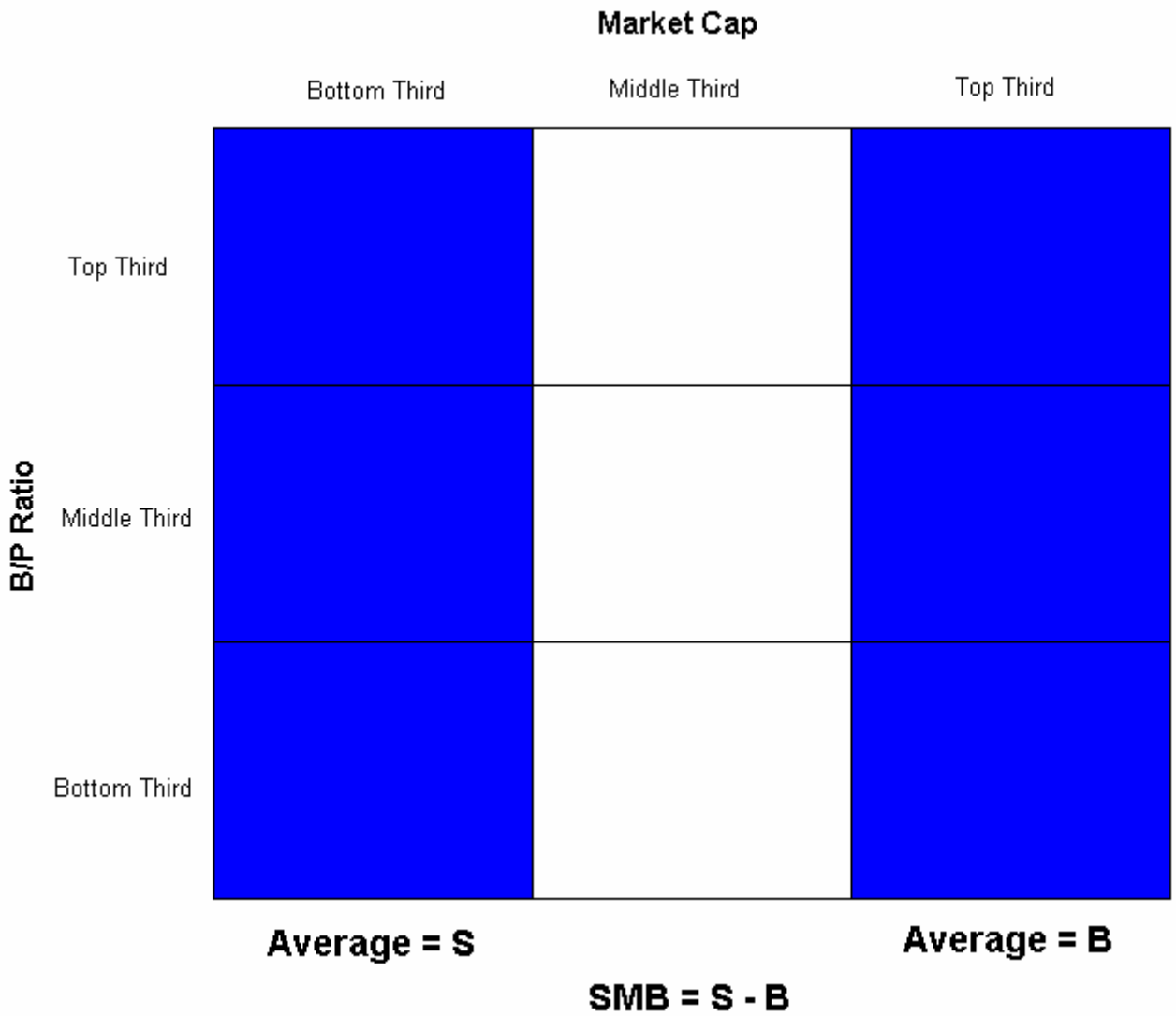
Fama & French Three-Factor Model

- Three-factor model of Fama & French (1993)
  - Asset – Cash = Alpha + Beta x (Market – Cash)
    + Gamma x SMB + Delta x HML
- Three factors
  - Beta – “volatility”
  - SMB – “small minus big” market capitalization
  - HML – “high minus low” book to price
- Model is very good with R-squares at around 92%!

HML Factor

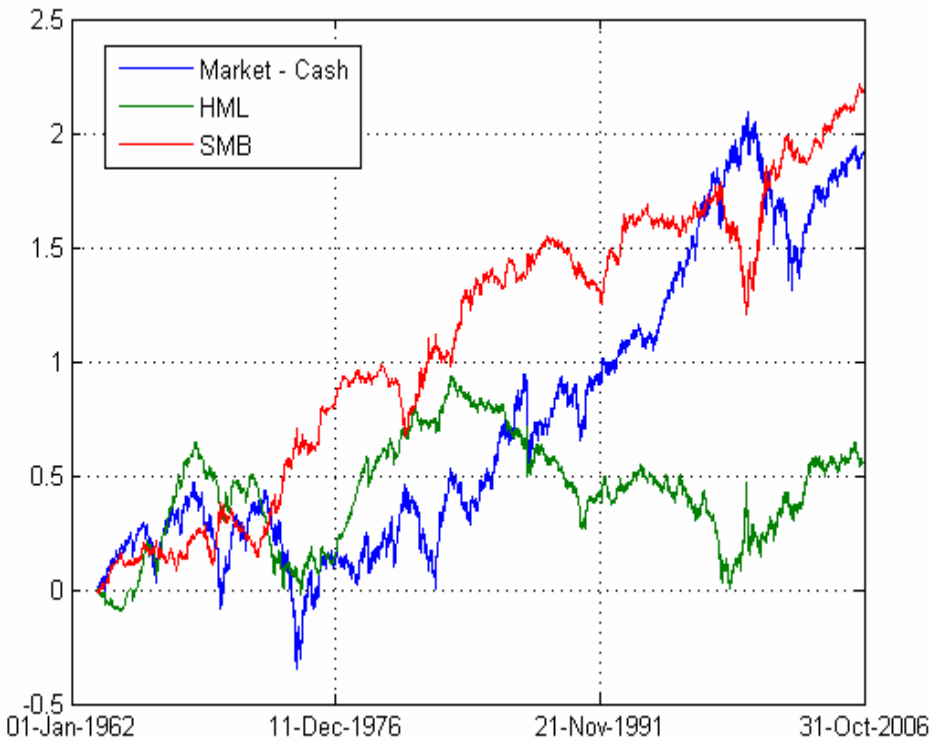

Market Cap
Bottom Third
Middle Third
Top Third
B/P Ratio
Top Third
Middle Third
Bottom Third
Average = H
Average = L
HML = H - L

SMB Factor


Market Cap
Bottom Third
Middle Third
Top Third
Top Third
Middle Third
Bottom Third
B/P Ratio
Average = S
Average = B
SMB = S - B

Fama & French Factor Returns

- Market Minus Cash
  - Growth of market over time minus risk-free rate
- High Minus Low Book/Price
  - Distressed companies tend to have high B/P ratios
  - Healthy companies tend to have negative "loadings" during bull markets
- Small Minus Large Cap
  - Small cap returns usually greater than large cap returns (growth vs value)
  - Large-cap shocks



Source: Factor returns courtesy of Kenneth French, Dartmouth University, 2006.

Estimation of the Model

- Multivariate normal regression
  - $$\begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_m \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{H}_1 \mathbf{b} \\ \vdots \\ \mathbf{H}_m \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{C} & \cdots & \mathbf{0} \\ \vdots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{C} \end{bmatrix} \right)$$
- Model
  - $\mathbf{Z}_k$ is an n-vector of observations of n series for $k = 1, \ldots, m$ samples
  - $\mathbf{H}_k$ is an n by p design matrix for sample k with n series by p parameters
  - $\mathbf{b}$ is a vector of p model parameters
  - $\mathbf{C}$ is an n by n covariance matrix for each sample
- Assumption
  - Residuals are i.i.d. across samples

Maximum Likelihood Estimation

- Method of choice is maximum likelihood (ML) estimation
- Given
  - A collection of *n*-dimensional observations $\mathbf{z}_1, \ldots, \mathbf{z}_m$
  - A collection of parameters θ to characterize the model and probability distribution of model residuals
    - θ contains **b** and distinct parameters from **C**
  - A joint probability density function $f(\mathbf{z}_1, \ldots, \mathbf{z}_m; \theta)$ for the observations given the collection of parameters
  - Log-likelihood function is
    - $L(\mathbf{z}_1, \ldots, \mathbf{z}_m; \theta) = \log f(\mathbf{z}_1, \ldots, \mathbf{z}_m; \theta)$
- ML estimate is value of θ that maximizes log-likelihood function given observations $\mathbf{z}_1, \ldots, \mathbf{z}_m$

Standard Errors

- Parameter estimates θ converge asymptotically for large samples to a multivariate normal distribution
- Standard errors are obtained from diagonal of $\mathrm{cov}(\theta)$
  - $\sigma(\theta_i) = \sqrt{\mathrm{cov}(\theta)_{ii}}$
- Fisher information matrix
  - $I(\theta) = -E[\nabla^2 L(\mathbf{z}_1, \ldots, \mathbf{z}_m; \theta)]$
- Cramér-Rao lower bound
  - $\sigma^2(\theta_i) = (I^{-1}(\theta))_{ii}$

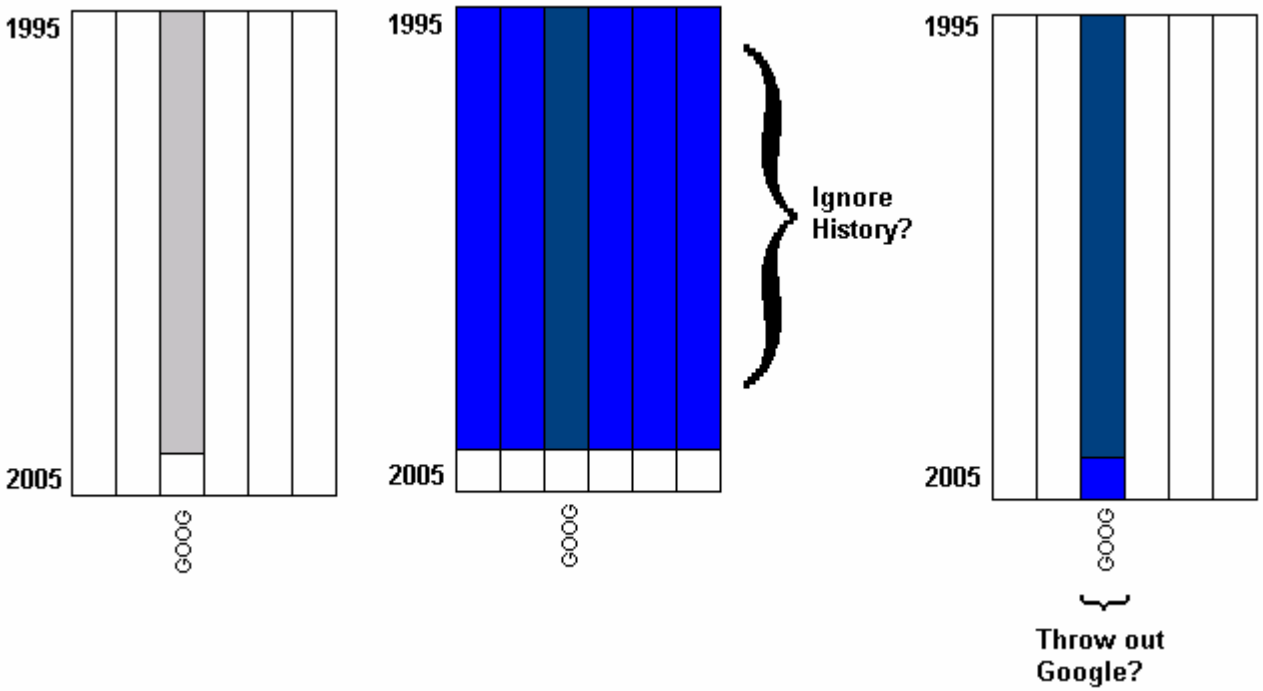
Seemingly-Unrelated Regression (SUR)

- Start with a multivariate normal regression model
- Common parameters across all series in regression model
- Select groups of series in regression model
- Duplicate parameters across groups of series
- Forms a “diagonal” structure of design matrices
  - Financial Toolbox has a function to do this
- In the Fama & French model
  - Start with 4 parameters common to all assets in universe
  - For the full model, convert to an SUR model with 4 parameters for each asset in universe

Missing Data

- Economic, financial, and market data
  - Dates never “line up”
    - Different holiday schedules for different exchanges
    - Mixed periodicities
  - Stale prices (“fair-value” pricing)
  - Delayed reporting of results
  - IPOs
  - Lack of availability
  - Missing values, data errors, non-existent values

What Should You Do?



Financial Time Series

- Financial Toolbox has enhanced functions to handle financial time series
  - Date math operations
  - Periodicity conversions
  - Merge and split time series
  - Lines up dates for data from multiple sources
    - Fills missing values with NaNs
  - Vectorized operations

Estimation with Missing Data

- Regression functions in the Financial Toolbox that handle missing data
  - Maximum likelihood estimation
  - ECM algorithm
  - Guaranteed to obtain a maximum of the log-likelihood
- Robust, stable, and accurate
- Estimation by direct numerical computation, not by simulation
  - Parameter uncertainties come from the data, not the estimation procedure
  - No need to “fill in” the data

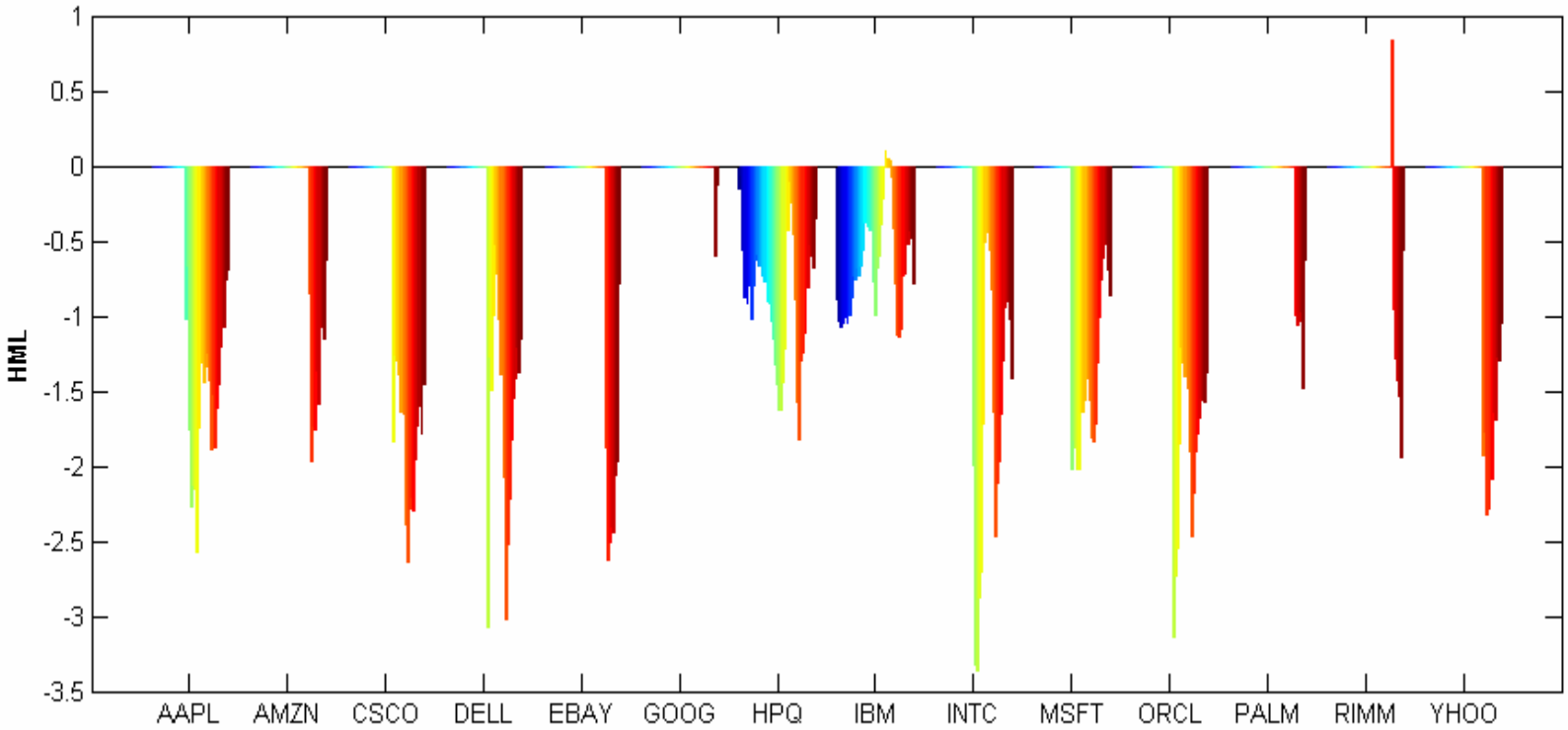
Technology Stocks

- Compare CAPM and Fama & French three-factor model
  - Analyze 14 technology stocks
    - AAPL, AMZN, CSCO, DELL, EBAY, GOOG, HPQ,
    - IBM, INTC, MSFT, ORCL, PALM, RIMM, YHOO
  - Obtain 40+ years of daily data for stocks, market, cash, and Fama & French factor returns
  - MATLAB Datafeed Toolbox to obtain stock price, dividend, and split data
  - Fama & French factor returns courtesy of Kenneth French, Dartmouth University, 2006
- Set up SUR model to estimate Alphas, Betas, Fama & French factor loadings, and residual covariance
- HPQ and IBM only companies around for entire analysis period
  - Provides an “anchor” for subsequent IPOs

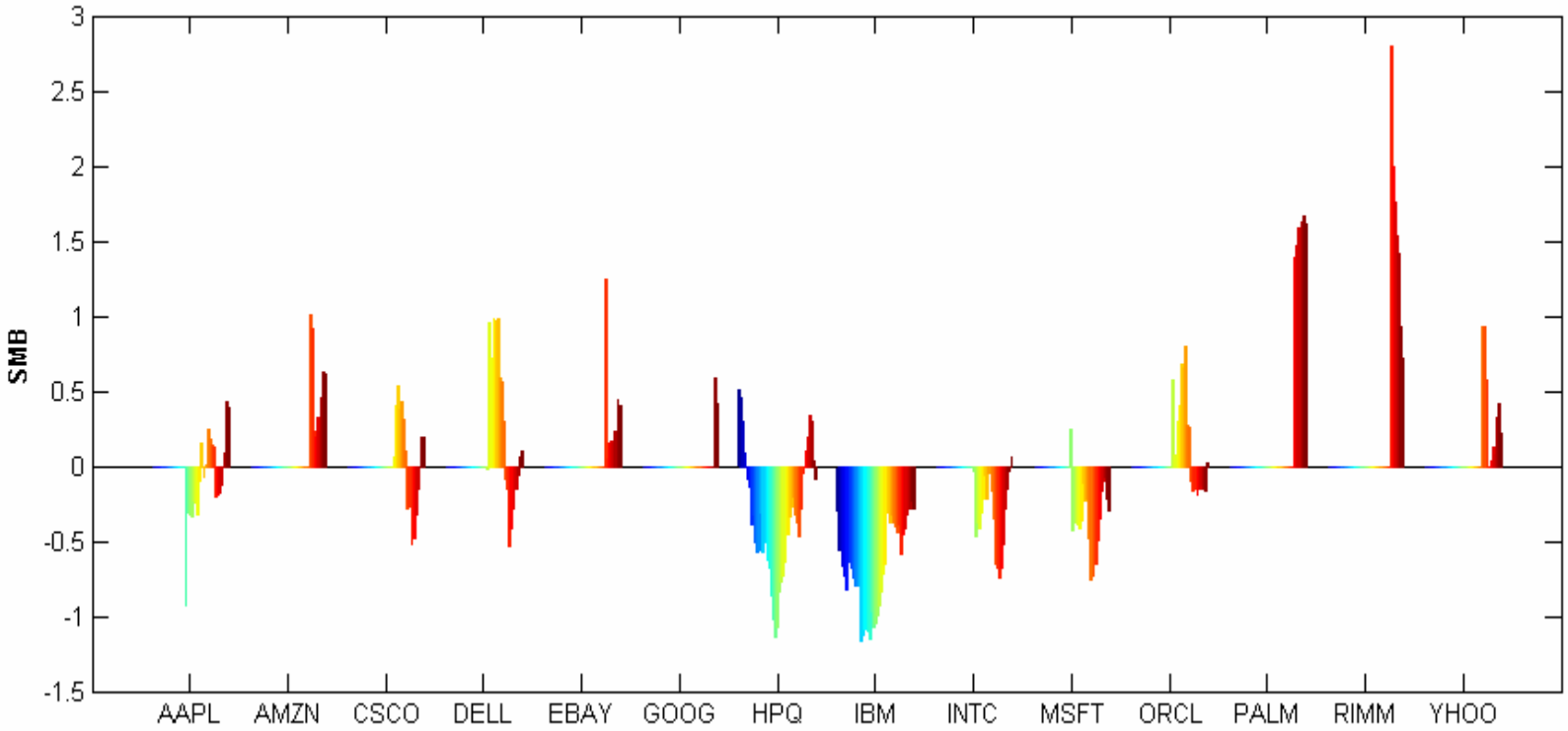
Model Comparison

- Likelihood ratio test
  - H_0: HML = 0 and SMB = 0 in Fama & French model
  - If H_0 accepted, then CAPM is sufficient to explain asset returns among this sample of assets
- Results
  - Test statistics exceed critical values by an order of magnitude
  - Complete rejection of the null hypothesis H_0
- Fama & French three-factor model explains asset returns better than the CAPM over all time periods in the analysis

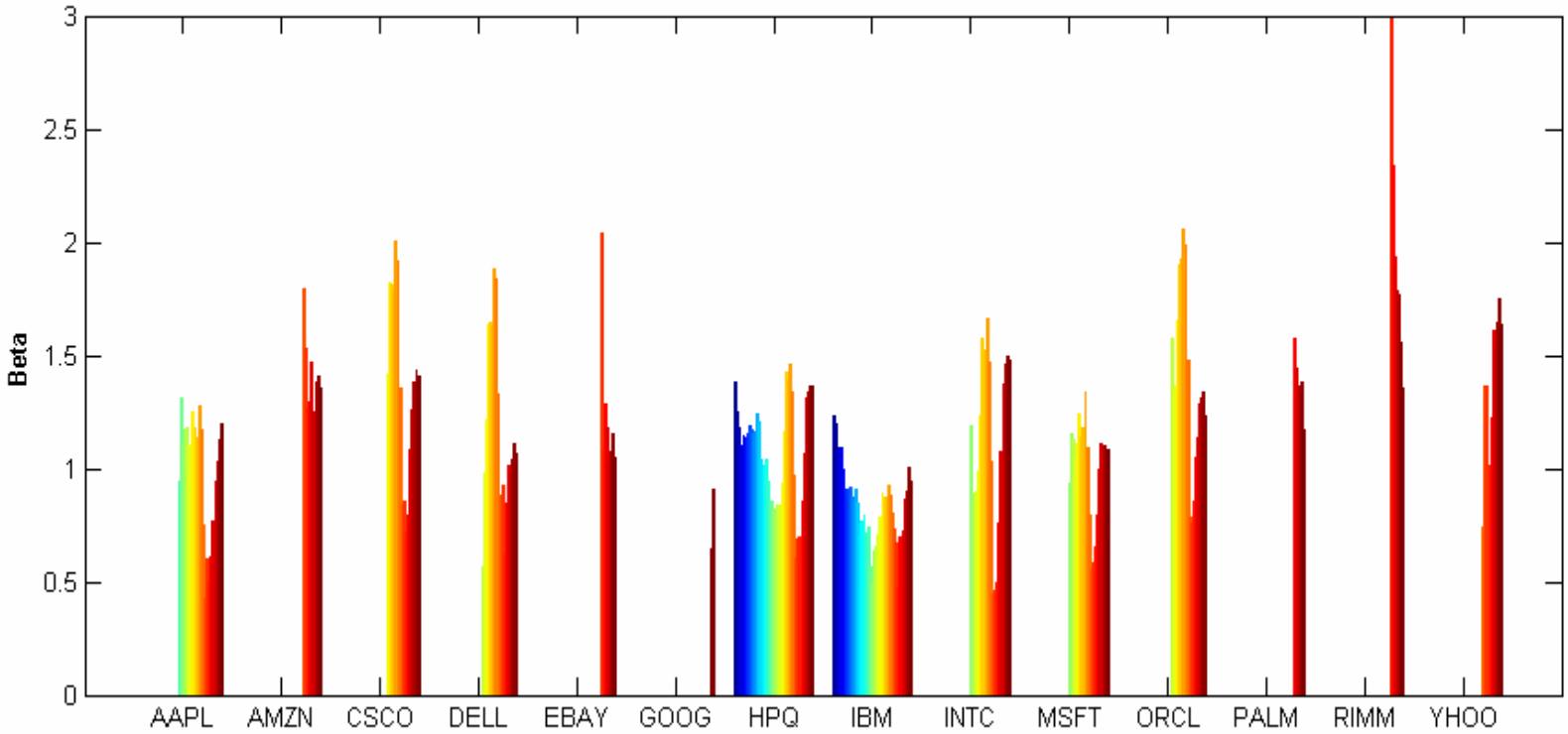
Factor Exposures - HML



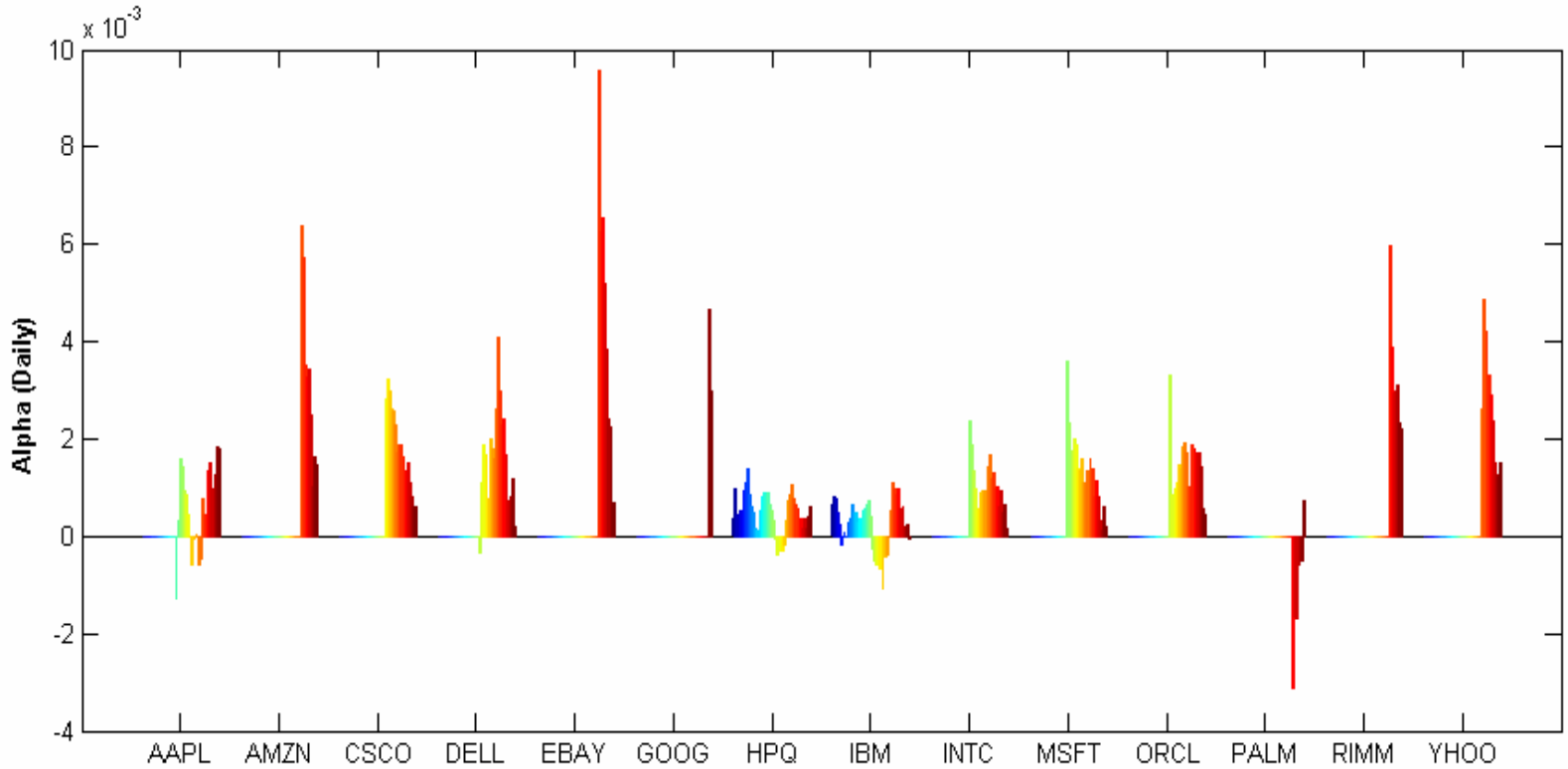
Factor Exposures - SMB



Factor Exposures - Beta



Factor Exposures - Alpha



IPOs and Market Bubbles

- Alphas and IPOs
  - Statistical tests show that significant Alphas exist shortly after IPOs
    - t-tests on individual Alphas
    - Likelihood ratio tests on collection of Alphas
  - Inconclusive, however, to “capture” Alphas through neutral hedges
  - Not an artifact of the missing data algorithms
  - Implies an unknown mechanism or even a possible “factor” that decays from an IPO onward
- Beta variations during a technology bubble
  - For a brief period, tech was uncorrelated with the broader market so lower betas
  - Possible additional factor

Discussion

- Dealing with missing data
  - Current approaches require about 3 years of data to estimate factor models
  - Missing data approach works with about 6 months of data
  - No need to get bogged down in handling NaNs
- Timely estimates
  - Can look at IPOs
  - Able to get beta estimates before anyone else
    - Estimated correct GOOG beta in 2005
- Fama & French model seems to explain most asset returns
  - Still need to find predictive models for factor returns
  - Still need to explain persistent Alphas immediately after an IPO

References

- Peter E. Caines, *Linear Stochastic Systems*, John Wiley & Sons, Inc., 1988.
- John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press, 1997.
- Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
- A. P. Dempster, N.M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, *Series B*, Vol. 39, No. 1, 1977, pp. 1-37.
- Eugene F. Fama and Kenneth R. French, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, Vol. 47, 1992, pp. 427-465.
- Eugene F. Fama and Kenneth R. French, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, Vol. 33, No. 1, 1993, pp. 3-56.
- Eugene F. Fama and Kenneth R. French, "Size and Book-to-Market Factors in Earnings and Returns," *Journal of Finance*, Vol. 50, 1995, pp. 131-155.
- Eugene F. Fama and Kenneth R. French, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, Vol. 51, No. 1, March 1996, pp. 55-84.
- Kenneth R. French, Data Library, Dartmouth, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
- William H. Greene, *Econometric Analysis*, 5th ed., Pearson Education, Inc., 2003.
- James D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.
- Robert A. Jarrow, *Finance Theory*, Prentice-Hall, Inc., 1988.
- John Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stocks," *Review of Economics and Statistics*, Vol. 14, 1965, pp. 13-37.
- Roderick J. A. Little and Donald B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed., John Wiley & Sons, Inc., 2002.
- Xiao-Li Meng and Donald B. Rubin, "Maximum Likelihood Estimation via the ECM Algorithm," *Biometrika*, Vol. 80, No. 2, 1993, pp. 267-278.
- Barr Rosenberg, "Extra-Market Components of Covariance in Security Returns," *Journal of Financial and Quantitative Analysis*, Vol. 9, No. 2, March 1974, pp. 263-274.
- Barr Rosenberg and Walt McKibben, "The Prediction of Systematic and Specific Risk in Common Stocks," *Journal of Financial and Quantitative Analysis*, Vol. 8, No. 2, March 1973, pp. 317-333.
- Stephen A. Ross, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, Vol. 13, No. 3, December 1976, pp. 341-360.
- Joe Sexton and Anders Rygh Swensen, "ECM Algorithms that Converge at the Rate of EM," *Biometrika*, Vol. 87, No. 3, 2000, pp. 651-662.
- William F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, Vol. 19, 1964, pp. 425-442.