Do Common Factors Really Explain the Cross-Section of Stock Returns?

Alejandro Lopez-Lira[1] Nikolai Roussanov[2]

[1]University of Florida

[2]The Wharton School, University of Pennsylvania and NBER

July 2022

Big Picture

- Arbitrage pricing theory (APT): Securities earn higher expected returns **only** because they are more exposed to common (i.e., undiversifiable) risk factors
- The expected excess returns of diversified portfolios that hedge against all systematic risk should be **zero** (at least in the limit)
- APT is the foundation of empirical multi-factor models for the cross-section of stock returns

Why is it hard to test APT?

- Identifying common factors
  - Maybe we haven't found the right factors?
  - Identifying all factors requires estimating very high-dimensional covariance matrices
- Expected returns
  - Unobservable

Our Contribution

We combine:

- An efficient method to measure **individual stocks' dynamic exposure to every common latent factor**
  - Singular value decomposition of the return matrix
  - No need to estimate the covariance matrix
  - No need to estimate the factors
- Out-of-sample measure of expected returns for each stock
  - Broad set of public and firm-specific signals
  - Rolling predictive regression
  - Machine learning
  - Linear Regression
- Pure-plays to hedge all systematic risk while sorting on expected returns.

Main Findings Summary

- Portfolios that hedge systematic risk perform better than similar portfolios that take more systematic risk
- Latent factors explain a considerable fraction of the time-series variation of stock returns
- But, all the latent factors carry either negligible or zero price of risk on average
    - Latent factors explain almost none of the cross-sectional variation in excess returns
- Profitable trading strategies with zero ex-ante systematic risk exposure

Related Literature

- Factor Models
  - Roll and Ross (1980), Chamberlain and Rothschild (1983), Connor and Korajczyk (1986), Connor and Korajczyk (1988), Shukla and Trzcinka (1990), Kozak, Nagel, and Santosh (2018), Pelger (2020), Giglio, Xiu, and Zhang (2021), among many others.
- Cross-sectional Return Prediction (Regression or Machine Learning)
  - Fama and MacBeth (1973), Jacobs and Levy (1988), Lewellen (2015),...
  - Gu, Kelly, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), ...
- Characteristics vs. Covariances
  - Daniel and Titman (1997), Back, Kapadia, and Ostdiek (2015), Kelly, Pruitt, and Su (2019), Kirby (2019), Lettau and Pelger (2020), Fama and French (2020), Daniel, Mota, Rottke, and Santos (2020), Cooper, Ma, Maio, and Philip (2021), Kim, Korajczyk, and Neuhierl (2021), Chen, Roussanov, and Wang (2021) among many others.

APT

$$r^e_{i,t+1} = \mu_{i,t} + \beta'_{i,t} F_{t+1} + \varepsilon_{i,t+1} \quad (1)$$

- $\mu_{i,t} = \alpha_{i,t} + \lambda'_t \beta_{i,t}$ (assuming F are mean zero)
- $E[\epsilon_{i,t+1}] = E[\epsilon_{i,t+1}\epsilon_{j,t+1}] = 0$
- $\alpha_{i,t} = 0$
- Otherwise we can build a diversified long-short portfolio with an extremely high Sharpe Ratio ("near-arbitrage")

Theory

The realized return of a long-short portfolio is:

$$r_{h,t+1} - r_{l,t+1} = \mu_{h,t} - \mu_{l,t} + (\beta_h - \beta_l)F_{t+1} + \bar{\varepsilon}_{h-l,t+1} \tag{2}$$

If $\beta_h = \beta_l$:

$$r_{h,t+1} - r_{l,t+1} = \alpha_{h,t} - \alpha_{l,t} + \bar{\varepsilon}_{h-l,t+1}. \tag{3}$$

If the long-short portfolio has N stocks and the variance of $\varepsilon_{i,t+1} \equiv \sigma^2$:

$$SR(r_{h,t+1} - r_{l,t+1}) \approx \frac{\sqrt{N}(\alpha_{h,t} - \alpha_{l,t})}{\sigma} \tag{4}$$

Strategy

Zero-cost portfolios with high $\mu_{i,t}$ and zero β_t

- We need expected returns: μ
  - Predictive Regressions
  - Machine Learning

- We need covariances: β
  - High Dimensional Covariance Estimation

Expected Returns

- We forecast returns at time $t+1$ using a subset of variables available at time t
- We model the conditional expectation as a (possibly non-linear) function of characteristics available at time t using random forest regression

$$E[r_{i,t+1}|c_{i,t}] = f(c_{i,t}) \equiv \mu_{i,t} \tag{5}$$

- Out-of-sample prediction with similar structure as rolling regressions.
  - Fit the non-linear function using the information up until time t.
  - Past 5 years to predict next month.

ML Details

Supervised Machine Learning

- Prediction Machine: Receives public information available at the time and returns a forecast.
- The best forecast available at every period is the conditional expectation.
- Machine Learning is a flexible and objective technique to approximate conditional expectations.

Conditional Betas

- We use betas with respect to the latent systematic factors to absorb as much of the undiversifiable variation as possible

- We use one-year rolling windows of daily returns to address time-varying betas

(Avoiding) Covariance Estimation

- Let R_t be the $T \times N$ matrix of (demeaned) daily excess returns
- Singular Value Decomposition: $R = USV^\top$
- $U^\top U = V^\top V = I_T$ (identity matrix), S diagonal
- $C = \frac{1}{N} R^\top R$, with size $N \times N$ is the covariance matrix, singular if $N > T$
- Eigendecomposition: $C = \frac{1}{N} V \; S^2 \; V^\top = V \Lambda V^\top$

(Avoiding) Covariance Estimation

- Eigendecomposition: $C = \frac{1}{N} V \; S^2 \; V^\top = V \Lambda V^\top$
- Principal Components: RV
- Covariance of Returns with Principal components:

$$\frac{1}{T} R^\top (US) = \frac{1}{T} VSU^\top US = V \frac{S^2}{T} = V\Lambda. \tag{6}$$

- Variance of Principal Components: $\frac{1}{T}(US)^\top US = \Lambda$
- Punchline: $\beta = V$

Betas: Identification and Time consistency

- Latent factors are only unique up to rotation and the same applies to betas
  - We normalize them so that the cross-sectional average beta at each period is positive for every factor to be sure we are tracking the same factors through time
- In any case, the null space is invariant to rotations
  - We make the portfolio betas equal to zero so it does not matter which rotation we pick
- Betas are measurable at the end of the period
- We forecast them with a 3 year rolling AR(12) panel model to obtain out-of-sample results

$$\beta_{i,k,t} = a_{k,t} + \sum_{j=1}^{12} \lambda_{k,t-j} \beta_{i,k,t-j} + u_{i,k,t}, \tag{7}$$

So far

- Proxy for conditional expected returns for every stock at every point in time
- Betas with respect to the latent factors
- We want diversified portfolios sorted on expected returns and with beta of zero

Diversified Hedged Portfolios

$$\underset{w}{\text{minimize}} \ \frac{1}{2} w' w \tag{8}$$

$$s.t. \ w' \iota = 0 \ (\text{or } \ = 1)$$

$$s.t. \ w' \mu = \mu_0$$

$$s.t. \ w' \beta_k = 0 \ (\text{or } \ = 1), \ \ k = 1, ..., K$$

$$A = [\iota, \mu, B]$$

$$a = [0, \mu_0, \mathbf{0}_{K \times 1}]' \tag{9}$$

$$w^* = A(A'A)^{-1} a \tag{10}$$

Data

- CRSP Monthly and Daily Returns
  - We omit stocks whose size falls below the 20th percentile of the NYSE following Fama and French (2008) and Kirby (2019) to avoid any concerns about liquidity.
- Compustat
- 1965-2014
- We use as predictors the variables from Freyberger, Neuhierl, and Weber (2020)
  - Standardized cross-sectionally each month

Factors

- Factors are latent
- To study them we project them on the return space using pure-plays
- The projection portfolios are zero-weight and have a beta of one with its respective factor and zero otherwise
- Very similar to long high beta, short low beta
- If we make them unit cost, the first portfolio is 95% correlated with the market

Factor Pure Plays

$$\underset{w}{\text{minimize}} \ \frac{1}{2} w' w \tag{11}$$

$$s.t. \ w'\iota = 0 \ (\text{or} \ = 1)$$

$$s.t. \ w'\beta_j = 1$$

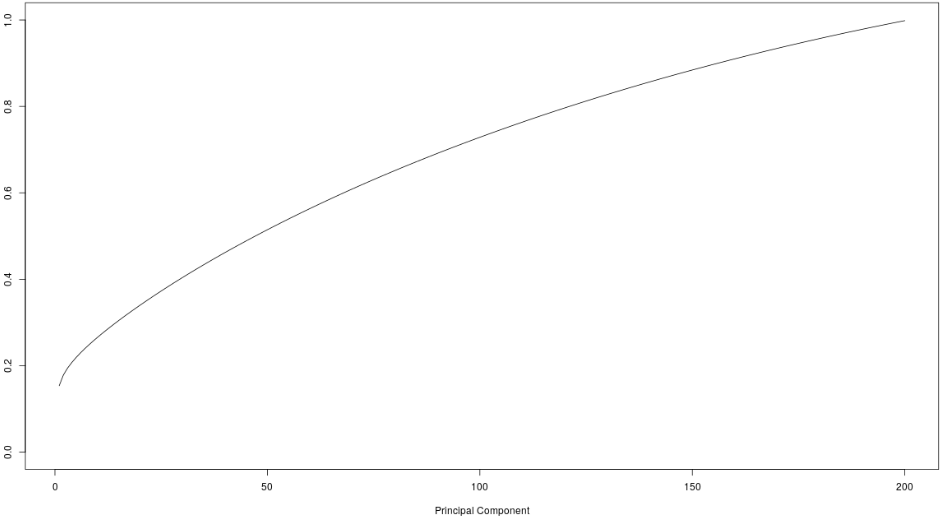
$$s.t. \ w'\beta_k = 0 \ k \neq j$$

$$A = [\iota, B]$$

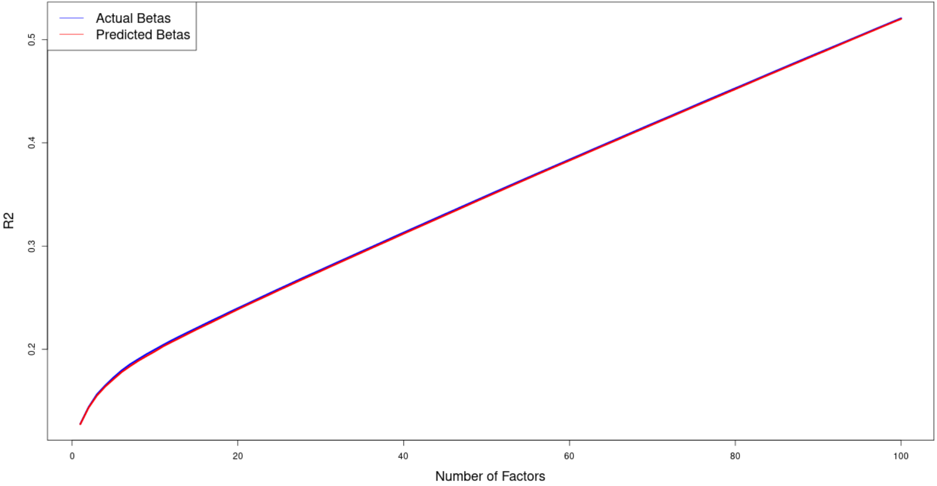
$$a = [0, \boldsymbol{e}_j']', \ \mathrm{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0) \tag{12}$$

$$w^* = A(A'A)^{-1}a \tag{13}$$

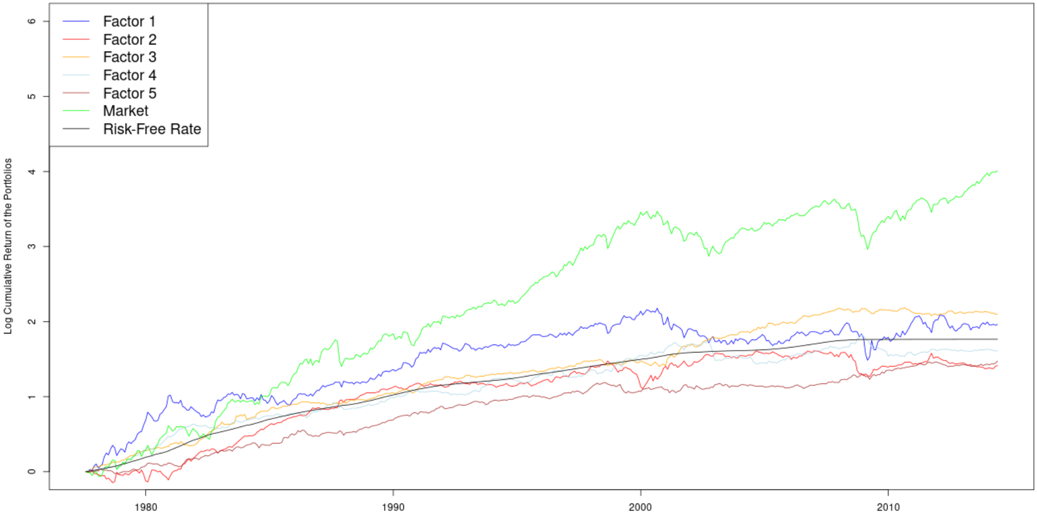
Cumulative Percentage of the Variance Explained by the Principal Components



Average Time-Series R2 of Stock Returns against Reconstructed Latent Factors



Cumulative Log Returns of the Zero-Cost Latent Factors:1974–2014



Descriptive Statistics of the Latent Factors: 1974–2014

Panel A: Value-Weighted Zero-Cost Portfolios

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Mean	-0.04	-0.01	0.07	-0.11	0.09
t-statistic	-0.23	-0.10	1.01	-1.23	1.04
Std. dev	3.50	2.43	1.64	1.96	1.97
Sharpe ratio	-0.04	-0.02	0.16	-0.19	0.16

Panel B: Value-Weighted Unit-Cost Portfolios

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Mean	0.64	0.67	0.75	0.57	0.77
t-statistic	2.60	3.29	3.41	2.52	3.33
Std. dev	5.47	4.49	4.89	5.01	5.13
Sharpe ratio	0.41	0.51	0.53	0.39	0.52
CAPM alpha	-0.24	0.21	0.11	-0.07	0.13
t-stat alpha	-0.04	1.22	0.80	-0.44	0.92

Intuition

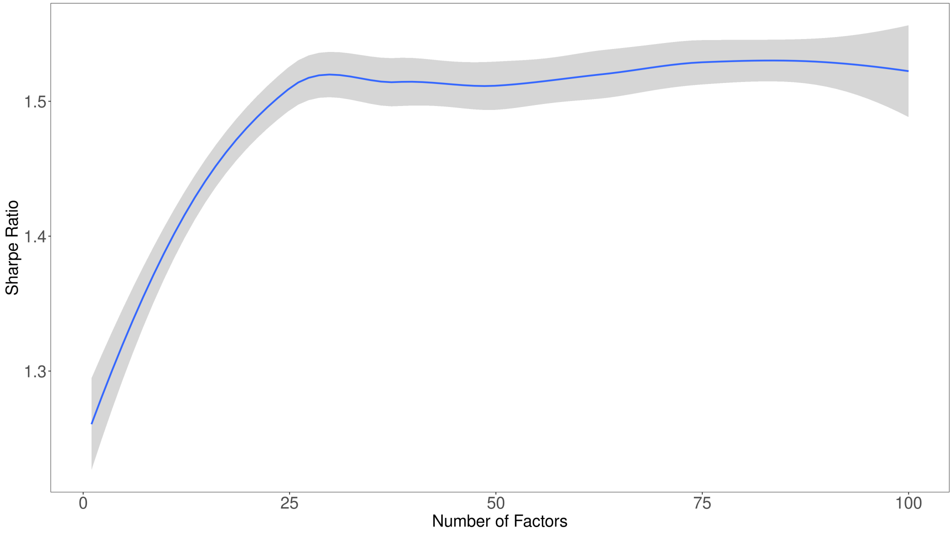
- Latent factors carry a high variance by design
- Latent factors do not carry a substantial risk premium
- Hedging portfolios against all systematic risk **reduces** their variation **without** reducing their returns
- Hedging portfolios against all systematic risk **increases** their Sharpe ratios

Descriptive Statistics of the Portfolios: 1974–2014

Panel A: 1974–2014

	Market	Long-short	Beta-neutral (hedging 50 factors)
Mean	0.52	0.72	0.74
Std. dev	4.65	3.34	1.70
Sharpe ratio	0.39	0.75	1.51

Sharpe Ratio as a function of the Number of Factors Hedged



Cumulative Log Returns: 1974–2014

Descriptive Statistics of the Portfolios: 1974–2014

Panel B: 1974–1999

	Market	Long-short	Beta-neutral
Mean	0.65	1.12	0.99
Std. dev	4.64	2.61	1.48
Sharpe ratio	0.48	1.49	2.32

Panel C: 2000–2014

	Market	Long-short	Beta-neutral
Mean	0.28	-0.06	0.25
Std. dev	4.67	4.31	1.97
Sharpe ratio	0.21	-0.04	0.44

Remarks

- Sharpe ratios almost 4 times the market's
  - Hansen–Jagannathan bound makes it extremely complicated for consumption asset pricing based models to explain them
- Latent Factors cannot explain the returns
- Reduced-form factor models?

Time Series Regressions: $Portfolio_t = \alpha + \sum \beta_i F_{i,t} + \varepsilon_t$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.75^{***}	0.76^{***}	0.53^{***}	0.69^{***}
	(4.41)	(8.85)	(2.95)	(8.44)
Mkt-RF	-0.06	-0.05^{**}	-0.04	-0.04
	(-1.20)	(-2.35)	(-0.94)	(-1.54)
SMB			0.10	-0.00
			(1.23)	(-0.04)
HML			0.01	-0.02
			(0.14)	(-0.36)
RMW			-0.19	-0.15^{*}
			(-1.36)	(-1.95)
CMA			-0.00	0.07
			(-0.00)	(0.99)
Mom			0.40^{***}	0.15^{***}
			(5.10)	(5.18)
Adj. R^2	0.01	0.02	0.28	0.19
Num. obs.	493	493	493	493

Time Series Regressions: $Portfolio_t = \alpha + \sum \beta_i F_{i,t} + \varepsilon_t$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.30	0.59***	0.37*	0.63***
	(1.43)	(6.85)	(1.85)	(6.47)
Mkt-RF	−0.02	−0.03	0.03	−0.01
	(−0.38)	(−0.94)	(0.59)	(−0.42)
R_ME	0.21***	0.06**		
	(1.79)	(0.87)		
R_IA	−0.03	0.03		
	(−0.20)	(0.49)		
R_ROE	0.13*	0.01		
	(0.77)	(0.12)		
R_EG	0.38***	0.16***		
	(2.52)	(2.17)		
SMB			0.15***	0.03
			(1.41)	(0.48)
MGMT			0.12**	0.07**
			(1.41)	(1.96)
PERF			0.32***	0.09***
			(5.65)	(4.80)
Adj. R^2	0.07	0.04	0.13	0.05

Characteristics and Covariances

If returns are driven by characteristics:

$$r^e_{i,t+1} = ac_{i,t} + \varepsilon_{i,t+1}, \tag{14}$$

or by factors, with the betas depending on characteristics:

$$r^e_{i,t+1} = c_{i,t} f_{t+1} + \varepsilon_{i,t+1}, \tag{15}$$

Models that use characteristics as covariances (like IPCA) are so effective at describing the cross-sectional variation that they can deal with both scenarios. So we can't use them to distinguish about APT.

No free lunch

But there is an empirical trade-off between explaining time-series variation and cross-sectional variation (in contrast with APT prediction).

Test		K				
assets	Statistic	1	3	4	5	6
Panel A: IPCA						
r_t	Total R^2	14.9	17.6	18.2	18.7	19
	Pred. R^2	0.36	0.43	0.43	0.70	0.70
Panel D: Principal components						
r_t	Total R^2	16.8	26.2	29.0	31.5	33.8
	Pred. R^2	<0	<0	<0	<0	<0

Total $R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - \hat{\beta}_{i,t} f_{t+1})^2}{\sum_{i,t} r_{i,t+1}^2}$, Predictive $R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - \hat{\beta}_{i,t} E[f_{t+1}])^2}{\sum_{i,t} r_{i,t+1}^2}$

Conclusion

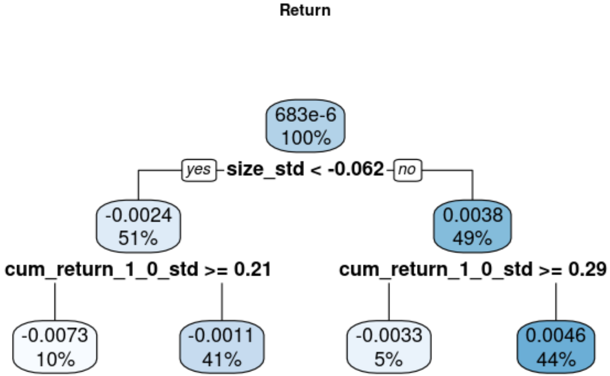
- Taking zero systematic risk should produce zero returns but creates returns at least as high as when taking systematic risk
- Sharpe Ratios are 2-4 times as large as the market
- Challenges the APT
  - as well as the consumption-based asset pricing models (Hansen–Jagannathan bound)
- Popular factor models cannot explain the returns
- But predictability seems to decline...
  - McLean and Pontiff (2016)

Appendix

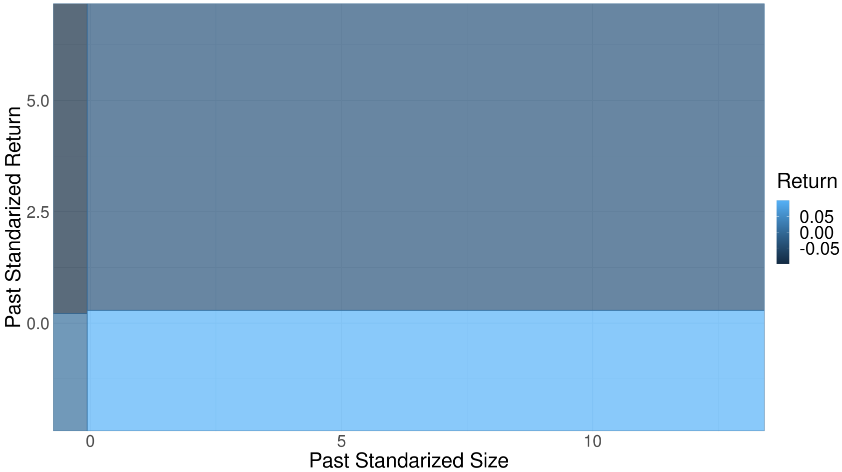
Supervised Machine Learning: Decision Trees

- Recursively split the data into non-intersecting regions one variable at a time.

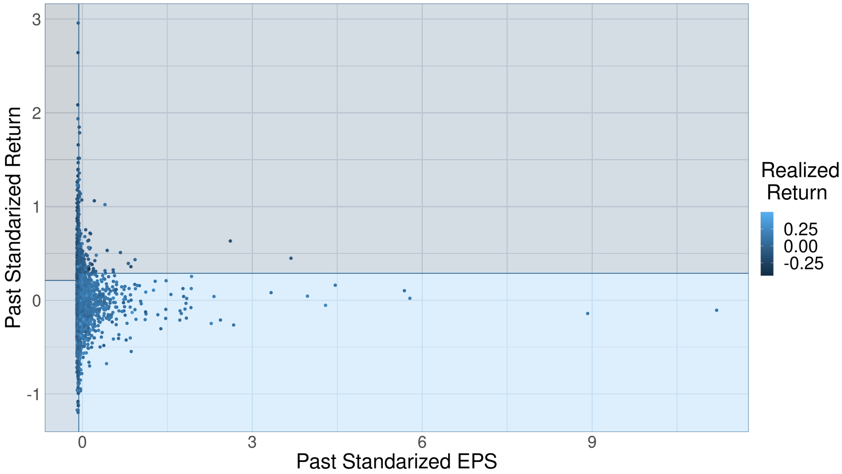
Decision Trees: Example of Depth 2



Decision Trees: Regions



Decision Trees: Regions



Supervised Machine Learning: Decision Trees

- Only one parameter: depth (length of the decision tree (2 in the previous example)).

- Tends to over-fit when depth is too large.
  - Does not give good out-of-sample predictions.

- Tends to under-fit when depth is too small.
  - Bad in-sample and out-of-sample predictions.

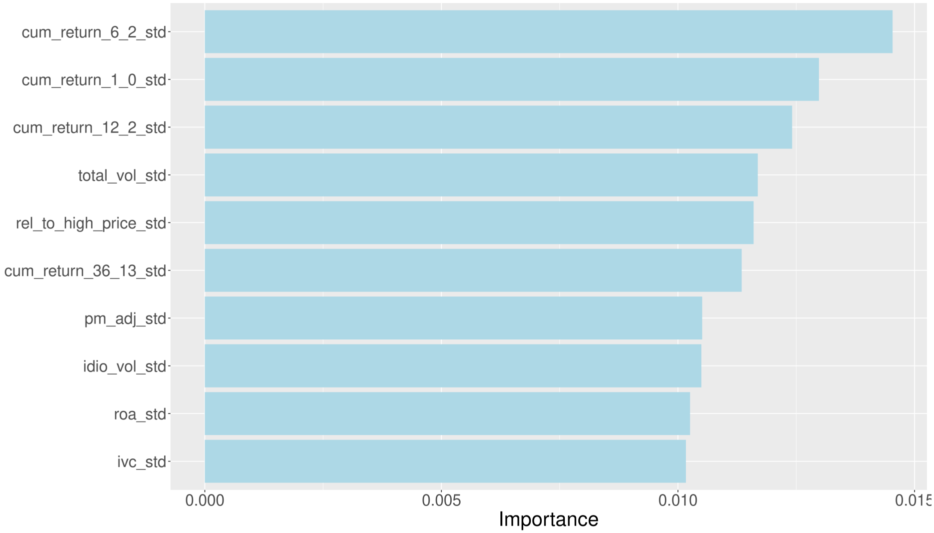
Supervised Machine Learning: Random Forest Regression

- Bootstrap of small decision trees.
- Each decision tree is trained on a different sub-sample and with different predictors.
- Flexible, non-parametric, and robust to over-fitting.
- Choose parameters in a data-driven way (cross-validation) before the forecasting period.
- Train using rolling windows.

Random Forest Regression: Interpretability

- Feature Importance
  - How much each variable decreases the mean squared error
  - Normalized to sum up to one

Feature importance of the one-quarter-ahead forecast

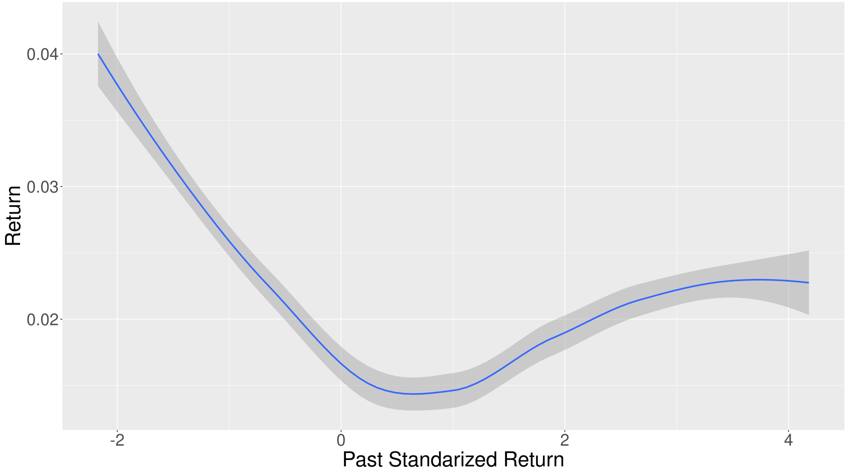


Interpretability: Impact of Variables

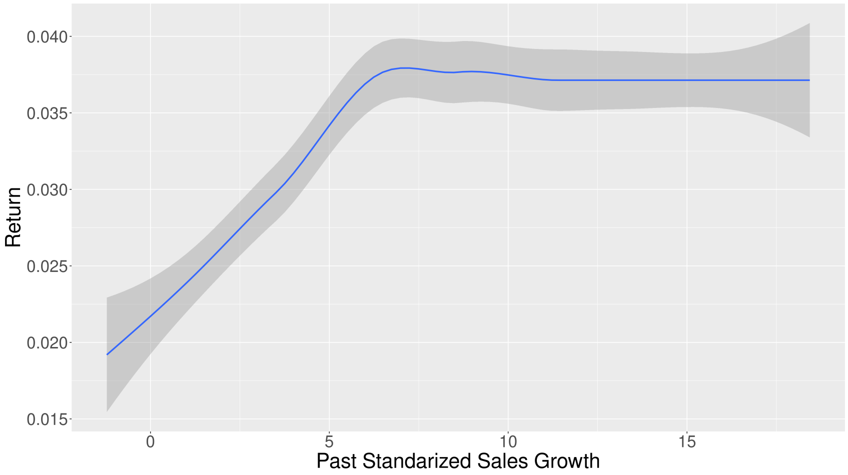
- How do features influence the predictions?

- Partial Dependence Plot
  - For given value(s) of features S what is the average marginal effect on the prediction.
  - What the model predicts on average when each data instance has a fixed value for that feature
  - $\hat{f}_{x_S}(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_S, x_C^{(i)}) \approx E_{x_C}\left[\hat{f}(x_S, x_C)\right] = \int \hat{f}(x_S, x_C) d\mathbb{P}(x_C)$

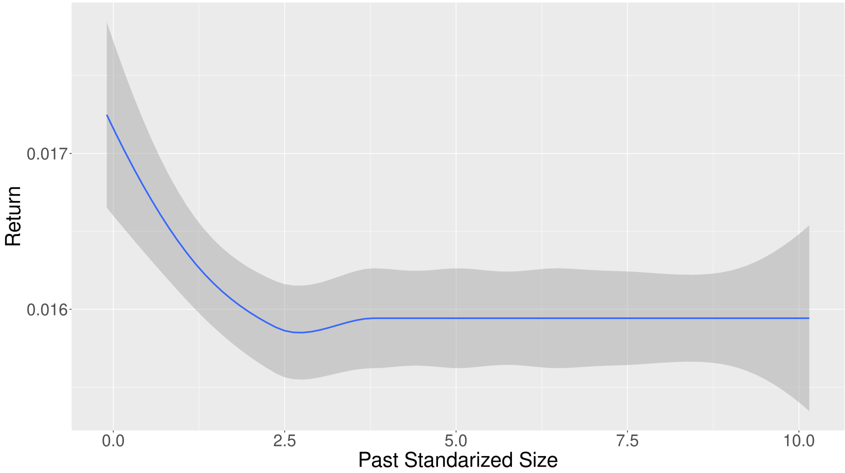
Partial Dependence Plot: Effect of Standardized Past Return on Prediction



Partial Dependence Plot: Effect of Standardized Sales on Prediction



Partial Dependence Plot: Effect of Standardized Size on Prediction



Back

Value-Weighted Time Series Regressions

$$Portfolio_t = \alpha + \sum \beta_i F_{i,t} + \varepsilon_t$$

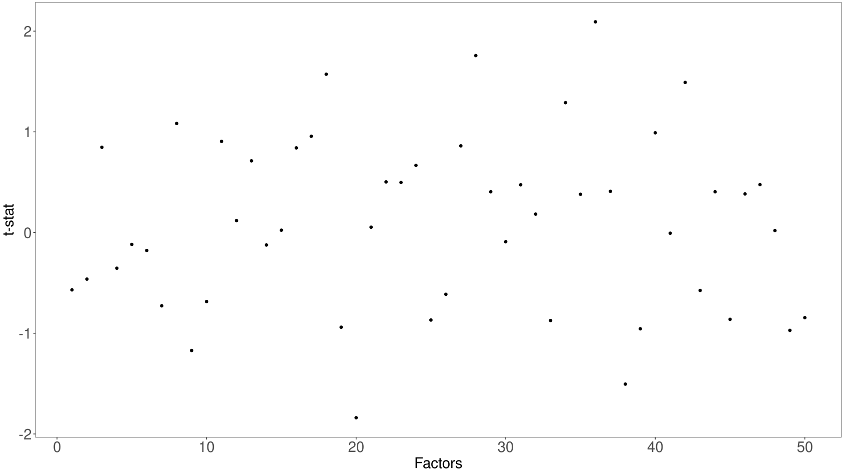
	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.60***	0.55***	0.42**	0.52***
	(3.95)	(7.77)	(2.59)	(7.17)
Mkt-RF	−0.03	−0.03	−0.04	−0.04*
	(−0.56)	(−1.81)	(−1.05)	(−2.02)
SMB			0.08	0.01
			(1.35)	(0.34)
HML			0.07	0.01
			(0.93)	(0.29)
RMW			−0.25*	−0.13
			(−1.98)	(−1.91)
CMA			−0.24	−0.06
			(−1.88)	(−1.19)
Mom			0.48***	0.14***
			(6.14)	(4.92)
Adj. R^2	−0.00	0.01	0.36	0.18
Num. obs.	493	493	493	493

Value-Weighted Time Series Regressions

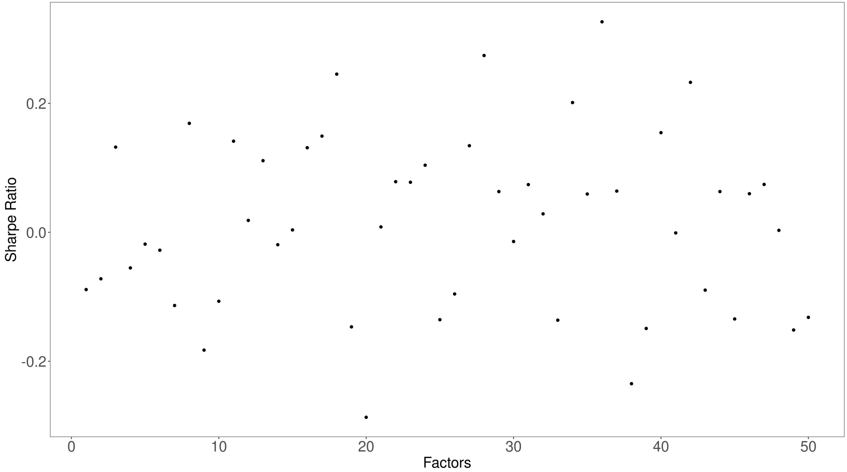
$$Portfolio_t = \alpha + \sum \beta_i F_{i,t} + \varepsilon_t$$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.28	0.45***	0.25	0.46***
	(1.33)	(5.84)	(1.49)	(5.82)
Mkt-RF	−0.02	−0.03	0.04	−0.01
	(−0.44)	(−1.30)	(1.02)	(−0.43)
R_ME	0.18	0.05		
	(1.71)	(1.06)		
R_IA	−0.22	−0.07		
	(−0.96)	(−0.92)		
R_ROE	0.13	−0.01		
	(0.75)	(−0.09)		
R_EG	0.35*	0.14*		
	(2.19)	(1.98)		
SMB			0.13	0.03
			(1.29)	(0.64)
MGMT			0.04	0.02
			(0.41)	(0.68)
PERF			0.37***	0.09***
			(5.21)	(3.66)
Adj. R^2	0.05	0.03	0.16	0.05
Num. obs.	493	493	493	493

t-statistics of Latent Factors



Sharpe Ratios of Latent Factors



Time-Series Regression on Latent Factors

$$Portfolio_t = \alpha + \sum \beta_i F_{i,t} + \varepsilon_t$$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
(Intercept)	0.71^{***}	0.73^{***}	0.74^{***}	0.73^{***}
	(4.16)	(8.08)	(5.05)	(9.30)
# Factors in Controls	10	50	10	50
Adj. R^2	0.09	0.05	0.13	0.11
Num. obs.	493	493	493	493

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$