Deep conditional portfolio sorts: The relation between past and future stock returns

Benjamin Moritz[1] Tom Zimmermann[2]

[1]Ludwig Maximilian University Munich

[2]Harvard University

December 1, 2014

INTRODUCTION

Motivation

Consider an investor who wants to forecast next period's stock returns. Challenging problem:

- Number of candidate predictor variables is large

Motivation

Consider an investor who wants to forecast next period's stock returns. Challenging problem:

- Number of candidate predictor variables is large
- Even more after accounting for potential non-linearities or interactions

Motivation

Consider an investor who wants to forecast next period's stock returns. Challenging problem:

- Number of candidate predictor variables is large
- Even more after accounting for potential non-linearities or interactions
- Many signals highly correlated

Motivation

Consider an investor who wants to forecast next period's stock returns. Challenging problem:

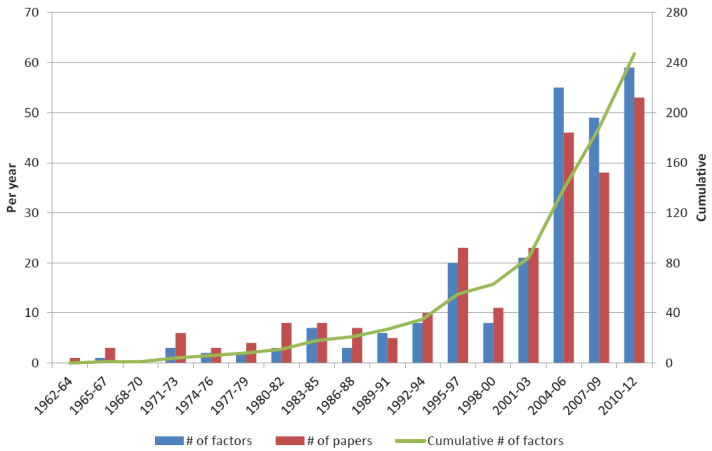
- Number of candidate predictor variables is large
- Even more after accounting for potential non-linearities or interactions
- Many signals highly correlated
- Number of observations relatively small

Motivation

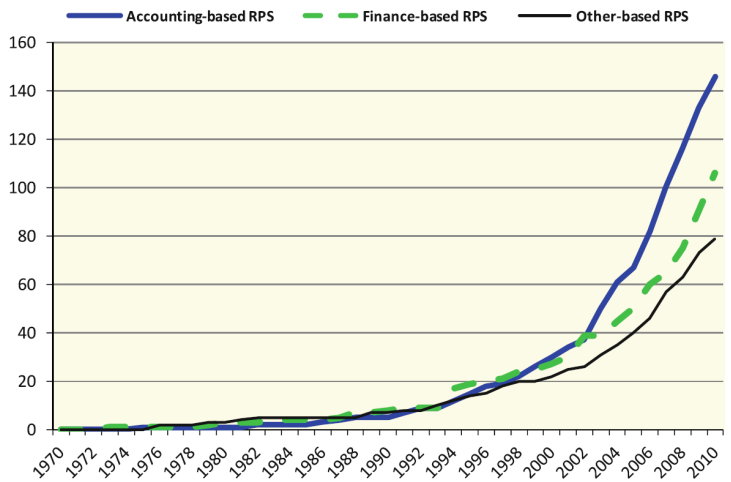
Consider an investor who wants to forecast next period's stock returns. Challenging problem:

- Number of candidate predictor variables is large
- Even more after accounting for potential non-linearities or interactions
- Many signals highly correlated
- Number of observations relatively small
- Concern about overfitting and bad out-of-sample predictions

Motivation: Ever-expanding set of anomalies



(a) Harvey et al (2013)



(b) Green et al (2014)

Cochrane (AFA Presidential Adress, 2011):

We are going to have to repeat Fama and Frenchs anomaly digestion, but with many more dimensions. *We have a lot of questions to answer:*

First, which characteristics really provide independent information about average returns? Which are subsumed by others?

Second, does each new anomaly variable also correspond to a new factor formed on those same anomalies?

Third, how many of these new factors are really important?

Cochrane (three years later):

Why dont I have an answer to "what are, really the important sorting variables?" and "how many factors do we really need to account for the known sorting variables?". [...]The technique of sorting into 1-10 portfolios and running time-series regressions on the factors [...] just is not up to the task of handling 5 or more variables. ***We need technical innovation as well as time to go look at the data.***

Contribution

1. Adopt techniques from machine learning literature to deal with these questions

Contribution

1. Adopt techniques from machine learning literature to deal with these questions
    - Desirable properties: Works well

Contribution

1. Adopt techniques from machine learning literature to deal with these questions
   - Desirable properties: Works well
     - in the presence of many variables

Contribution

1. Adopt techniques from machine learning literature to deal with these questions
    - Desirable properties: Works well
        - in the presence of many variables
        - in the presence of interactions

Contribution

1. Adopt techniques from machine learning literature to deal with these questions
   - Desirable properties: Works well
     - in the presence of many variables
     - in the presence of interactions
     - out-of-sample

Contribution

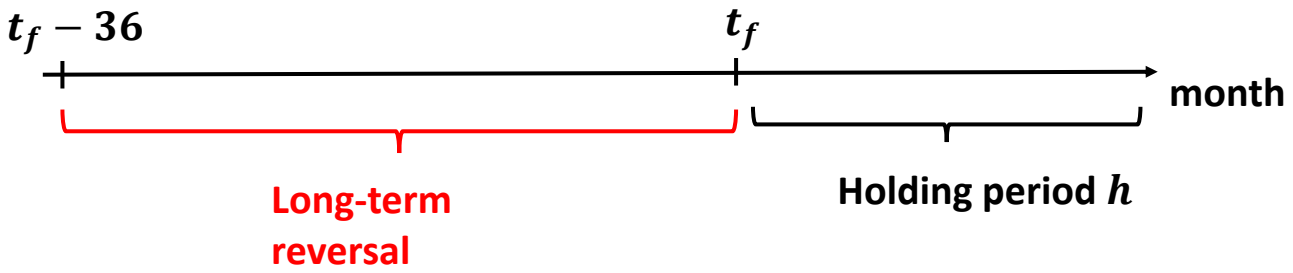
1. Adopt techniques from machine learning literature to deal with these questions
   - Desirable properties: Works well
     - in the presence of many variables
     - in the presence of interactions
     - out-of-sample
2. Tailor to an asset pricing application

Contribution

1. Adopt techniques from machine learning literature to deal with these questions
   - Desirable properties: Works well
     - in the presence of many variables
     - in the presence of interactions
     - out-of-sample
2. Tailor to an asset pricing application
3. Apply to uncover the relationship between past and future returns in US stock market

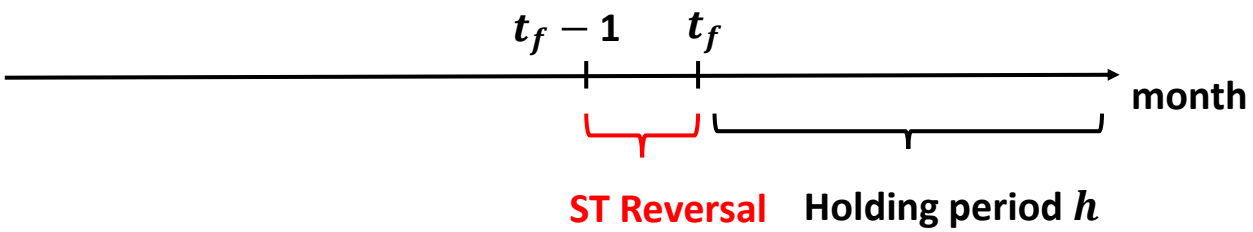
Running Example: Relation between past and future returns

Stylized facts:



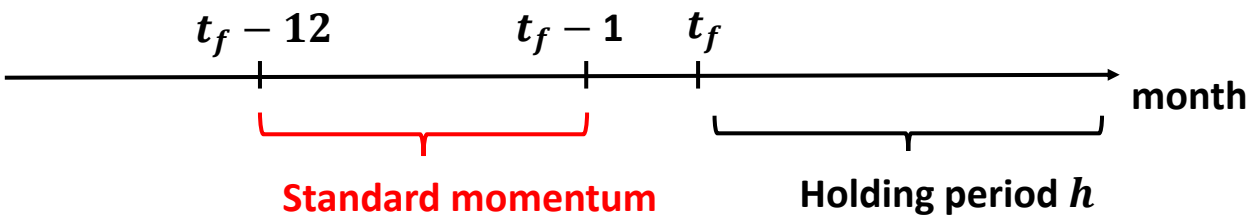
Running Example: Relation between past and future returns

Stylized facts:



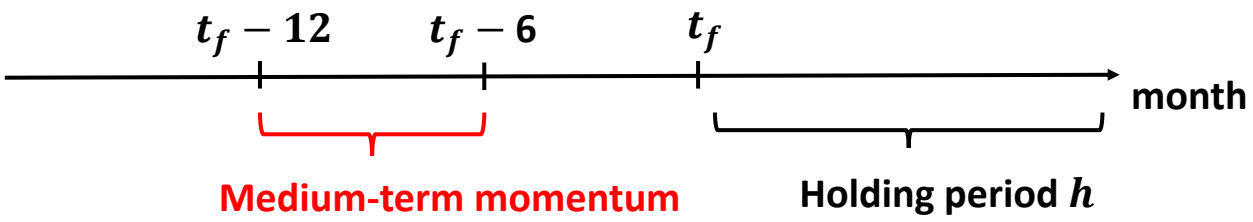
Running Example: Relation between past and future returns

Stylized facts:



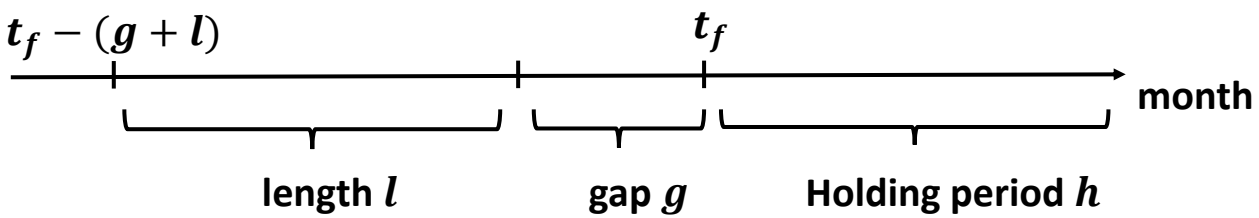
Running Example: Relation between past and future returns

Stylized facts:



Running Example: Relation between past and future returns

- In general, a plethora of past return strategies could be constructed.
- Is there a systematic way to identify the relevant predictors from an ignorant search?



Main empirical results

Predicting future returns based on past returns

1. Using more complex technique significantly improves prediction

Main empirical results

Predicting future returns based on past returns

1. Using more complex technique significantly improves prediction
   - Average monthly return of more than 2%

Main empirical results

Predicting future returns based on past returns

1. Using more complex technique significantly improves prediction
   - Average monthly return of more than 2%
   - Sharpe ratio and information ratios almost twice as high as in a Fama-MacBeth setup

Main empirical results

Predicting future returns based on past returns

1. Using more complex technique significantly improves prediction
   - Average monthly return of more than 2%
   - Sharpe ratio and information ratios almost twice as high as in a Fama-MacBeth setup
2. The relation between past and future returns seems more complex than can be captured by a linear model

Main empirical results

Predicting future returns based on past returns

1. Using more complex technique significantly improves prediction
   - Average monthly return of more than 2%
   - Sharpe ratio and information ratios almost twice as high as in a Fama-MacBeth setup
2. The relation between past and future returns seems more complex than can be captured by a linear model
   - More recent past returns are more important than more distant ones

Main empirical results

Predicting future returns based on past returns

1. Using more complex technique significantly improves prediction
   - Average monthly return of more than 2%
   - Sharpe ratio and information ratios almost twice as high as in a Fama-MacBeth setup
2. The relation between past and future returns seems more complex than can be captured by a linear model
   - More recent past returns are more important than more distant ones
   - Returns at different horizons interact in non-trivial ways

Outline

1. INTRODUCTION
2. SETUP
3. DEEP CONDITIONAL PORTFOLIO SORTS
4. PERFORMANCE RESULTS
5. EVALUATING THE MECHANISM
6. ROBUSTNESS
7. CONCLUSION

SETUP

Investor problem

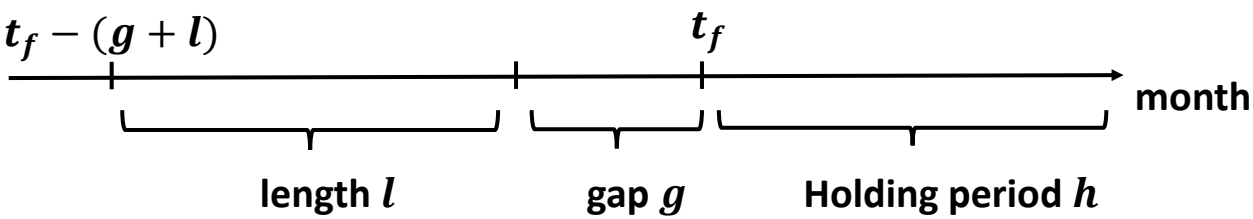
- Use past return information available in month t_f to forecast stock returns in month $t + 1$

Investor problem

- Use past return information available in month t_f to forecast stock returns in month $t + 1$
- Consider decile rankings of all one-month returns over the previous two years as candidate predictor variables

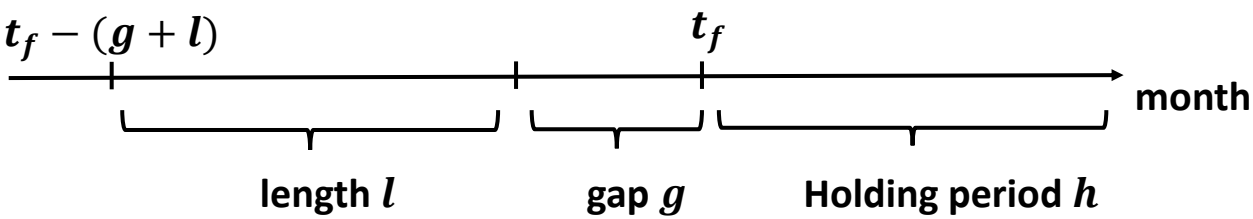
Investor problem

- Use past return information available in month t_f to forecast stock returns in month $t + 1$
- Consider decile rankings of all one-month returns over the previous two years as candidate predictor variables
- E.g.
  - $R_{i,t_f}(g, 1)$: Decile rank of firm i, g months before month t_f



Investor problem

- Use past return information available in month t_f to forecast stock returns in month $t + 1$
- Consider decile rankings of all one-month returns over the previous two years as candidate predictor variables
- E.g.
  - $R_{i,t_f}(g, 1)$: Decile rank of firm i, g months before month t_f
  - $R_{i,t_f}(0, 1) = 1$: Firm i in lowest decile of returns in previous month
  - $R_{i,t_f}(1, 1) = 10$: Firm i in highest decile of returns in second-to-last month



Investor problem

- General model of expected returns

$$E_t[r_{i,t+1}|R_{i,t}(0,1),\ldots,R_{i,t}(24,1)] = f_t(R_{i,t}(0,1),\ldots,R_{i,t}(24,1))$$

Investor problem

- General model of expected returns

$$E_t[r_{i,t+1}|R_{i,t}(0,1),\ldots,R_{i,t}(24,1)] = f_t(R_{i,t}(0,1),\ldots,R_{i,t}(24,1))$$

- Next: Discuss several approximations to this function

Investor problem

- General model of expected returns

$$E_t[r_{i,t+1}|R_{i,t}(0,1),\ldots,R_{i,t}(24,1)] = f_t(R_{i,t}(0,1),\ldots,R_{i,t}(24,1))$$

- Next: Discuss several approximations to this function
- Investment strategy will always be: Go long high predictions, go short low predictions (Lewellen (2013), Green et al (2014))

Investor problem: Modeling devices

1. Portfolio sort (single variable)

Investor problem: Modeling devices

1. Portfolio sort (single variable)
   - Can be thought of as nonparametric regression

$$r_{i,t+1} = E_t[r_{i,t+1}|R_{i,t}(0,1)]$$

Investor problem: Modeling devices

Portfolio sort (single variable)

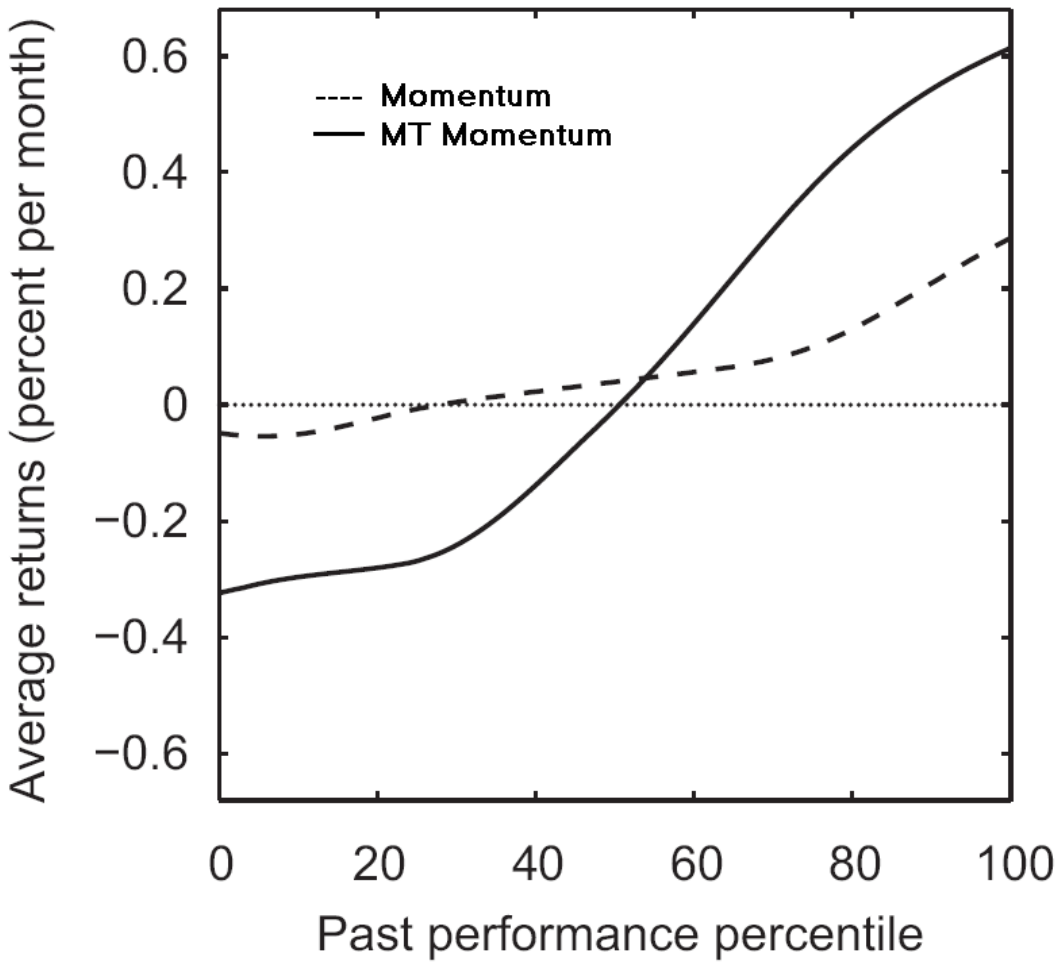


Figure: Example from Novy-Marx (2012)

Investor problem: Modeling devices

1. Portfolio sort (single variable)
   - Can be thought of as nonparametric regression

$$r_{i,t+1} = E_t[r_{i,t+1}|R_{i,t}(0,1)]$$

Investor problem: Modeling devices

1. Portfolio sort (single variable)
   - Can be thought of as nonparametric regression

$$r_{i,t+1} = E_t[r_{i,t+1}|R_{i,t}(0,1)]$$

   - In practice: Condition only on past information

Investor problem: Modeling devices

1. Portfolio sort (single variable)
   - Can be thought of as nonparametric regression

$$r_{i,t+1} = E_t[r_{i,t+1}|R_{i,t}(0,1)]$$

   - In practice: Condition only on past information
   - Pick sorting variable that would have produced the highest Sharpe ratio if it had been used over the previous m months

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
   - Use linear model

$$r_{i,t+1} = a + \sum_{g=0}^{24} \beta_g^t R_{i,t}(g,1) + \epsilon_{i,t}$$

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
   - Use linear model

$$r_{i,t+1} = a + \sum_{g=0}^{24} \beta_g^t R_{i,t}(g,1) + \epsilon_{i,t}$$

   - Make predictions based on coefficients estimated and averaged over the past m months ▸ Details

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
3. Fama-MacBeth regression (with two-way interactions)

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
3. Fama-MacBeth regression (with two-way interactions)
   - Analogous to before, but add two-way interactions between all variables ▸ Details

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
3. Fama-MacBeth regression (with two-way interactions)

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
3. Fama-MacBeth regression (with two-way interactions)
4. Deep conditional portfolio sorts

Investor problem: Modeling devices

1. Portfolio sort (single variable)
2. Fama-MacBeth regression (levels only)
3. Fama-MacBeth regression (with two-way interactions)
4. Deep conditional portfolio sorts
   - Why stop at two-way interactions?
   - Can we estimate the prediction equation more flexibly?
   - Adopt a machine learning approach that can deal with these questions $\rightarrow$ next section

DEEP CONDITIONAL PORTFOLIO SORTS

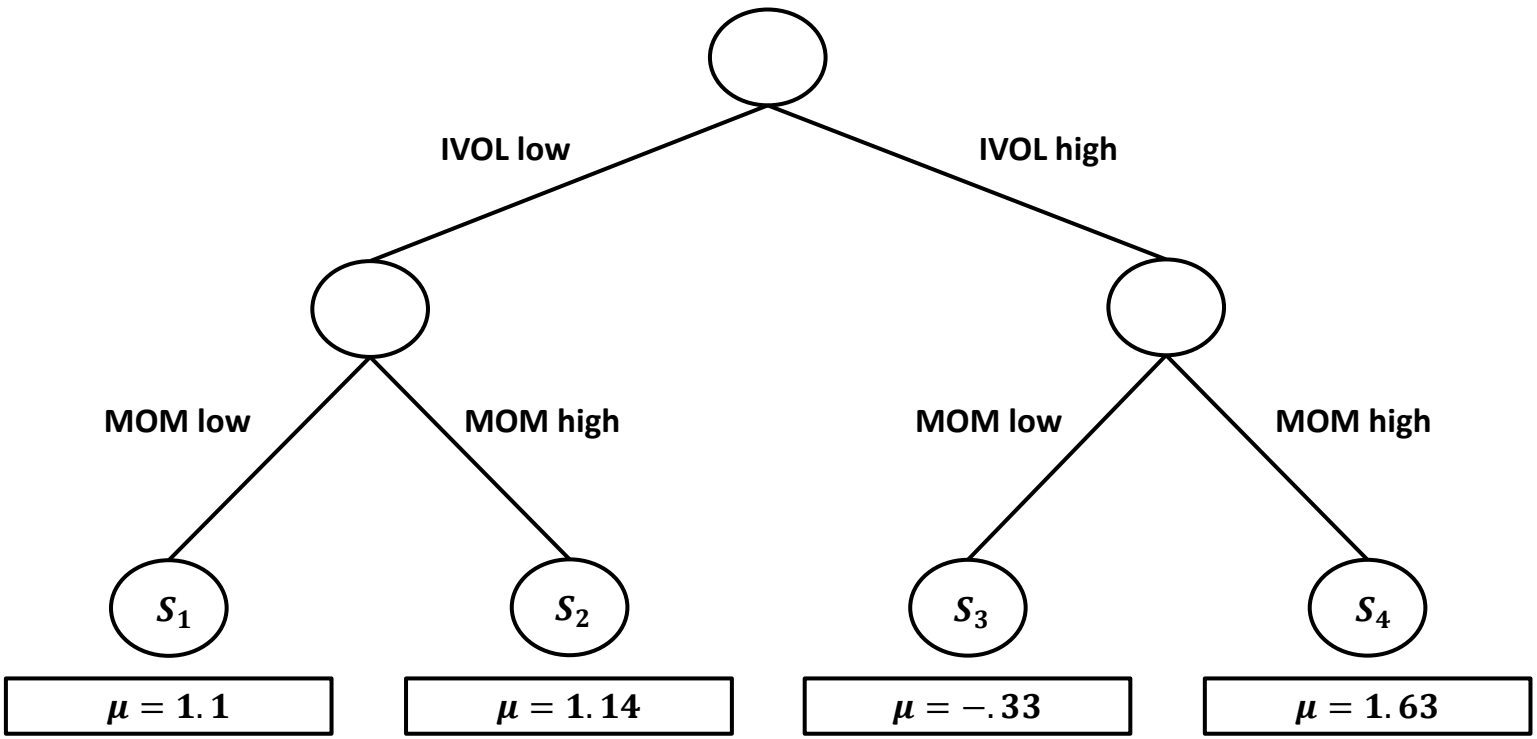
Deep conditional portfolio sort

Two steps

1. Extend the method of conditional portfolio sorts
2. Use model averaging to improve estimates

Conditional portfolio sort: Example

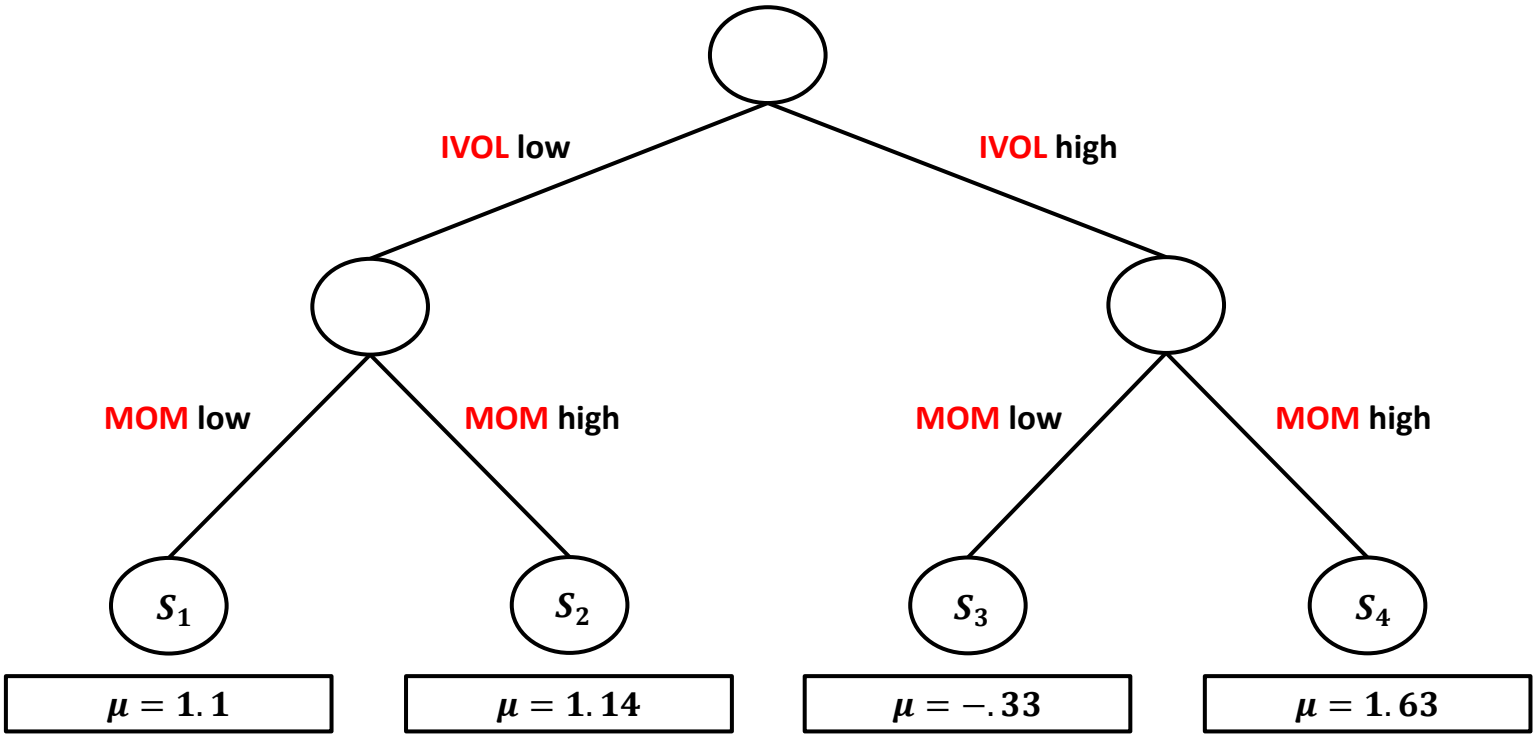
From Bandarchuk and Hilscher (2012)



Conditional portfolio sort: Example

From Bandarchuk and Hilscher (2012)

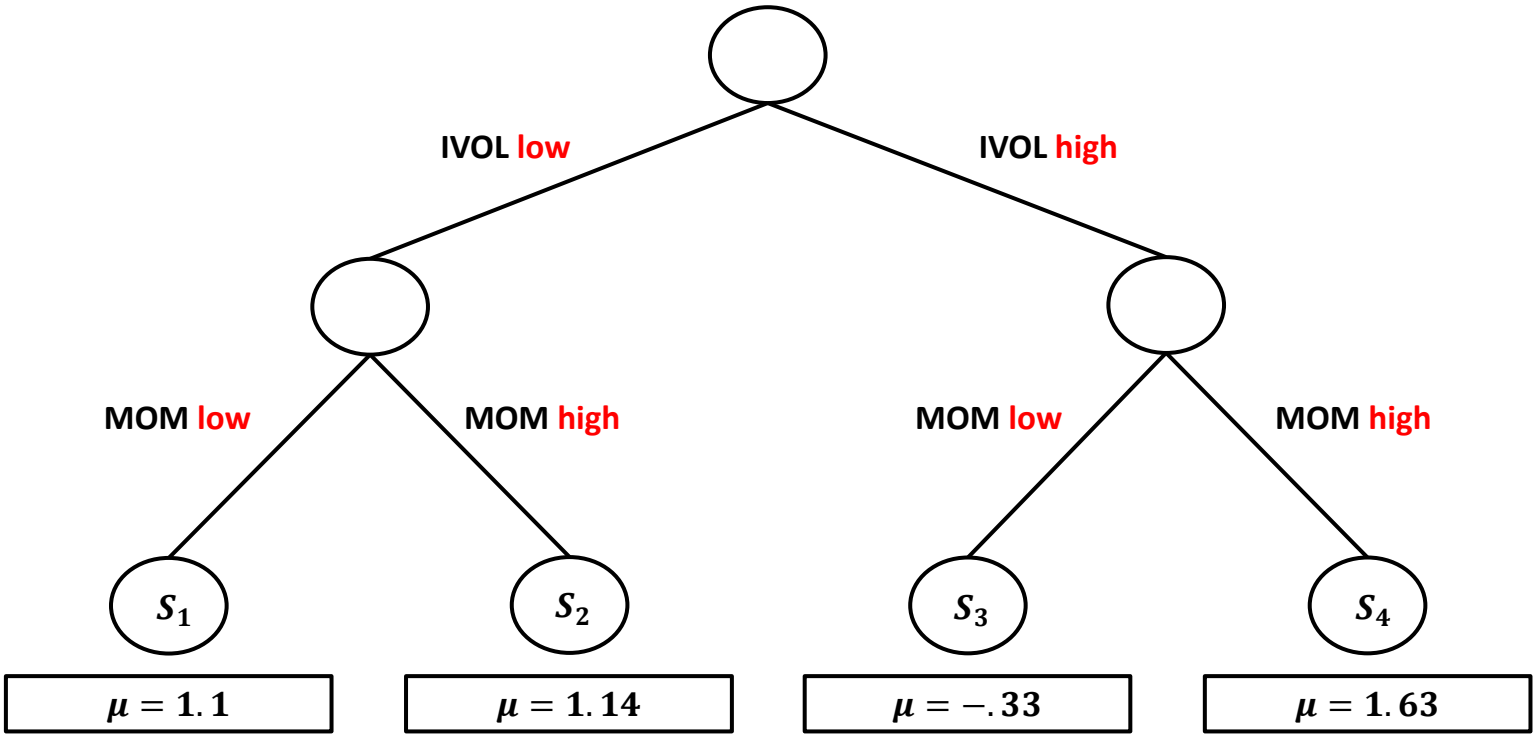
Sorting variables selected ex-ante



Conditional portfolio sort: Example

From Bandarchuk and Hilscher (2012)

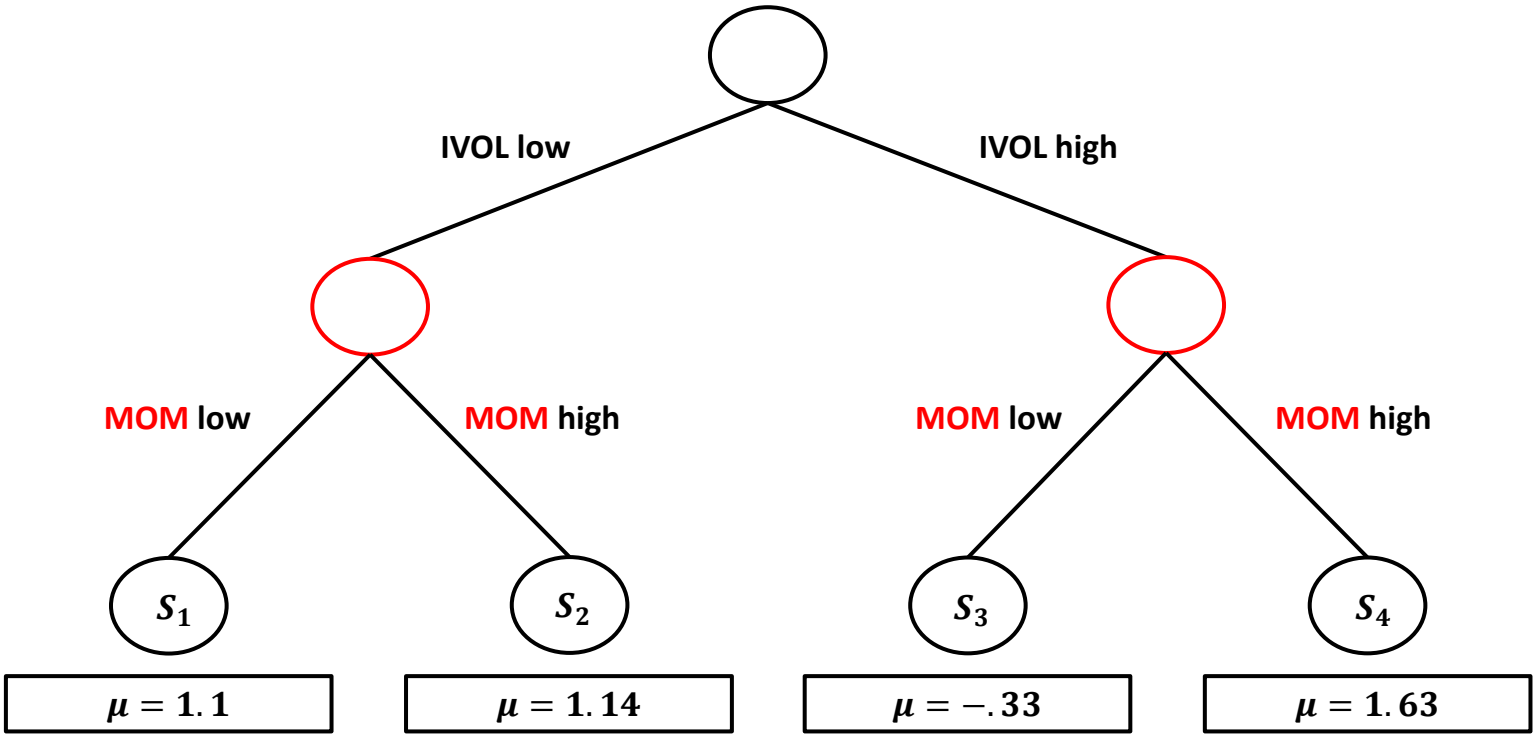
Sorting values selected ex-ante



Conditional portfolio sort: Example

From Bandarchuk and Hilscher (2012)

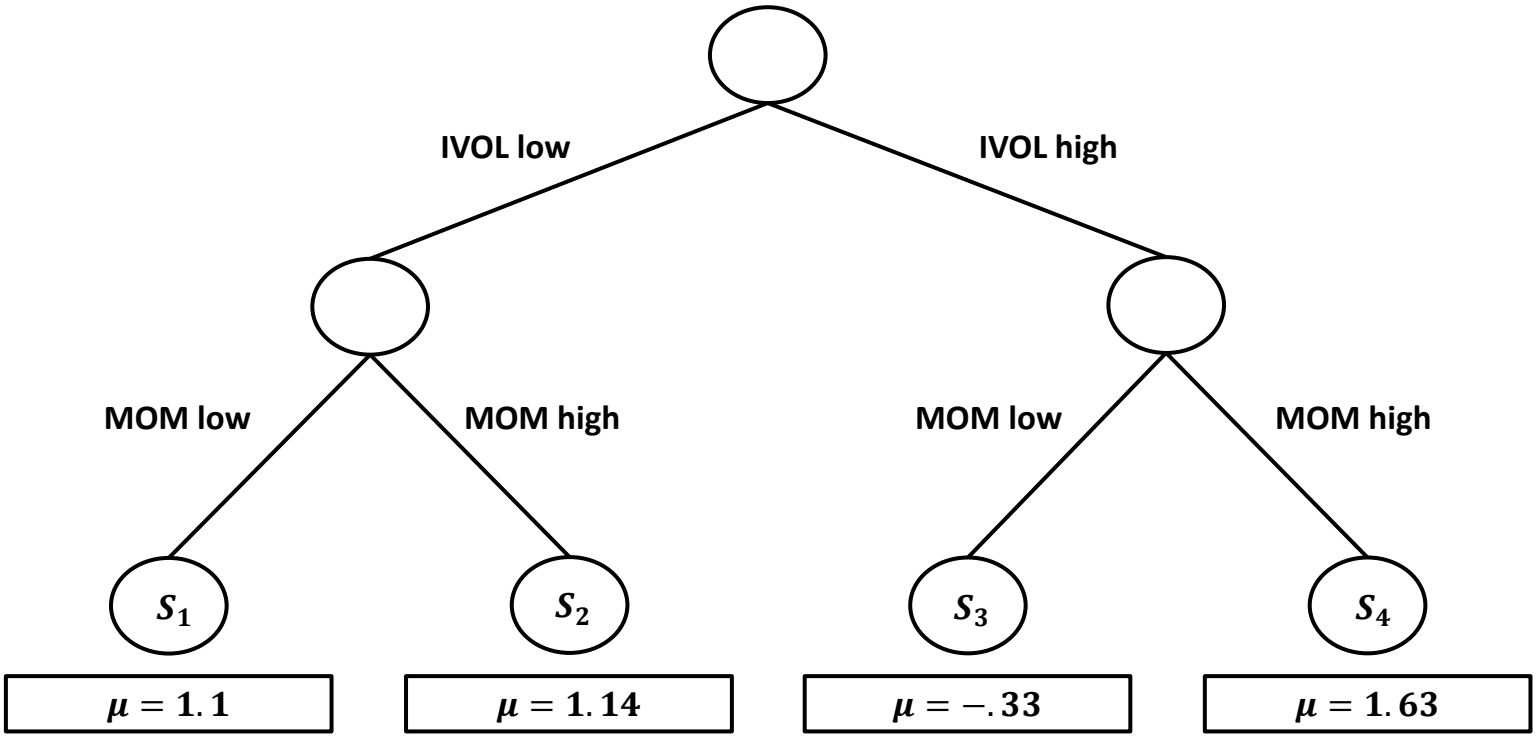
Break on the same variable in all branches



Conditional portfolio sort: Example

From Bandarchuk and Hilscher (2012)

Stop after two levels



Conditional portfolio sort

- This paper estimates

Conditional portfolio sort

- This paper estimates
  - the optimal sorting variables

Conditional portfolio sort

- This paper estimates
  - the optimal sorting variables
  - the optimal sorting values

Conditional portfolio sort

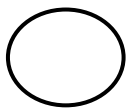
- This paper estimates
  - the optimal sorting variables
  - the optimal sorting values
  - the conditional portfolio sort for deeper levels

Conditional portfolio sort

- This paper estimates
  - the optimal sorting variables
  - the optimal sorting values
  - the conditional portfolio sort for deeper levels
- Analogous to the decision-tree literature in machine learning

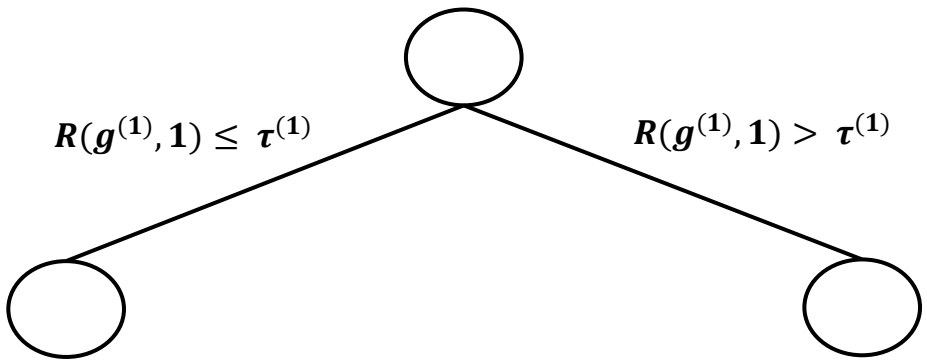
Deep conditional portfolio sort

Start with all data and find the optimal sorting variable $R(g^{(1)}, 1)$ and the associated threshold $\tau^{(1)}$



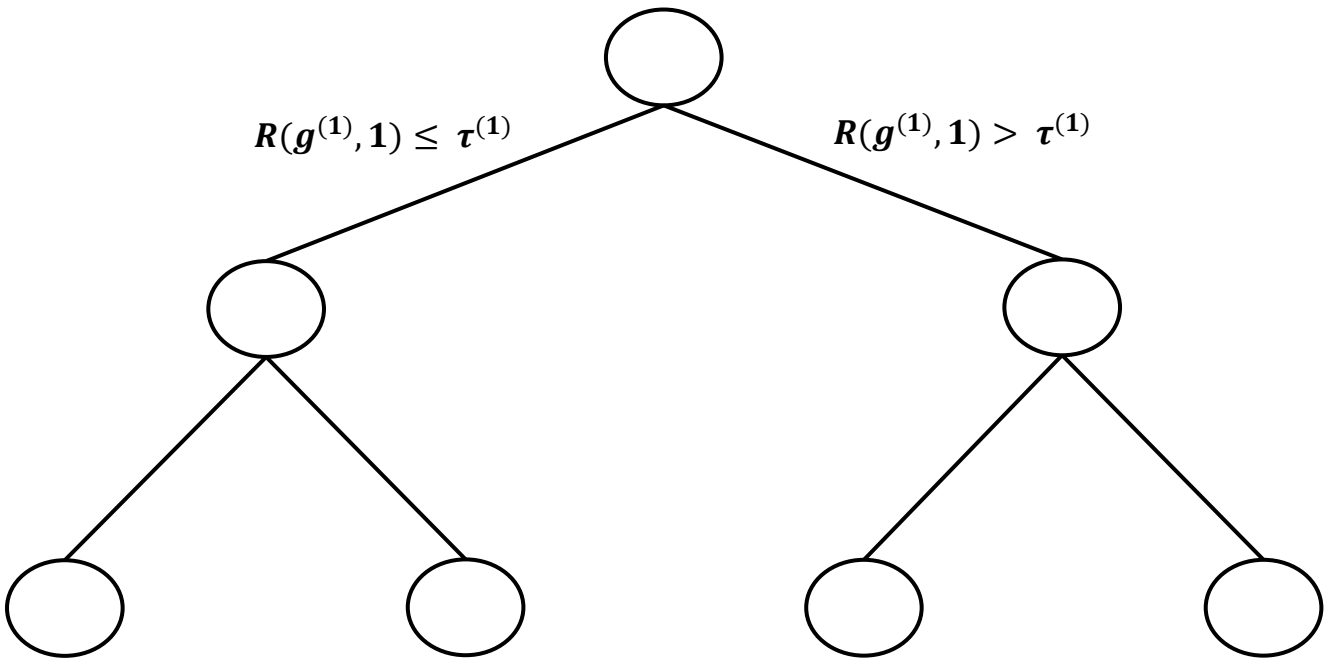
Deep conditional portfolio sort

Start with all data and find the optimal sorting variable $R(g^{(1)}, 1)$ and the associated threshold $\tau^{(1)}$



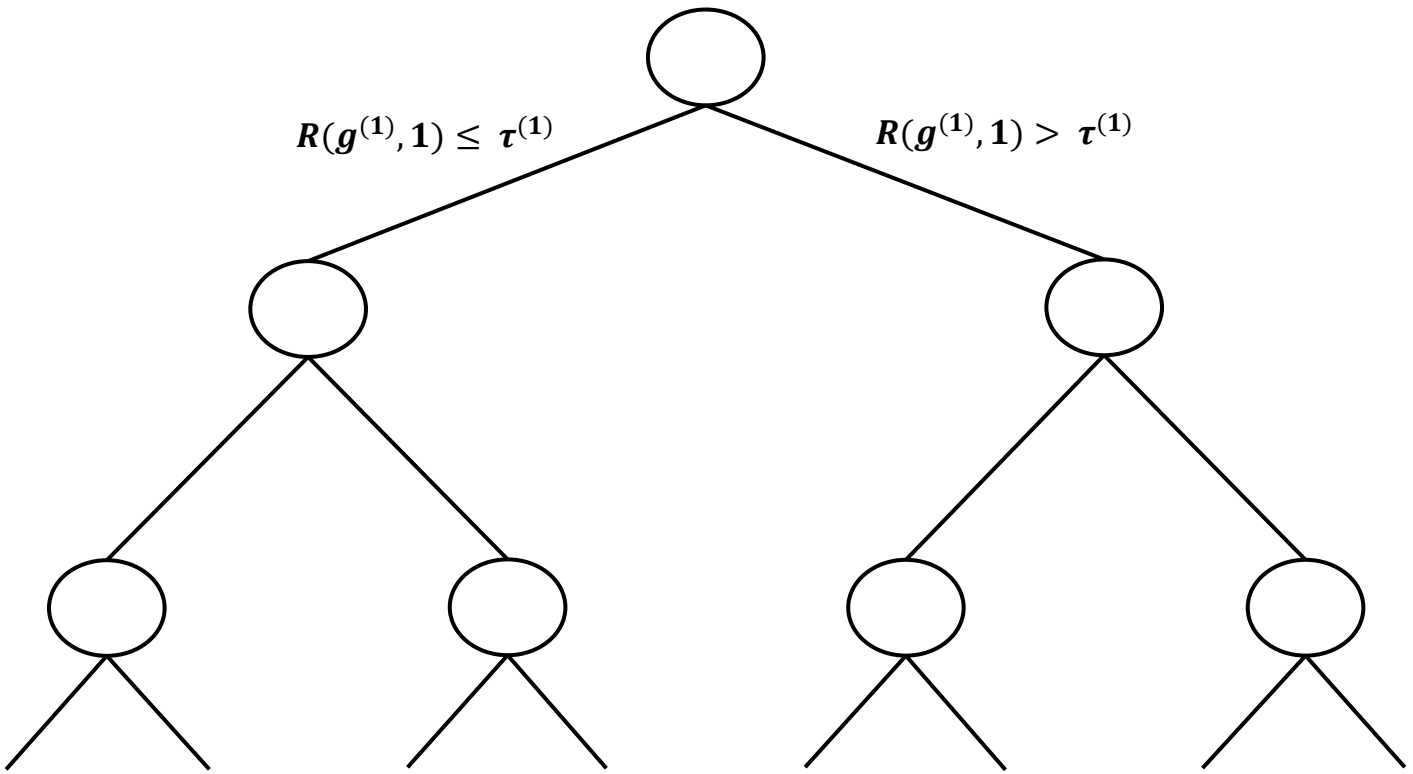
Deep conditional portfolio sort

In each subset of observations, find the optimal sorting variable and associated threshold



Deep conditional portfolio sort

Repeat until there are few observations in a node or the improvement becomes small. ▸ Details



Deep conditional portfolio sort

- In each terminal node, estimate expected return as

$$\hat{\mu}_l = \text{Mean}(r_{i,t+1} | \text{Firm i } \in S_l \text{ in period t})$$

- Prediction function is then

$$\hat{r}_{i,t+1} = f(z_{it}) = \sum_{l=1}^{L} \hat{\mu}_l \mathbb{1}(\text{Firm i } \in S_l \text{ in period t})$$

Summary of deep conditional portfolio sorts

- Method mirrors the estimation of a decision-tree in the machine learning literature

Summary of deep conditional portfolio sorts

- Method mirrors the estimation of a decision-tree in the machine learning literature
  - Can detect relevant interactions and therefore increases the type of relations that can be captured

Summary of deep conditional portfolio sorts

- Method mirrors the estimation of a decision-tree in the machine learning literature
  - Can detect relevant interactions and therefore increases the type of relations that can be captured
  - Local approximation of the dependent variable over parts of the regressor space $\neq$ linear regression

Summary of deep conditional portfolio sorts

- Method mirrors the estimation of a decision-tree in the machine learning literature
  - Can detect relevant interactions and therefore increases the type of relations that can be captured
  - Local approximation of the dependent variable over parts of the regressor space $\neq$ linear regression
- In practice, these advantages come at the cost of weak predictive power out-of-sample $\rightarrow$ model averaging as a solution

Deep conditional portfolio sort

Two steps

1. Extend the method of conditional portfolio sorts
2. Use model averaging to improve the estimates

Model averaging: Illustration

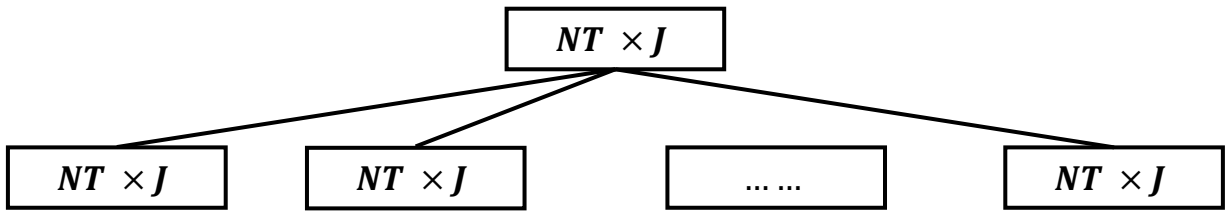
1. Estimation data

$$NT \times J$$

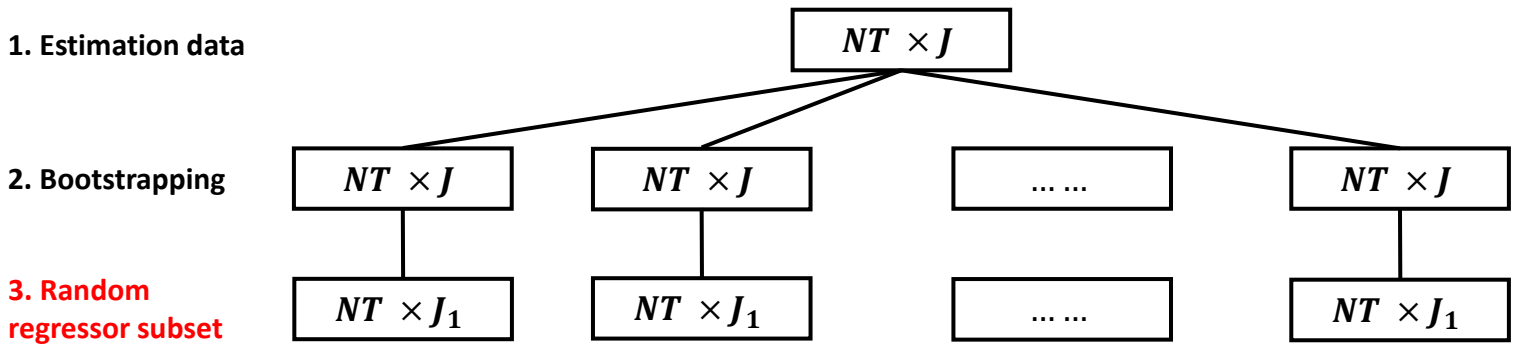
Model averaging: Illustration

1. Estimation data

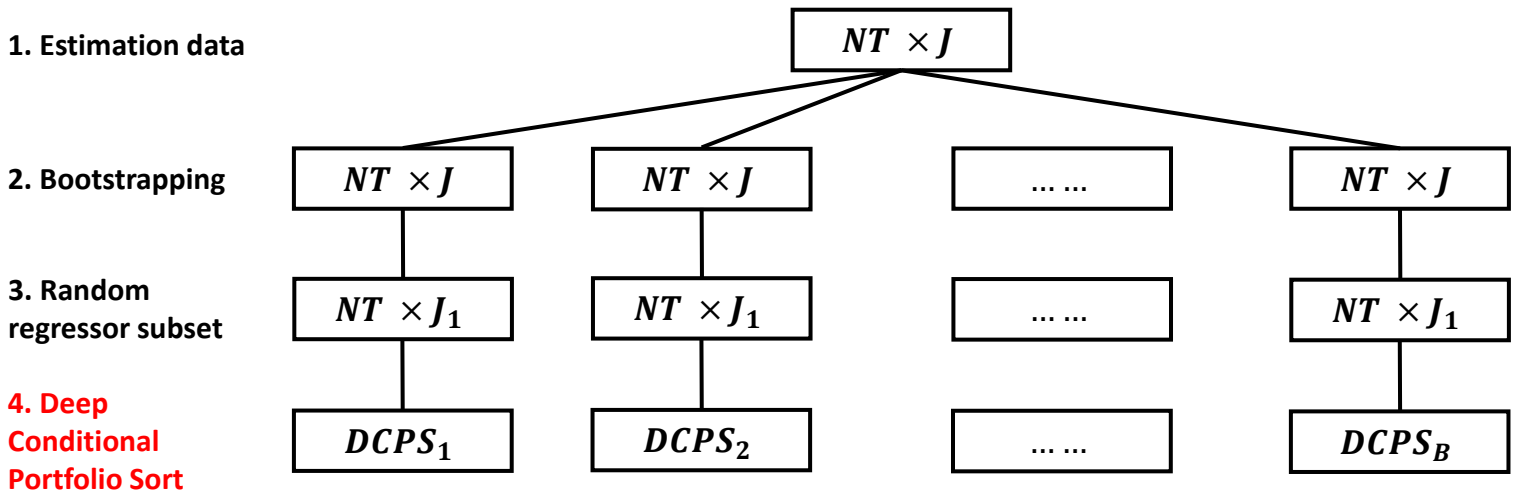
2. Bootstrapping



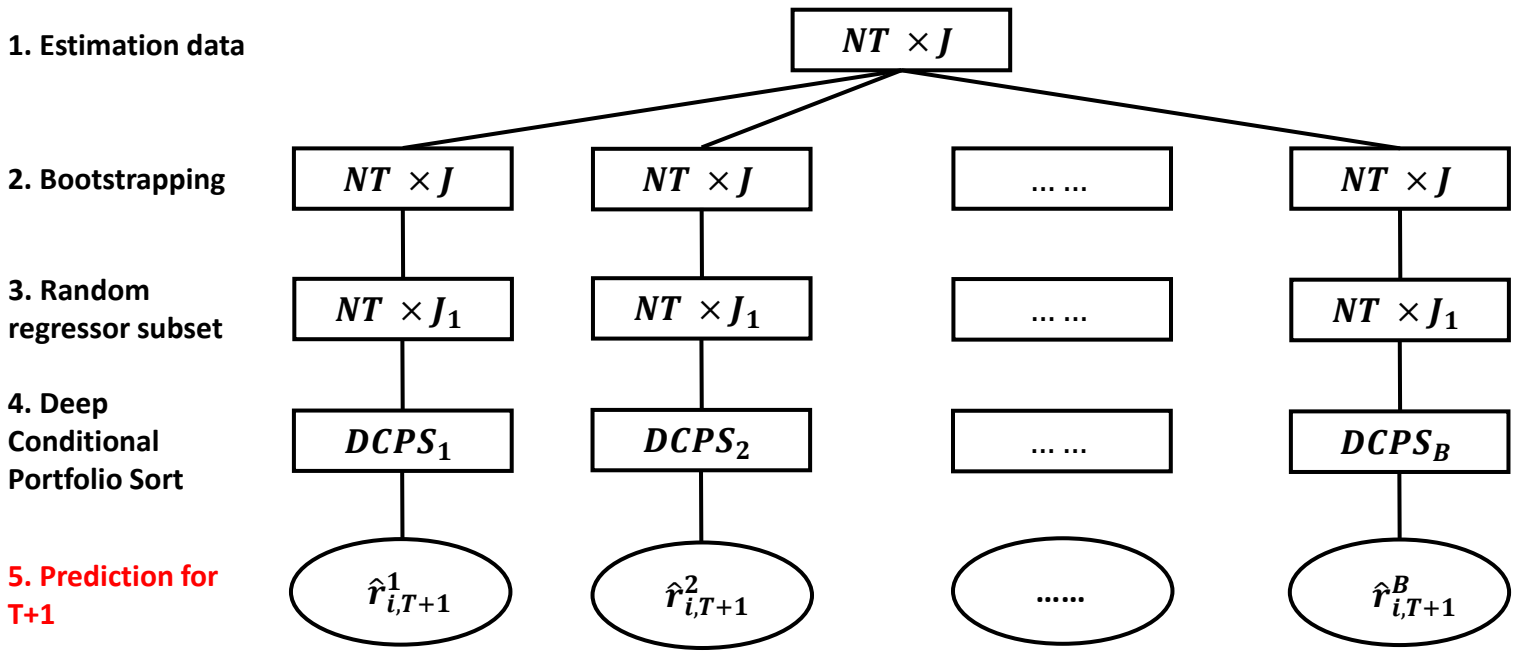
Model averaging: Illustration



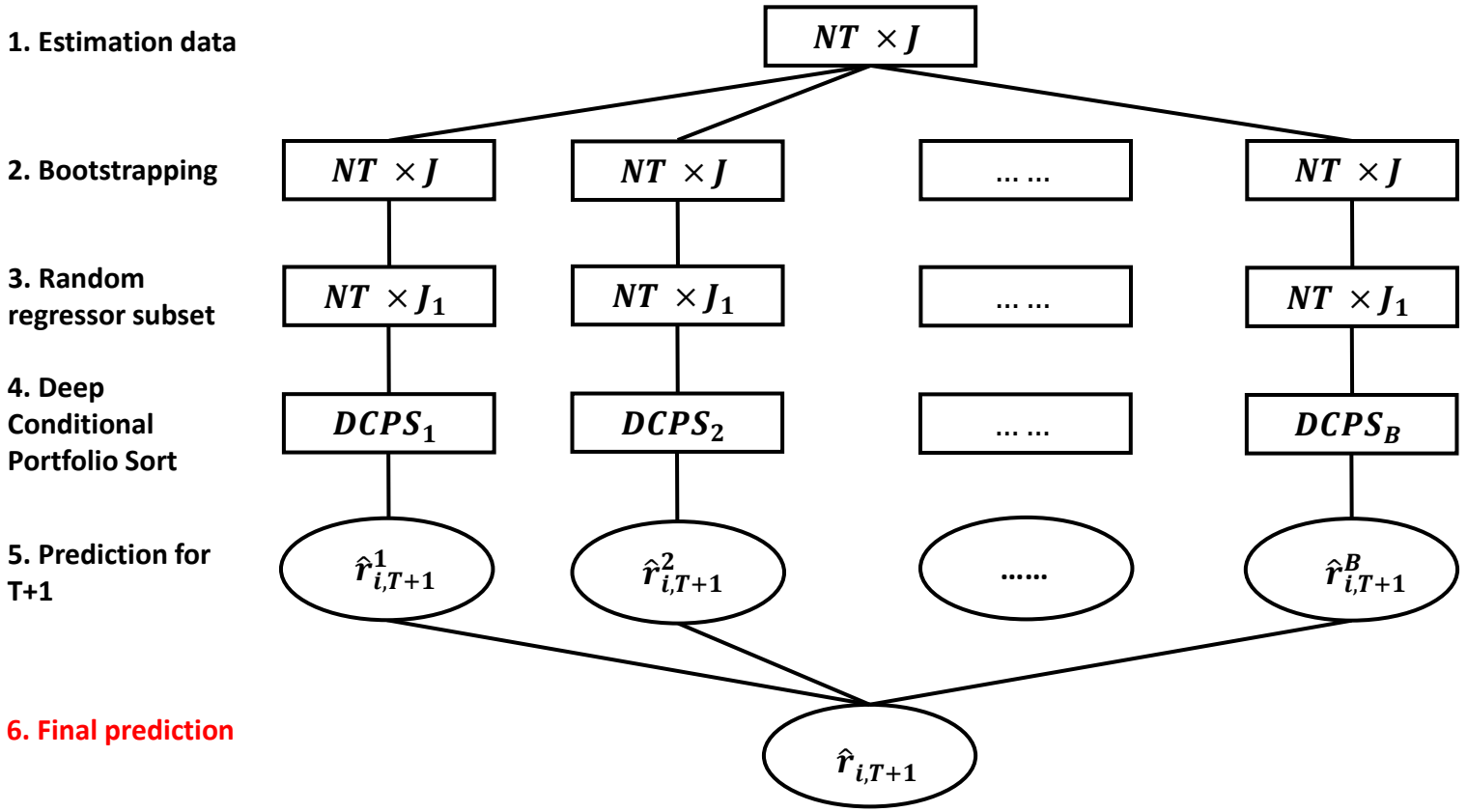
Model averaging: Illustration



Model averaging: Illustration



Model averaging: Illustration



▸ Details

RESULTS

Data

- CRSP monthly stock file
- Exclude penny stocks
- Ordinary common shares only
- Truncate .5% and 99.5% monthly return percentile
- FF factors and UMD
- Compustat and IBES for robustness checks
- Matched sample period 1963-2012

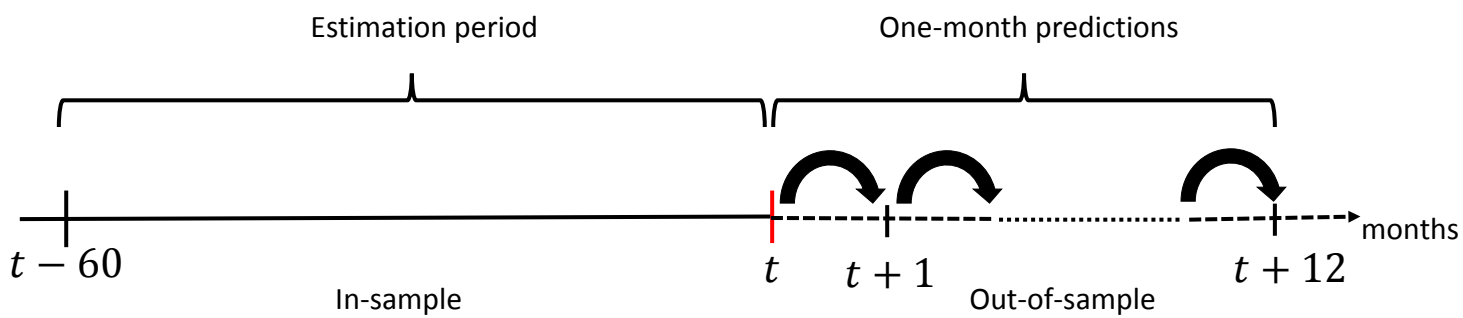
Recall the four methods that we evaluate:

1. Simple portfolio sort
2. Fama-MacBeth regression (levels only)
3. Fama-MacBeth regression (with two-way interactions)
4. Deep conditional portfolio sorts

Model evaluation: Quality of return prediction

Out-of-sample testing procedure

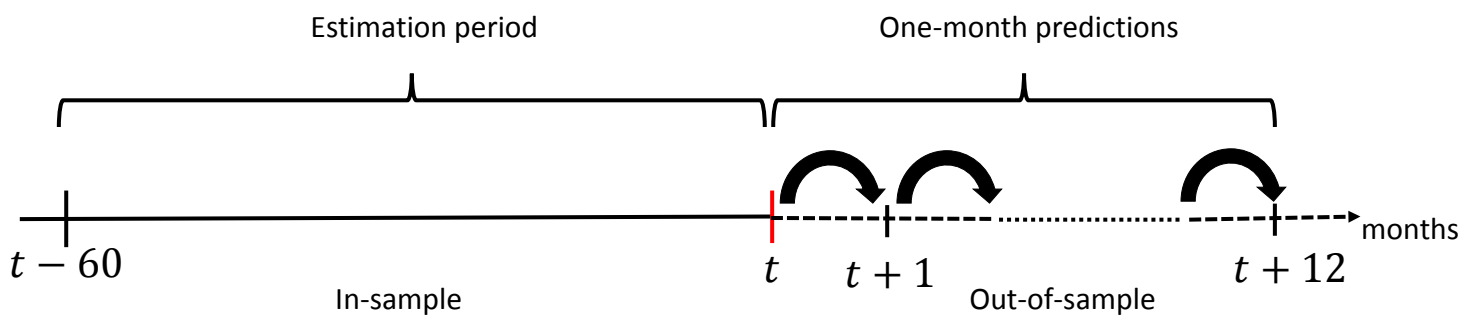
1. At beginning of each year, estimate model with data over past 60 months



Model evaluation: Quality of return prediction

Out-of-sample testing procedure

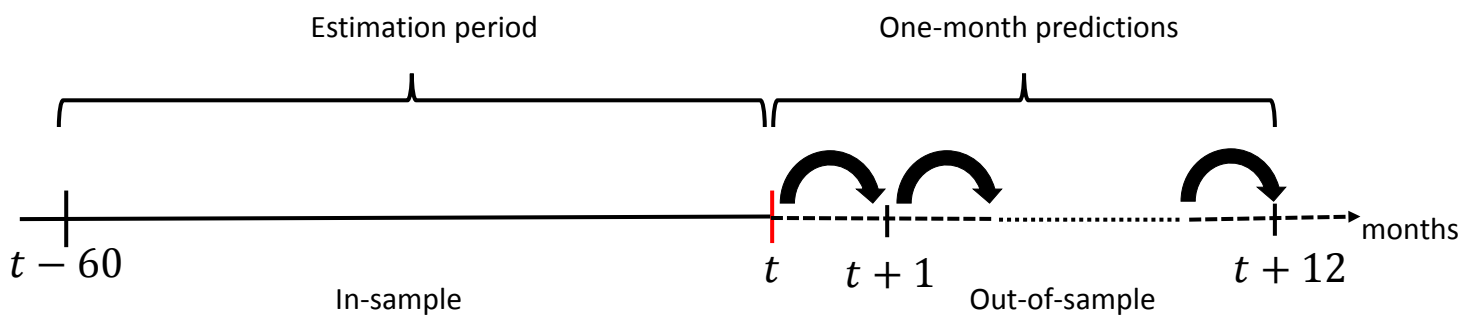
1. At beginning of each year, estimate model with data over past 60 months
2. Fix model estimates over the next year



Model evaluation: Quality of return prediction

Out-of-sample testing procedure

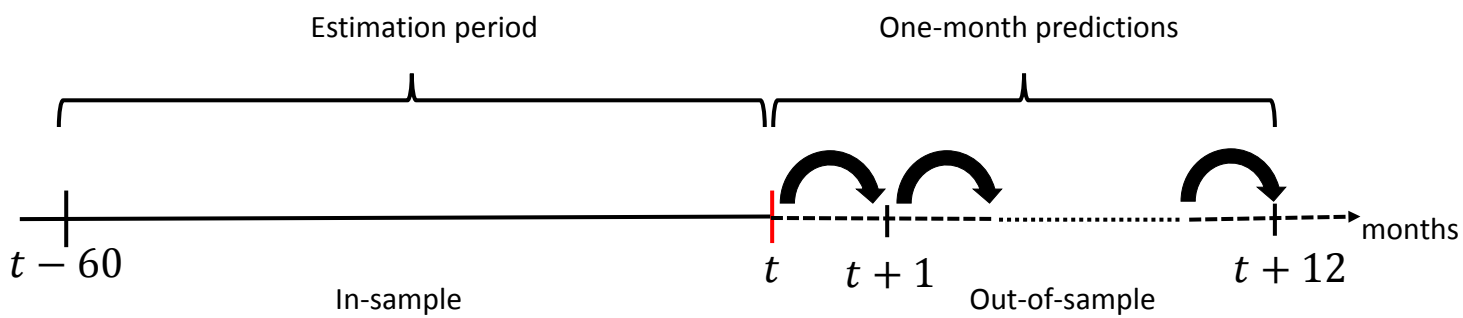
1. At beginning of each year, estimate model with data over past 60 months
2. Fix model estimates over the next year
3. Predict next months' returns each month: buy highest predictions, sell lowest predictions



Model evaluation: Quality of return prediction

Out-of-sample testing procedure

1. At beginning of each year, estimate model with data over past 60 months
2. Fix model estimates over the next year
3. Predict next months' returns each month: buy highest predictions, sell lowest predictions
4. Re-calibrate model at beginning of next year



Results: Portfolio sorts

Average return of about .7 percent per month

	(1)	(2)	(3)	(4)
Intercept	0.71	0.74	0.71	0.72
	(6.45)	(6.77)	(6.60)	(6.67)
MKT		-0.03	-0.03	-0.03
		(-1.51)	(-1.28)	(-1.28)
SMB			0.03	0.03
			(0.73)	(0.73)
HML			0.04	0.04
			(1.17)	(1.16)
UMD				-0.00
				(-0.11)
R^2		0.00	0.01	0.01
IR		0.92	0.89	0.89
SR	0.88			
N	540	540	540	540

t-statistics in parentheses.

Results: Portfolio sorts

Information ratio of around .9, Sharpe ratio of .88

	(1)	(2)	(3)	(4)
Intercept	0.71	0.74	0.71	0.72
	(6.45)	(6.77)	(6.60)	(6.67)
MKT		-0.03	-0.03	-0.03
		(-1.51)	(-1.28)	(-1.28)
SMB			0.03	0.03
			(0.73)	(0.73)
HML			0.04	0.04
			(1.17)	(1.16)
UMD				-0.00
				(-0.11)
R^2		0.00	0.01	0.01
IR		0.92	0.89	0.89
SR	0.88			
N	540	540	540	540

t-statistics in parentheses.

Results: Fama-Macbeth (levels only)

Average return of 1.5 percent per month

	(1)	(2)	(3)	(4)
Intercept	1.50	1.31	1.28	1.00
	(8.63)	(7.60)	(7.79)	(6.16)
MKT		0.21	0.21	0.26
		(3.29)	(3.34)	(4.51)
SMB			0.05	0.05
			(0.53)	(0.47)
HML			0.05	0.14
			(0.44)	(1.31)
UMD				0.27
				(3.76)
R^2		0.06	0.07	0.16
IR		1.20	1.17	0.97
SR	1.33			
N	540	540	540	540

t-statistics in parentheses.

Results: Fama-Macbeth (levels only)

Information ratio of 1-1.2, Sharpe ratio of 1.33

	(1)	(2)	(3)	(4)
Intercept	1.50	1.31	1.28	1.00
	(8.63)	(7.60)	(7.79)	(6.16)
MKT		0.21	0.21	0.26
		(3.29)	(3.34)	(4.51)
SMB			0.05	0.05
			(0.53)	(0.47)
HML			0.05	0.14
			(0.44)	(1.31)
UMD				0.27
				(3.76)
R^2		0.06	0.07	0.16
IR		1.20	1.17	0.97
SR	1.33			
N	540	540	540	540

t-statistics in parentheses.

Results: Fama-Macbeth (with two-way interactions)

Average return of 1.46 percent per month

	(1)	(2)	(3)	(4)
Intercept	1.46	1.27	1.28	1.13
	(9.62)	(8.29)	(8.35)	(7.55)
MKT		0.21	0.18	0.21
		(4.38)	(3.82)	(4.40)
SMB			0.12	0.12
			(1.63)	(1.48)
HML			-0.03	0.02
			(-0.33)	(0.20)
UMD				0.15
				(2.91)
R^2		0.09	0.11	0.15
IR		1.43	1.46	1.32
SR	1.57			
N	540	540	540	540

t-statistics in parentheses.

Results: Fama-Macbeth (with two-way interactions)

Information ratio of 1.3-1.4, Sharpe ratio of 1.57

	(1)	(2)	(3)	(4)
Intercept	1.46	1.27	1.28	1.13
	(9.62)	(8.29)	(8.35)	(7.55)
MKT		0.21	0.18	0.21
		(4.38)	(3.82)	(4.40)
SMB			0.12	0.12
			(1.63)	(1.48)
HML			-0.03	0.02
			(-0.33)	(0.20)
UMD				0.15
				(2.91)
R^2		0.09	0.11	0.15
IR		1.43	1.46	1.32
SR	1.57			
N	540	540	540	540

t-statistics in parentheses.

Results: Deep conditional portfolio sort

Average return of 2.3 percent per month

	(1)	(2)	(3)	(4)
Intercept	2.30	2.23	2.25	2.05
	(16.75)	(16.04)	(16.51)	(14.54)
MKT		0.07	0.05	0.09
		(2.14)	(1.53)	(2.78)
SMB			0.08	0.09
			(1.40)	(1.69)
HML			-0.03	0.04
			(-0.39)	(0.61)
UMD				0.20
				(5.57)
R^2		0.02	0.03	0.13
IR		2.90	2.93	2.82
SR	2.96			
N	540	540	540	540

t-statistics in parentheses.

Results: Deep conditional portfolio sort

Extremely high information ratios of 2.8-2.9 and Sharpe ratio of 2.96

	(1)	(2)	(3)	(4)
Intercept	2.30	2.23	2.25	2.05
	(16.75)	(16.04)	(16.51)	(14.54)
MKT		0.07	0.05	0.09
		(2.14)	(1.53)	(2.78)
SMB			0.08	0.09
			(1.40)	(1.69)
HML			-0.03	0.04
			(-0.39)	(0.61)
UMD				0.20
				(5.57)
R^2		0.02	0.03	0.13
IR		2.90	2.93	2.82
SR	2.96			
N	540	540	540	540

t-statistics in parentheses.

Results: Deep conditional portfolio sort

- Average annual return positive in all years of the sample ▸ Figure
- Improved return spread throughout all deciles ▸ Figure
- Decile portfolios do not differ much in their loadings on risk factors ▸ Table

EVALUATING DEEP CONDITIONAL PORTFOLIO SORTS

Evaluation

- Deep conditional portfolio sort appears to perform really well
- Not as easy to summarize as a linear regression

Evaluation

- Deep conditional portfolio sort appears to perform really well
- Not as easy to summarize as a linear regression
- Answer two questions:

Evaluation

- Deep conditional portfolio sort appears to perform really well
- Not as easy to summarize as a linear regression
- Answer two questions:
  1. Which variables are most important for predictions? (”t-statistics”)

Evaluation

- Deep conditional portfolio sort appears to perform really well
- Not as easy to summarize as a linear regression
- Answer two questions:
  1. Which variables are most important for predictions? (”t-statistics”)
  2. How do these variables affect predictions? (”partial derivatives”)

Evaluation: Variable importance

1. Randomly permute values of one variable at a time
2. Compute predictions and increase in MSE
3. Rank variables by their relative increase

Evaluation: Variable importance

Sort by median rank over 45 years

	Median rank
R(0,1)	1
R(1,1)	4
R(2,1)	4
R(3,1)	6
R(11,1)	7
R(4,1)	8
R(5,1)	8
R(8,1)	11
R(10,1)	11
R(9,1)	12

R(g,1) = one-month return g months ago

Evaluation: Variable importance

Mostly recent returns show up

	Median rank
R(0,1)	1
R(1,1)	4
R(2,1)	4
R(3,1)	6
R(11,1)	7
R(4,1)	8
R(5,1)	8
R(8,1)	11
R(10,1)	11
R(9,1)	12

R(g,1) = one-month return g months ago

Evaluation: Variable importance

Two well-known past returns rank highly on the list

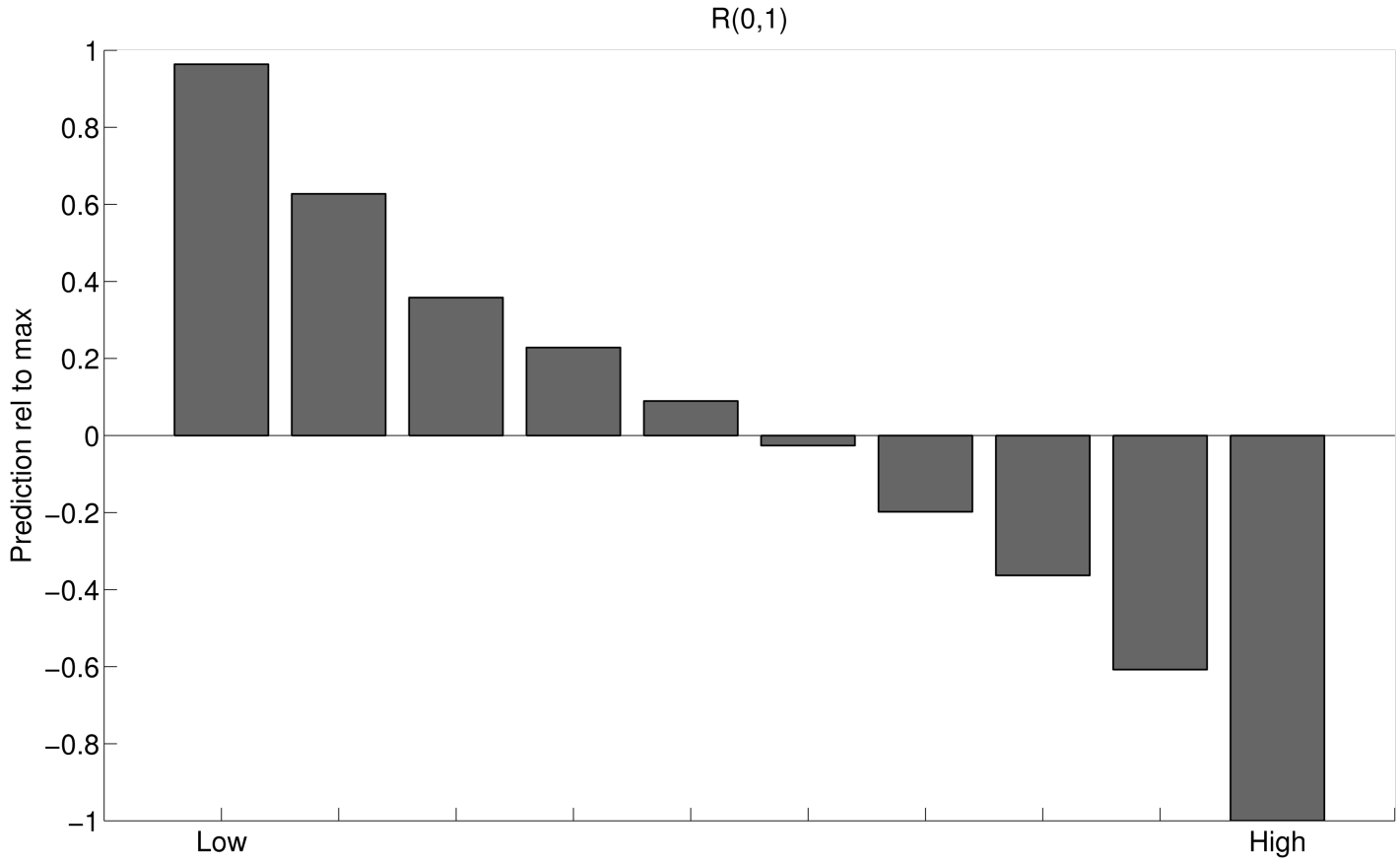
	Median rank
R(0,1)	1
R(1,1)	4
R(2,1)	4
R(3,1)	6
R(11,1)	7
R(4,1)	8
R(5,1)	8
R(8,1)	11
R(10,1)	11
R(9,1)	12

R(g,1) = one-month return g months ago

Ranking of variables relatively stable over time ▸ Spearman

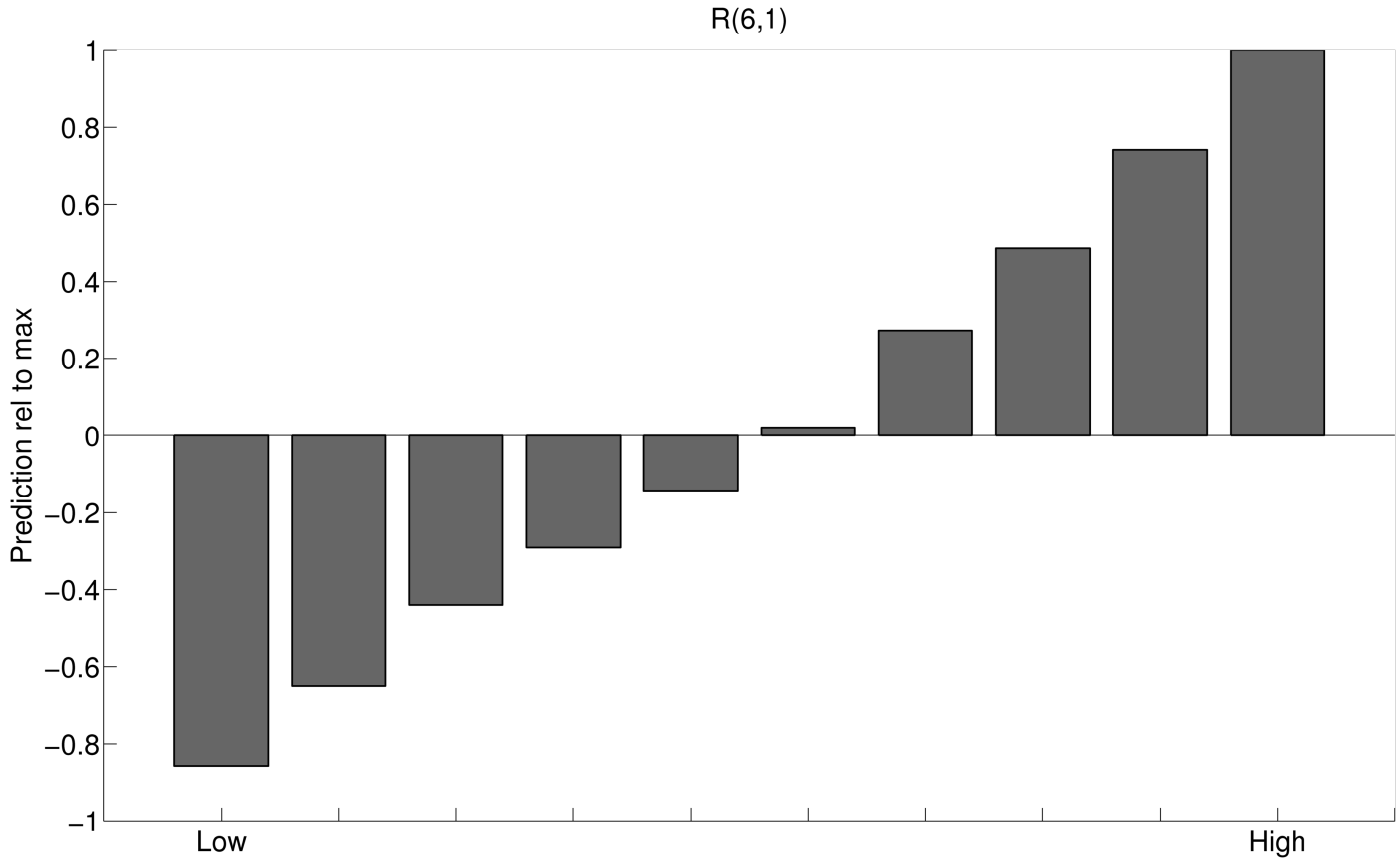
Evaluation: Partial derivatives

Find short-term reversal for R(0,1)



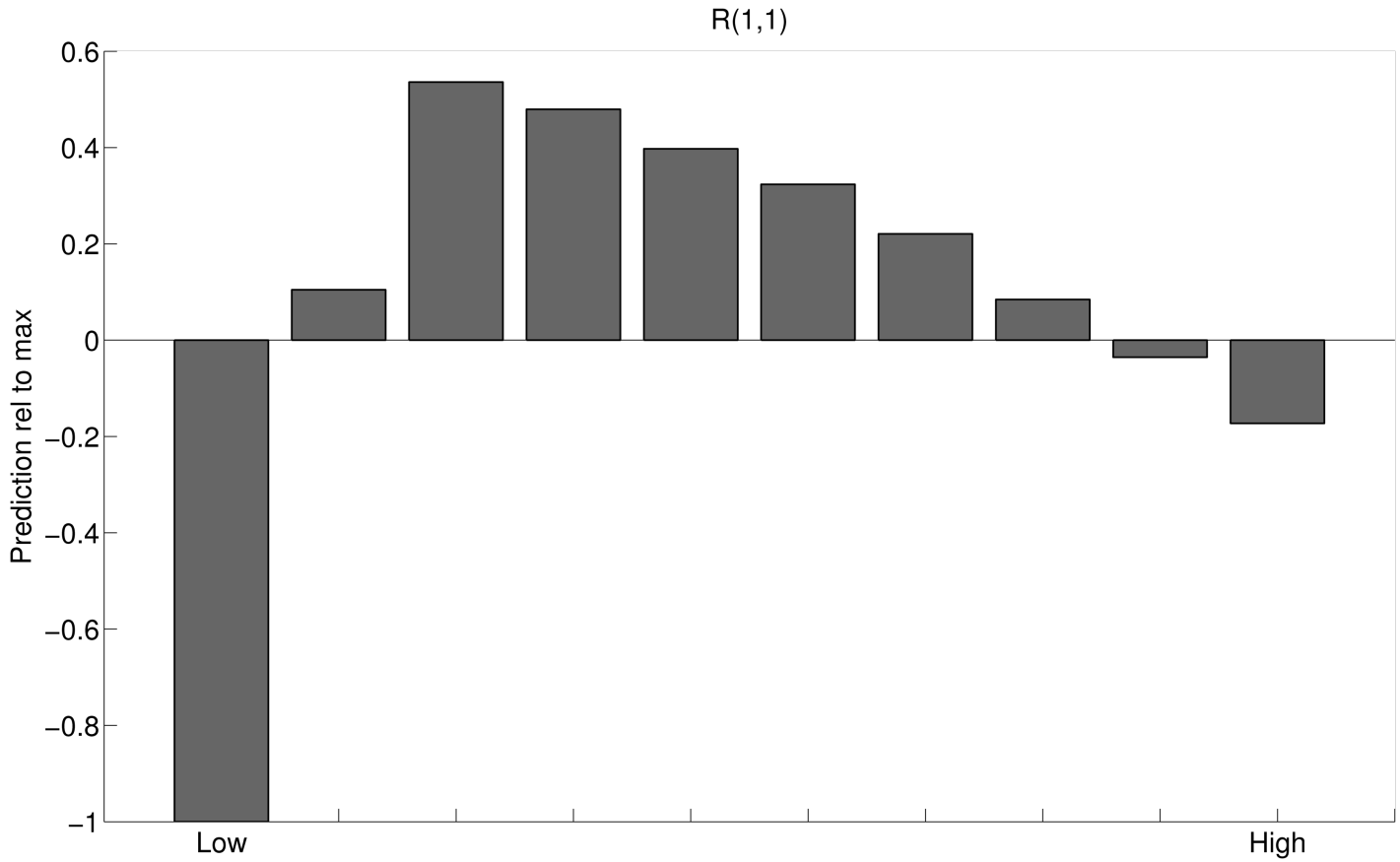
Evaluation: Partial derivatives

Find momentum for R(6,1)

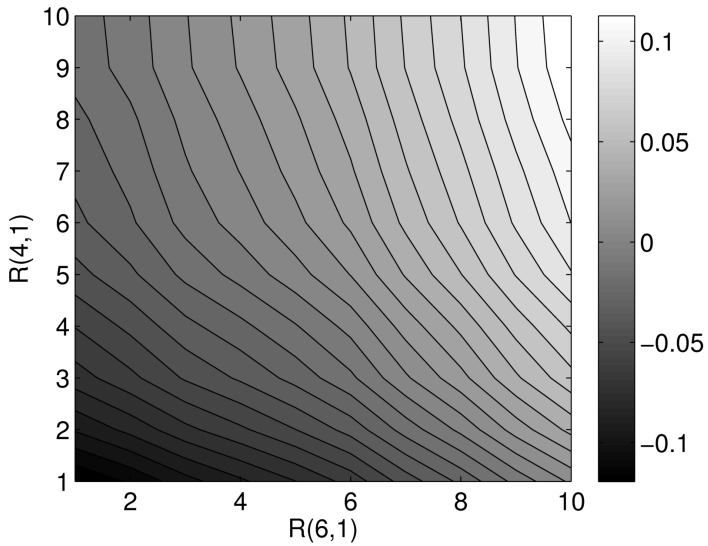


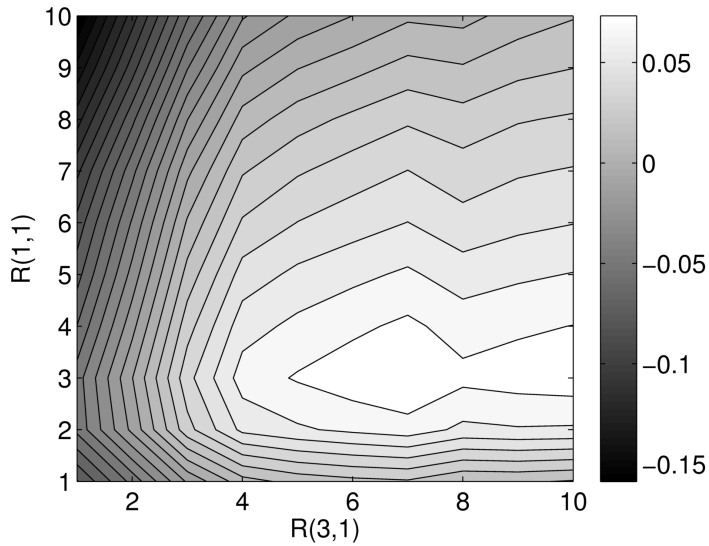
Evaluation: Partial derivatives

Find non-linear relationship for R(1,1)



Evaluation: Double partial derivatives





ROBUSTNESS

Robustness

- Further specifications
  - Include a set of 86 firm fundamentals ▸ Results
  - Estimate by size category ▸ Results
  - Estimate with larger set of correlated return variables ▸ Results

Robustness

- Further specifications
  - Include a set of 86 firm fundamentals ▸ Results
  - Estimate by size category ▸ Results
  - Estimate with larger set of correlated return variables ▸ Results
- Transaction costs
  - Approximation based on Frazzini, Israel and Moskowitz (2013) ▸ Results

Robustness: Predictive models

- All specifications point to the importance of more recent past returns
- What if we use only recent past returns?

Robustness: Predictive models

- All specifications point to the importance of more recent past returns
- What if we use only recent past returns?
  - Deep conditional portfolio sort: Performs just as well ▸ Table

Robustness: Predictive models

- All specifications point to the importance of more recent past returns
- What if we use only recent past returns?
  - Deep conditional portfolio sort: Performs just as well ▸ Table
  - Fama-MacBeth regression with interactions: further increase in average return and information ratios ▸ Table

CONCLUSION

Conclusion

- We tailor a method from the machine learning literature to stock return prediction

Conclusion

- We tailor a method from the machine learning literature to stock return prediction
- We apply the method to predict future stock returns based on past stock returns

Conclusion

- We tailor a method from the machine learning literature to stock return prediction
- We apply the method to predict future stock returns based on past stock returns
- We show that it performs better than standard methods along three dimensions:
  - Average return
  - Sharpe ratio
  - Information ratio

Conclusion

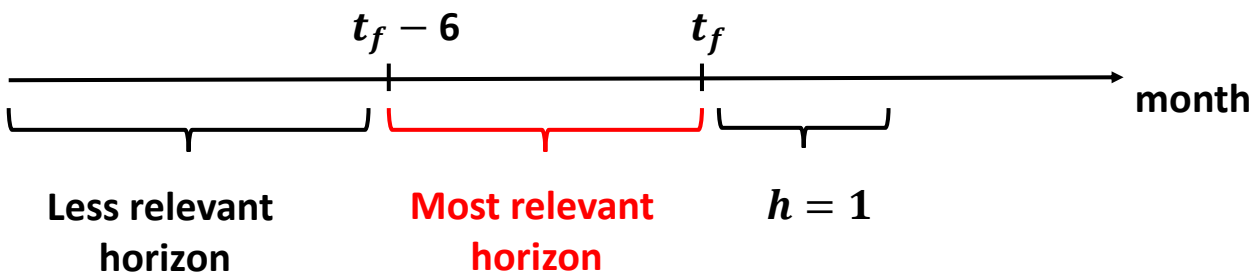
- Relationship between past and future returns seems to be more involved than can be captured by any single one past return.

Conclusion

- Relationship between past and future returns seems to be more involved than can be captured by any single one past return.
- Which past returns are most relevant to predict returns over the next month?

Conclusion

- Relationship between past and future returns seems to be more involved than can be captured by any single one past return.
- Which past returns are most relevant to predict returns over the next month?
- More recent returns turn out to be more relevant once all considered at the same time



Outlook

- Still remarkable lack of consensus about which variables can be robustly related to expected stock returns
- Work in progress: Use a large set of firm fundamentals and find the most relevant ones

APPENDIX

Fama-MacBeth approach

Fama-MacBeth approach as in Lewellen (2013):

Back

Fama-MacBeth approach

Fama-MacBeth approach as in Lewellen (2013):

1. In each cross-section, fit the regression

$$r_{i,t+1} = \beta_0^t + \sum_{g=1}^{24} \beta_g^t R_{it}(g,1) + \epsilon_{it}$$

Back

Fama-MacBeth approach

Fama-MacBeth approach as in Lewellen (2013):

1. In each cross-section, fit the regression

$$r_{i,t+1} = \beta_0^t + \sum_{g=1}^{24} \beta_g^t R_{it}(g,1) + \epsilon_{it}$$

2. Average coefficients over the past m months

$$\overline{\beta}_g^{t-1} = \frac{1}{m} \sum_{j=t-1-m}^{t-1} \beta_g^t$$

Back

Fama-MacBeth approach

Fama-MacBeth approach as in Lewellen (2013):

1. In each cross-section, fit the regression

$$r_{i,t+1} = \beta_0^t + \sum_{g=1}^{24} \beta_g^t R_{it}(g,1) + \epsilon_{it}$$

2. Average coefficients over the past m months

$$\overline{\beta}_g^{t-1} = \frac{1}{m} \sum_{j=t-1-m}^{t-1} \beta_g^t$$

3. Prediction is

$$\hat{r}_{i,t+1} = \overline{\beta}_0^{t-1} + \sum_{g=1}^{24} \overline{\beta}_g^{t-1} R_{it}(g,1),$$

Back

Fama-MacBeth approach

1. In each cross-section, fit the regression

Show results based on variable selection via LASSO

$$r_{i,t+1} = \beta_0^t + \sum_{g=1}^{24} \beta_g^t R_{it}(g,1) + \epsilon_{it}$$

2. Average coefficients over the past m months

$$\overline{\beta}_g^{t-1} = \frac{1}{m} \sum_{j=t-1-m}^{t-1} \beta_g^t$$

3. Prediction is

$$\hat{r}_{i,t+1} = \overline{\beta}_0^{t-1} + \sum_{g=1}^{24} \overline{\beta}_g^{t-1} R_{it}(g,1),$$

Back

Fama-MacBeth approach with interactions

- Add two way interactions

$$r_{i,t+1} = a + \sum_{g=1}^{24} \beta_m R_{i,t}(g,1) + \sum_{g=1}^{24} \sum_{j>g} \gamma_{jg} R_{i,t}(g,1) R_{i,t}(j,1) + \epsilon_{i,t}$$

- Use LASSO to regularize the model (350 regressors)

Back

Greedy conditional portfolio sort

How are sorting variables and sorting values estimated?

- Computationally hard to optimize globally
- Instead: Use a greedy algorithm
  - Within each subset of remaining observations, find the variable and sorting value that minimize mean squared error by brute-force

$$SC(g,\tau) = \min_{\mu_1} \left(\sum_{R_{it}(g,1) \in S_1(g,\tau)} (r_{i,t+1} - \mu_1)^2 \right) + \min_{\mu_2} \left(\sum_{R_{it}(g,1) \in S_2(g,\tau)} (r_{i,t+1} - \mu_2)^2 \right)$$

▸ Back

Model averaging

- Random regressor selection (Kleinberg, 1990): Estimate model many (B) times with only random subsets of regressors
- Bootstrap aggregating (Breiman 1996): Estimate model many (B) times on only subsets of the observations

Final prediction is computed as

$$\hat{r}_{i,t+1} = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(z_{i,t})$$

Back

Model averaging: Discussion

- ”Random forests” in the decision-tree literature in ML

Model averaging: Discussion

- ”Random forests” in the decision-tree literature in ML
- Helps with overfitting:

Model averaging: Discussion

- ”Random forests” in the decision-tree literature in ML
- Helps with overfitting:
  - individual models only fitted on subsets of the data

Model averaging: Discussion

- ”Random forests” in the decision-tree literature in ML
- Helps with overfitting:
  - individual models only fitted on subsets of the data
  - individual models only fitted on subset of regressors (not very complex)

Model averaging: Discussion

- ”Random forests” in the decision-tree literature in ML
- Helps with overfitting:
  - individual models only fitted on subsets of the data
  - individual models only fitted on subset of regressors (not very complex)
- Works well in many diverse settings (XBox, DNA sequencing, consumer credit risk predictions)

Model averaging: Discussion

- ”Random forests” in the decision-tree literature in ML
- Helps with overfitting:
  - individual models only fitted on subsets of the data
  - individual models only fitted on subset of regressors (not very complex)
- Works well in many diverse settings (XBox, DNA sequencing, consumer credit risk predictions)
- Theory lagging behind applications but some results on consistency and out-of-sample errors ▸ Details

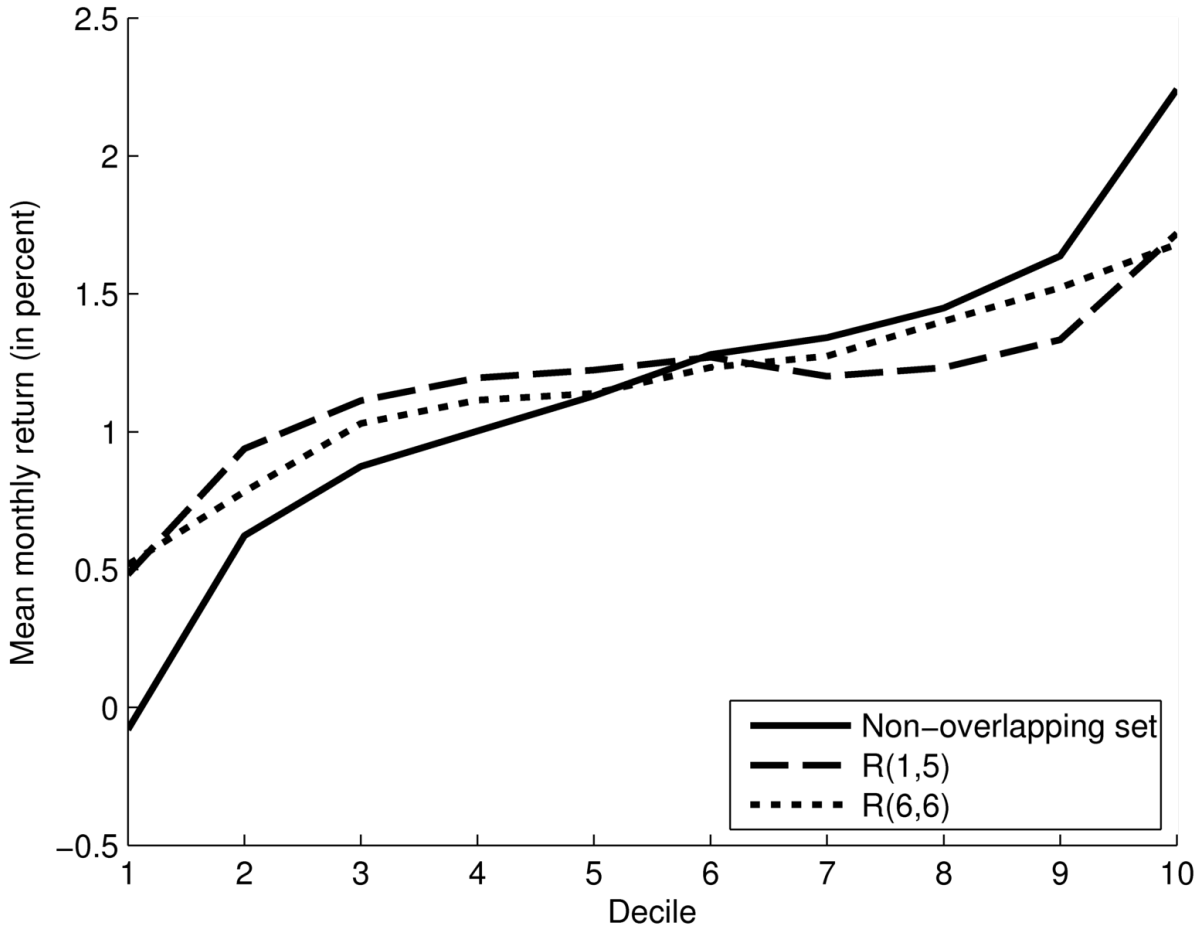
Model averaging: Theory results

- Breiman (2001): Bound on out-of-sample error depends on correlation between underlying models
- Büchlmann and Yu (2002): Model averaging helps when the underlying models involve hard thresholds
  - Error in threshold estimate makes models unstable
  - Averaging smooths out the decision boundary
  - Bagging reduces variance with small effect on bias
- Biau, Devroye and Lugosi (2008): Prediction of averaged decision-trees is consistent for true value
- Some theoretical questions still open

Back

Return spread

Back



Decile factor loadings

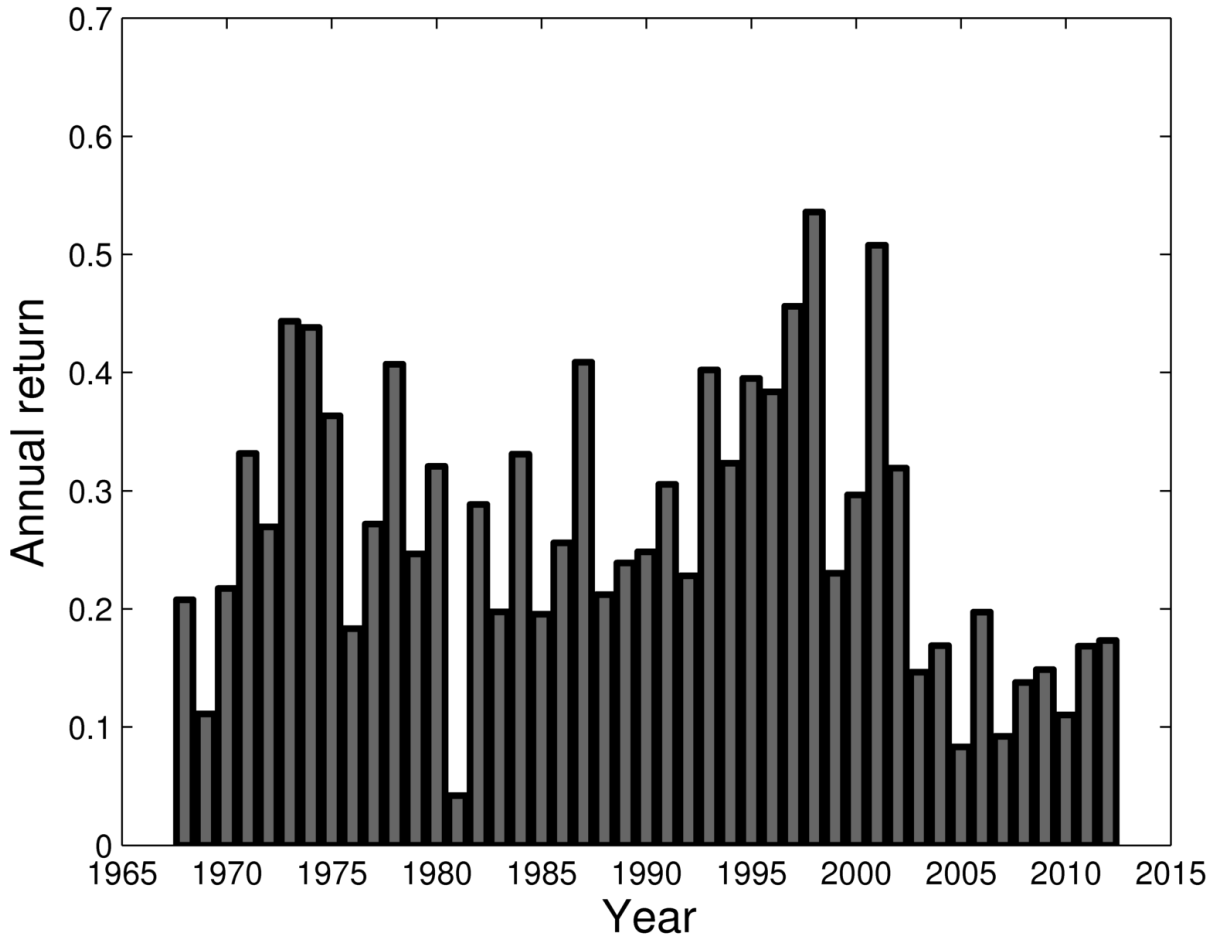
Back

	Low	2	3	4	5	6	7	8	9	High	High-Low
Average return	-0.53	0.21	0.44	0.58	0.68	0.80	0.94	1.01	1.22	1.76	2.30
	(-1.74)	(0.77)	(1.66)	(2.29)	(2.61)	(3.10)	(3.52)	(3.78)	(4.27)	(5.54)	(16.75)
						CAPM					
Intercept	-1.52	-0.73	-0.48	-0.32	-0.23	-0.13	0.01	0.06	0.23	0.72	2.23
	(-8.59)	(-4.97)	(-3.60)	(-2.47)	(-1.81)	(-1.03)	(0.08)	(0.42)	(1.55)	(3.97)	(16.04)
MKT	1.12	1.08	1.06	1.04	1.04	1.06	1.06	1.09	1.14	1.20	0.07
	(27.40)	(30.01)	(31.94)	(32.94)	(30.65)	(31.83)	(30.62)	(28.90)	(29.10)	(25.30)	(2.14)
						Three factor model					
Intercept	-1.64	-0.87	-0.63	-0.48	-0.38	-0.28	-0.13	-0.07	0.10	0.61	2.25
	(-13.12)	(-7.71)	(-6.80)	(-5.58)	(-4.66)	(-3.45)	(-1.42)	(-0.71)	(0.98)	(4.51)	(16.51)
MKT	0.99	0.97	0.97	0.96	0.97	0.98	0.98	0.99	1.02	1.04	0.05
	(27.70)	(29.22)	(34.92)	(36.04)	(38.43)	(38.76)	(32.98)	(29.14)	(32.43)	(27.55)	(1.53)
SMB	0.87	0.74	0.69	0.67	0.65	0.66	0.69	0.71	0.80	0.95	0.08
	(8.64)	(8.49)	(8.76)	(8.51)	(8.80)	(8.69)	(8.51)	(8.84)	(10.21)	(12.08)	(1.40)
HML	0.23	0.25	0.27	0.30	0.29	0.28	0.27	0.25	0.25	0.21	-0.03
	(2.84)	(3.58)	(4.44)	(4.73)	(4.64)	(4.44)	(4.24)	(3.50)	(3.68)	(2.52)	(-0.39)
						Four factor model					
Intercept	-1.37	-0.67	-0.48	-0.36	-0.29	-0.21	-0.07	-0.02	0.13	0.69	2.05
	(-12.77)	(-7.10)	(-6.32)	(-4.96)	(-4.11)	(-3.10)	(-0.90)	(-0.25)	(1.40)	(5.13)	(14.54)
MKT	0.94	0.94	0.94	0.93	0.95	0.97	0.96	0.98	1.02	1.03	0.09
	(30.12)	(31.11)	(37.89)	(39.00)	(41.63)	(44.66)	(36.30)	(31.58)	(35.46)	(27.00)	(2.78)
SMB	0.86	0.73	0.69	0.67	0.65	0.65	0.69	0.71	0.80	0.95	0.09
	(11.69)	(11.04)	(10.88)	(10.14)	(10.15)	(9.58)	(9.15)	(9.38)	(10.52)	(13.16)	(1.69)
HML	0.15	0.19	0.23	0.26	0.25	0.26	0.25	0.23	0.24	0.18	0.04
	(2.47)	(3.39)	(4.40)	(4.91)	(4.96)	(4.65)	(4.53)	(3.62)	(3.86)	(2.34)	(0.61)
UMD	-0.27	-0.20	-0.15	-0.12	-0.10	-0.07	-0.06	-0.05	-0.03	-0.08	0.20
	(-7.71)	(-6.29)	(-4.87)	(-3.82)	(-2.90)	(-2.26)	(-1.86)	(-1.49)	(-0.93)	(-2.34)	(5.57)

t-statistics in parentheses.

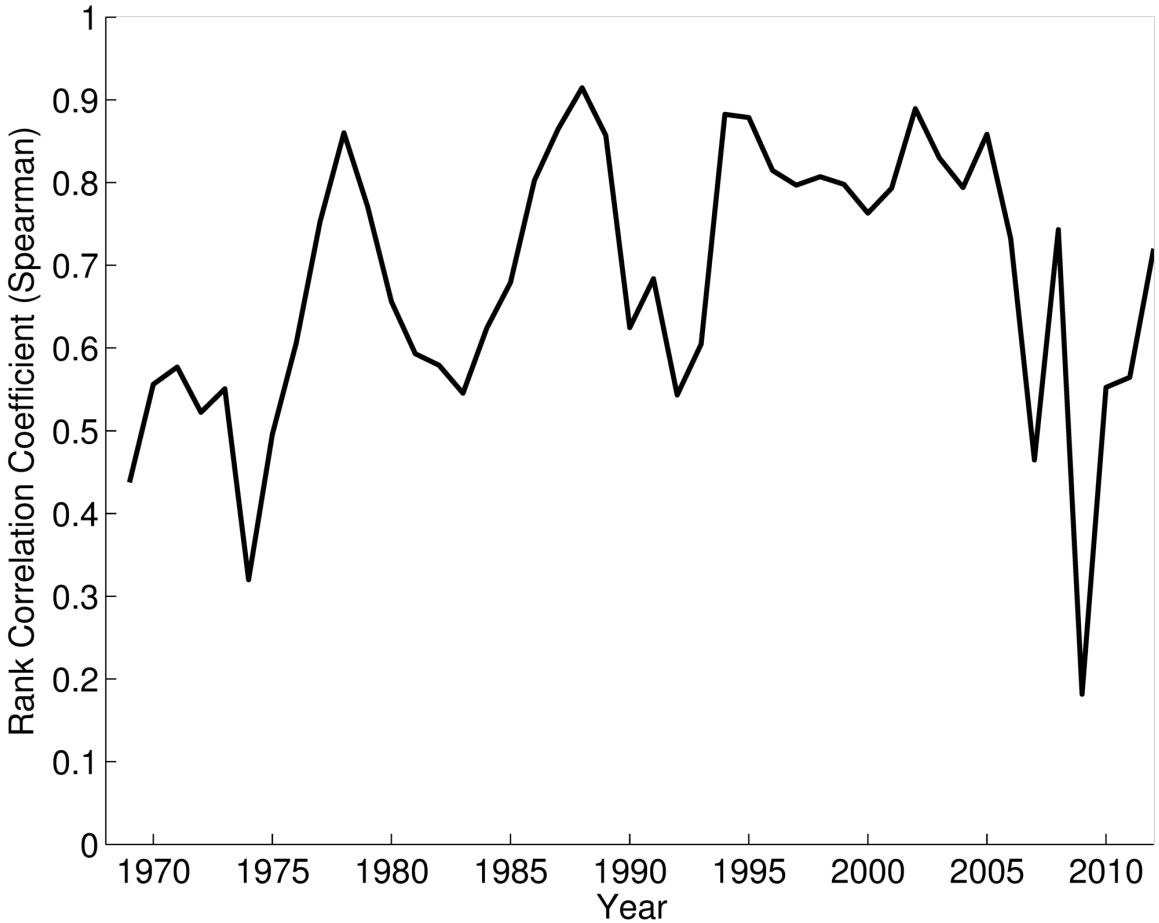
Average returns

Back



Variable importance: Spearman rank correlation

Median rank correlation: .7



Back

Variable importance

Including 86 firm fundamentals

- Construct firm fundamentals from Green, Hand and Zhang (2014)
- Popular characteristics that can be constructed with only Compustat, CRSP and IBES
- E.g. beta, size, value, earnings surprise, quality, volatility

Back

Variable importance

Including 86 firm fundamentals

No Fundamentals	With Fundamentals
R(0,1)	R(0,1)
R(1,1)	R(1,1)
R(2,1)	R(2,1)
R(3,1)	R(3,1)
R(11,1)	R(4,1)
R(4,1)	R(5,1)
R(5,1)	R(11,1)
R(8,1)	R(7,1)
R(10,1)	R(6,1)
R(9,1)	R(8,1)

Back

Variable importance

Including 86 firm fundamentals

No Fundamentals	With Fundamentals
R(0,1)	R(0,1)
R(1,1)	R(1,1)
R(2,1)	R(2,1)
R(3,1)	R(3,1)
R(11,1)	R(4,1)
R(4,1)	R(5,1)
R(5,1)	R(11,1)
R(8,1)	R(7,1)
R(10,1)	R(6,1)
R(9,1)	R(8,1)

Back

Variable importance

By size category

- Size breakpoints as in Fama and French (2008)
- Based on NYSE stocks
  - Micro: Lowest 20%
  - Small: 20% to 50%
  - Large: upper 50%

Back

Variable importance

By size category

All firms	Micro	Small	Large
R(0,1)	R(0,1)	R(0,1)	R(0,1)
R(1,1)	R(2,1)	R(2,1)	R(3,1)
R(2,1)	R(1,1)	R(1,1)	R(4,1)
R(3,1)	R(3,1)	R(3,1)	R(1,1)
R(11,1)	R(5,1)	R(4,1)	R(5,1)
R(4,1)	R(6,1)	R(5,1)	R(8,1)
R(5,1)	R(4,1)	R(6,1)	R(2,1)
R(8,1)	R(7,1)	R(7,1)	R(9,1)
R(10,1)	R(11,1)	R(8,1)	R(18,1)
R(9,1)	R(8,1)	R(11,1)	R(6,1)

Back

Variable importance

By size category

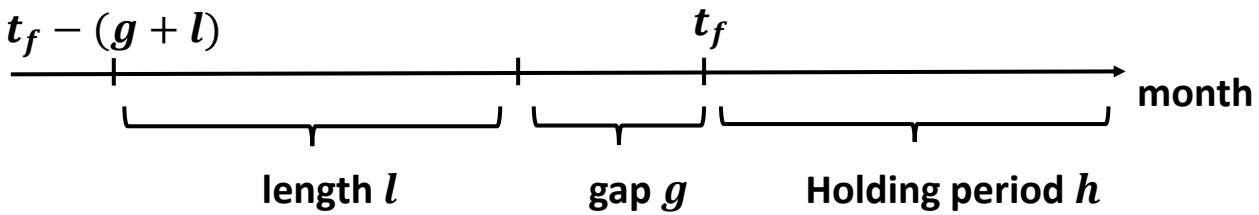
All firms	Micro	Small	Large
R(0,1)	R(0,1)	R(0,1)	R(0,1)
R(1,1)	R(2,1)	R(2,1)	R(3,1)
R(2,1)	R(1,1)	R(1,1)	R(4,1)
R(3,1)	R(3,1)	R(3,1)	R(1,1)
R(11,1)	R(5,1)	R(4,1)	R(5,1)
R(4,1)	R(6,1)	R(5,1)	R(8,1)
R(5,1)	R(4,1)	R(6,1)	R(2,1)
R(8,1)	R(7,1)	R(7,1)	R(9,1)
R(10,1)	R(11,1)	R(8,1)	R(18,1)
R(9,1)	R(8,1)	R(11,1)	R(6,1)

Back

Variable importance

Large set of correlated return variables

Consider the case: $h = 1; l = 1, \ldots, 6; g = 0, \ldots, 24$:



Back

Variable importance

Large set of correlated return variables

One-month returns	Correlated set
R(0,1)	R(0,1)
R(1,1)	R(1,1)
R(2,1)	R(2,1)
R(3,1)	R(3,1)
R(11,1)	R(6,1)
R(4,1)	R(5,1)
R(5,1)	R(4,1)
R(8,1)	R(0,2)
R(10,1)	R(1,2)
R(9,1)	R(3,2)

Back

Variable importance

Large set of correlated return variables

One-month returns	Correlated set
R(0,1)	R(0,1)
R(1,1)	R(1,1)
R(2,1)	R(2,1)
R(3,1)	R(3,1)
R(11,1)	R(6,1)
R(4,1)	R(5,1)
R(5,1)	R(4,1)
R(8,1)	R(0,2)
R(10,1)	R(1,2)
R(9,1)	R(3,2)

Back

Variable importance

Large set of correlated return variables

One-month returns	Correlated set
R(0,1)	R(0,1)
R(1,1)	R(1,1)
R(2,1)	R(2,1)
R(3,1)	R(3,1)
R(11,1)	R(6,1)
R(4,1)	R(5,1)
R(5,1)	R(4,1)
R(8,1)	R(0,2)
R(10,1)	R(1,2)
R(9,1)	R(3,2)

Back

Transaction costs

- Past-return based strategies have high turnover (De Groot et al, 2012), Frazzini et al (2013)
- Trading costs depend on investor type (Frazzini et al, 2013)
- We use the estimates from Frazzini et al to approximate the costs for an institutional investor
- In principle, these could be reduced by following an (s,S)-type strategy as in Novy-Marx and Velikov (2014)

Back

Transaction costs

Table: Turnover and trading costs

	Low	2	3	4	5	6	7	8	9	High	High-Low	9-2	8-3
Turnover (monthly)	1.56	1.74	1.76	1.78	1.78	1.8	1.78	1.76	1.76	1.62	3.18	3.5	3.52
Trading cost (annual)	3.71	4.09	4.13	4.17	4.17	4.21	4.17	4.13	4.13	3.83	7.13	7.81	7.85
Gross return (annual)	-6.18	2.55	5.41	7.19	8.47	10.03	11.88	12.82	15.66	23.29	31.37	12.82	7.06
Net return (annual)	-9.88	-1.54	1.28	3.02	4.30	5.82	7.71	8.69	11.53	19.46	24.24	5.01	-.79

Back

Robustness: Predictive models

Estimate deep conditional portfolio sort with one-month returns over most recent half year only

	(1)	(2)	(3)	(4)
Intercept	2.15	2.12	2.17	2.04
	(19.09)	(18.47)	(19.35)	(17.37)
MKT		0.03	0.01	0.03
		(1.38)	(0.46)	(1.64)
SMB			0.04	0.04
			(0.65)	(0.85)
HML			-0.08	-0.04
			(-1.52)	(-0.85)
UMD				0.13
				(3.80)
R^2		0.00	0.02	0.08
IR		3.38	3.46	3.37

Back

FMB: Short-horizon past returns

Average return of 1.58 percent per month

	(1)	(2)	(3)	(4)
Intercept	1.58	1.45	1.57	1.35
	(9.29)	(8.60)	(9.44)	(7.08)
MKT		0.14	0.08	0.13
		(3.11)	(1.92)	(3.18)
SMB		0.09	0.09	
			(1.01)	(1.27)
HML		-0.21	-0.14	
			(-2.21)	(-1.91)
UMD				0.22
				(2.59)
R^2		0.04	0.08	0.16
IR		1.49	1.65	1.48

Back