Machine learning applied to accounting variables yields the risk-return metrics of private company portfolios

Flavio Abdenur (SLQ)
Elias Cavalcante (FEA-USP)
Rodrigo de Losso (FEA-USP)

WAIAF (ITA)
2019
16/05/2019

SLQ } SOLUÇÕES QUANTITATIVAS
www.slq.com.br

{Roteiro

- A Fronteira Eficiente e a Carteira Ótima
- O Caso de Empresas Abertas
- O Caso de Empresas Fechadas: Métodos Tradicionais
- O Caso de Empresas Fechadas: Machine Learning
- Resultados
- Possíveis Aplicações

SLQ}

{Roteiro

- A Fronteira Eficiente e a Carteira Ótima
- O Caso de Empresas Abertas
- O Caso de Empresas Fechadas: Métodos Tradicionais
- O Caso de Empresas Fechadas: Machine Learning
- Resultados
- Possíveis Aplicações

Trabalho em conjunto com Elias Cavalcante (FEA-USP) e Rodrigo de Losso (FEA-USP)

SLQ}

A Fronteira Eficiente e a Carteira Ótima

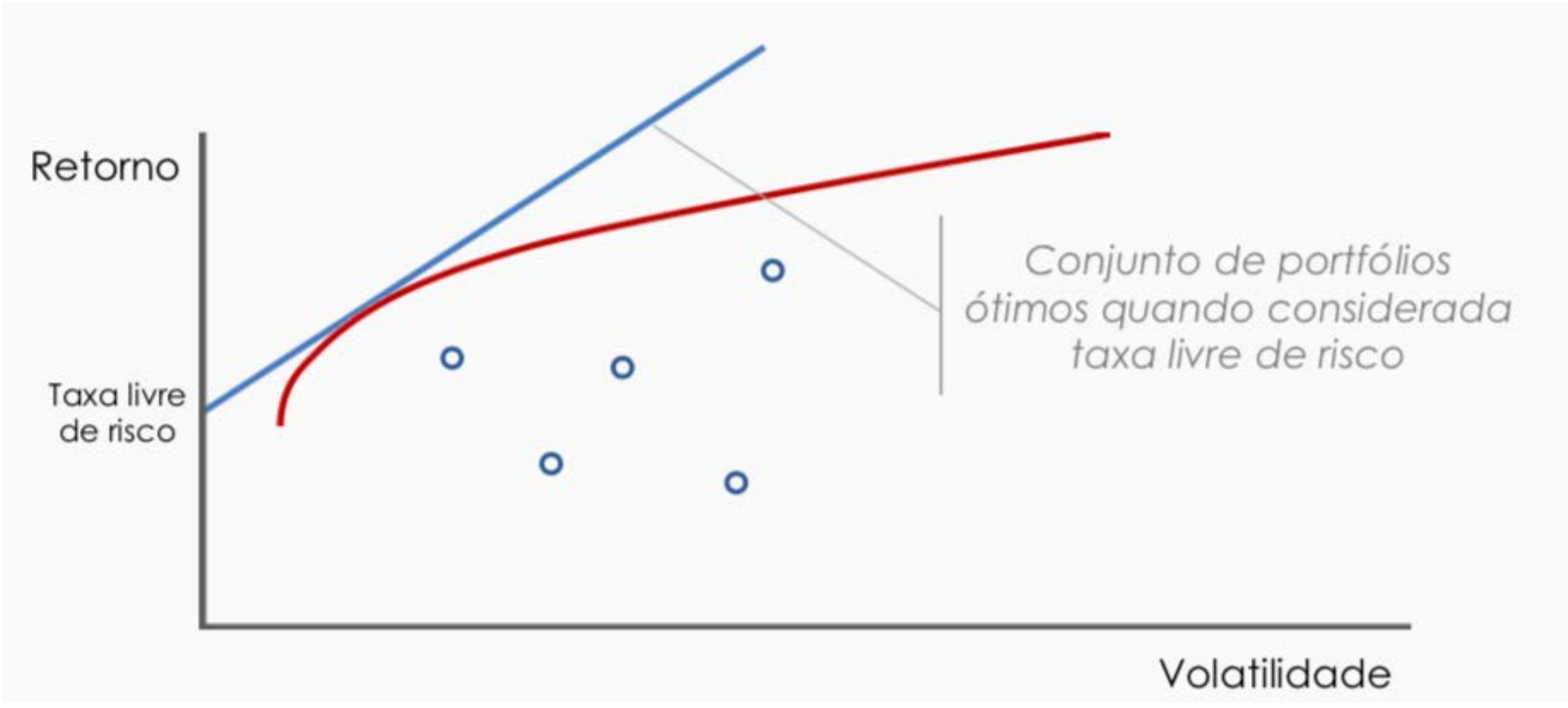


Figura por Elias Cavalcante

SLQ}

A Fronteira Eficiente e a Carteira Ótima

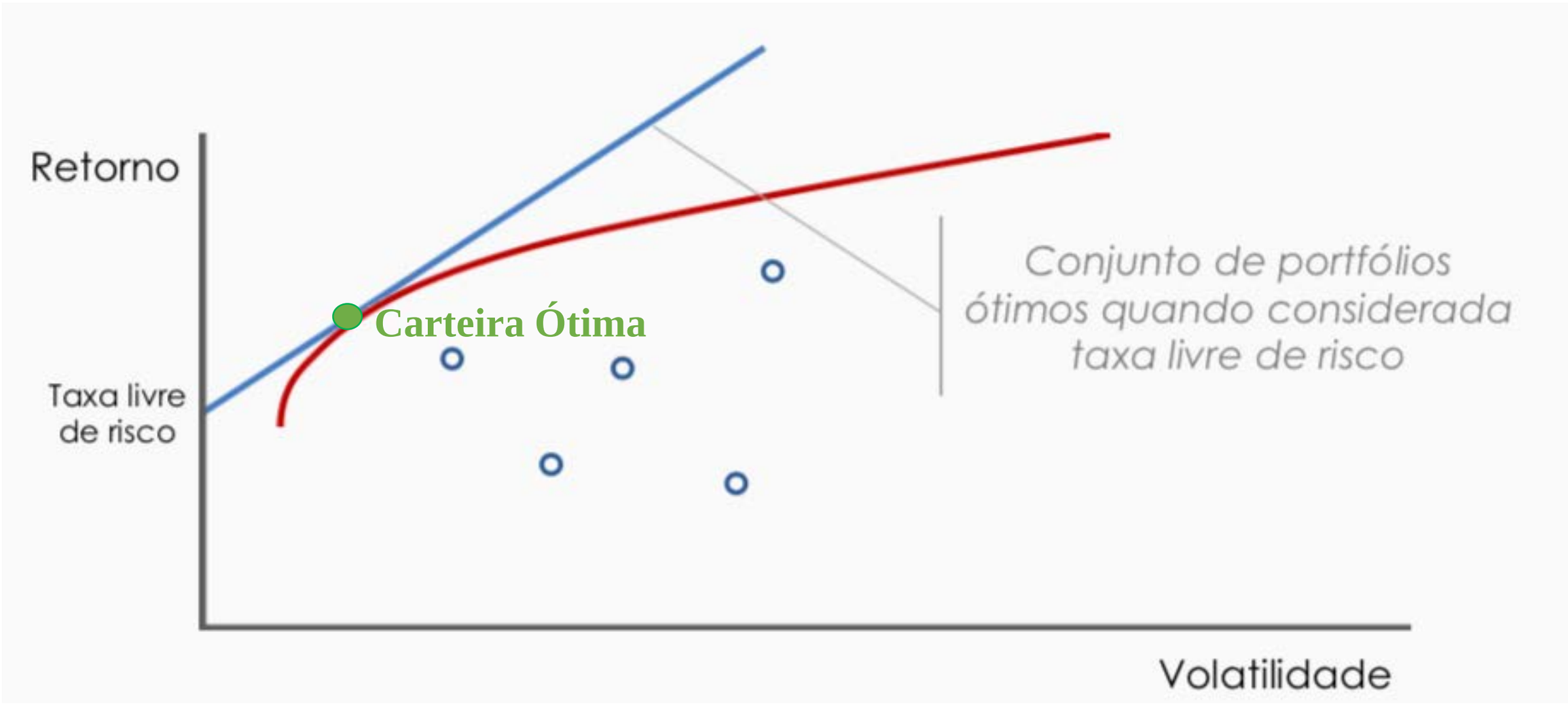


Figura por Elias Cavalcante

SLQ}

O Caso de Empresas Abertas:

Dada uma cesta $C_A = \{A_1, ..., A_n\}$ de empresas abertas, como obter a carteira ótima?

O Caso de Empresas Abertas:

Dada uma cesta $C_A = \{A_1, ..., A_n\}$ de empresas abertas, como obter a carteira ótima?

- Temos séries de preços (diários) publicamente disponíveis, logo obtemos diretamente volatilidade, retorno (em excesso) e covariâncias

O Caso de Empresas Abertas:

Dada uma cesta $C_A = \{A_1, ..., A_n\}$ de empresas abertas, como obter a carteira ótima?

- Temos séries de preços (diários) publicamente disponíveis, logo obtemos diretamente volatilidade, retorno (em excesso) e covariâncias
- Teoria de Markowitz permite construção de fronteira eficiente e portanto da carteira ótima lançando mão de retornos, volatilidades e covariâncias das ações

SLQ}

O Caso de Empresas Abertas:

Dada uma cesta $C_A = \{A_1, ..., A_n\}$ de empresas abertas, como obter a carteira ótima?

- Temos séries de preços (diários) publicamente disponíveis, logo obtemos diretamente volatilidade, retorno (em excesso) e covariâncias

- Teoria de Markowitz permite construção de fronteira eficiente e portanto da carteira ótima lançando mão de retornos, volatilidades e covariâncias das ações

- **Obs: “**desempenho passado não garante desempenho futuro”

O Caso de Empresas Abertas:

Dada uma cesta $C_A = \{A_1, ..., A_n\}$ de empresas abertas, como obter a carteira ótima?

- Temos séries de preços (diários) publicamente disponíveis, logo obtemos diretamente volatilidade, retorno (em excesso) e covariâncias

- Teoria de Markowitz permite construção de fronteira eficiente e portanto da carteira ótima lançando mão de retornos, volatilidades e covariâncias das ações

- **Obs: “**desempenho passado não garante desempenho futuro”... mas tudo bem

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima?

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos “artesanais” de *valuation* geram estimativas de valor

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos “artesanais” de *valuation* geram estimativas de valor, mas não séries diárias

SLQ}

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos “artesanais” de *valuation* geram estimativas de valor, mas não séries diárias
- Escolha de *peers* / “empresas comparáveis” gera as estatísticas

SLQ}

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos “artesanais” de *valuation* geram estimativas de valor, mas não séries diárias
- Escolha de *peers* / “empresas comparáveis” gera as estatísticas, mas é muito subjetiva

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos “artesanais” de *valuation* geram estimativas de valor, mas não séries diárias
- Escolha de *peers* / “empresas comparáveis” gera as estatísticas, mas é muito subjetiva
- “Controle sintético” é elegante

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos “artesanais” de *valuation* geram estimativas de valor, mas não séries diárias
- Escolha de *peers* / “empresas comparáveis” gera as estatísticas, mas é muito subjetiva
- “Controle sintético” é elegante, mas manipulável

O Caso de Empresas Fechadas:

Dada uma cesta $C_F = \{F_1, ..., F_n\}$ de empresas fechadas, como obter a carteira ótima? – não há preços diários!

Problema se reduz a: como obter retornos, vols, covs para gerar carteira ótima?

- Métodos "artesanais" de *valuation* geram estimativas de valor, mas não séries diárias
- Escolha de *peers* / "empresas comparáveis" gera as estatísticas, mas é muito subjetiva
- "Controle sintético" é elegante, mas manipulável (vide Ferman-Cristine-Possebom 2018)

SLQ}

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

- Base de ≈ 1000 empresas abertas de diversos setores, cerca de 77% nos EUA e 23% no BR

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

- Base de ≈ 1000 empresas abertas de diversos setores, cerca de 77% nos EUA e 23% no BR (obs: small- e large-caps foram purgadas)

SLQ}

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

- Base de ≈ 1000 empresas abertas de diversos setores, cerca de 77% nos EUA e 23% no BR (obs: small- e large-caps foram purgadas)

- Séries trimestrais de ≈ 55 características e variáveis contábeis de cada empresa entre 2005 e 2016, via Economática

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

- Base de ≈ 1000 empresas abertas de diversos setores, cerca de 77% nos EUA e 23% no BR (obs: small- e large-caps foram purgadas)

- Séries trimestrais de ≈ 55 características e variáveis contábeis de cada empresa entre 2005 e 2016, via Economática

- Preços (e portanto vols, rets e corrs) e juros, via Bloomberg / Nefin

{Machine Learning: Taxonomia

Variables	Description
Country	Country in which Asset is Traded
Sector	Specific Activity of Asset
FxA_to_Equity	Fixed Asset/Equity
Leverage	Total Asset/Equity
Leverage_smoo4	Mean total assets last 4 quarters/Mean Equity last 4 quarters
Leverage_smoo8	Mean Total Asset 8 quarters/Mean PL 8 quarters
TotAssets	Asset total
CapEmpl	Capital employed (Total Asset - Current Liability + Total Debt Short and Long term)
WorkingCap	Working Capital (Working Assets - Current Liability)
GrossDebt_to_Asset	Gross Debt/Total Asset
GrossDebt_to_Ebitda	Gross Debt/Ebitda
GrossDebt_to_Equity	Gross Debt/Equity
DivTlBr	Gross Debt (Total Debt Short and Long Term)
Net Debt	Net Debt (Total Debt - Cash assets - Capital Investments)
Debt_to_Equity	Net Debt/(Equity + Minority Shareholder Participation)
EBIT	Earnings Before Interest and Taxes
EBIT_to_NetDebt	EBIT/Net Debt
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortization
CapStru	Capital Structure [Total Gross Debt/(Total Gross Debt + Equity)]
Exg_to_TotA	(Total Asset - Equity - Minority Shareholder Participation)/Total Asset
ExgyTt	(Total Asset - Equity - Minority Shareholder Participation)
Exig_to_Equity	(Total Asset - Equity - Minority Shareholder Participation)/Equity
CFL	Corporate Financial Leverage
CFL_smoo4	Corporate Financial Leverage (mean 4 quarters)
CFL_smoo8	Corporate Financial leverage (mean 8 quarters)
COL	Corporate Operational Leverage
COL_CFL	GAO x GAF
COL_CFL_smoo4	GAO_smoo4 x GAF_smoo4
COL_CFL_smoo8	GAO_smoo8 x GAF_smoo8
COL_smoo4	Corporate Operational Leverage (mean 4 quarters)
GAO_smoo8	Corporate Operational Leverage (mean 8 quarters)
PIC	Permanent Investment Capital
InvestCap	Total Asset - Current Liability + Total Debt Short Term - Capital investments – Cash assets
IT	Income Tax
preTxprofit	Pre-Tax profit
preTxprofitFE	Pre-Tax Profit + Financial Expenses
LiqCor	Current Liquidity
Profit	Profit
Profit_to_Revenue	Net Profit/Revenue
Profit_to_Revenue_smoo4	Revenue (mean 4 quarters)/Revenue (mean 4 quarters)
Profit_to_Revenue_smoo8	Revenue (mean 8 quarters)/Revenue (mean 8 quarters)
NetProfit	Net Profit
ProfitCOp	Profit of Continued Operations
NetMargin	(Net Profit + Minority Shareholder Participation)/(Net Revenue)
MarginEBIT	EBIT Margin (EBIT/Net Revenue)
MarginEbitda	EBITDA Margin (EBITA/Net Revenue)
Payout0	Payout/Revenue
Payout0_smoo4	Payout (mean 4 quarters)
Payout0_smoo8	Payout (mean 8 quarters)
Equity	Equity
earnings0	Dividends + Payment of interest on Shareholders' Equity
Revenue	Revenue
Revenue_to_At	Revenue/Assets
Revenue_to_At_smoo4	Revenue/Assets (mean 4 quarters)
Revenue_to_At_smoo8	Revenue/Assets (mean 8 quarters)
Yield_end	Profit/Equity (end of the period)
Yield_begin	Profit/Equity (begin of the period end period)
Yield_middle	Profit/Equity (middle of the period end period)
Profitability	Profit/Assets
ROI	Return on Investment
ROI_smoo4	Return on Investment (mean 4 quarters)
ROI_smoo8	Return on Investment (mean 8 quarters)
ROIC_middle	[(1-Income Tax Rate)*EBIT]/Invested Capital (middle)

Variáveis incluem:

- País
- Setor
- EBIDTA
- Alavancagem financeira
- ROI
- etc etc

SLQ}

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

- Para cada uma das variáveis v ∈ {ret, vol, corr}, meta é obter função f_v({variáveis contábeis}) → $\mathbb{R}$

Machine Learning:

Abordagem via ML, usando companhias abertas como “proxies” de fechadas:

- Para cada uma das variáveis v ∈ {ret, vol, corr}, meta é obter função

$$f_v(\{\text{variáveis contábeis}\}) \rightarrow \mathbb{R}$$

Train / test split e tuning:

- Split 70% train / 30% test, balanceado por classes (país e setor)
- Tuning via 5-fold CV (no training set)

SLQ}

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

- Por um lado, isso “achatou” as séries em 1 observação por empresa

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

- Por um lado, isso “achatou” as séries em 1 observação por empresa
- Por outro, gerou algo da ordem de 14 x 50 = 700 *features...*

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

- Por um lado, isso “achatou” as séries em 1 observação por empresa
- Por outro, gerou algo da ordem de 14 x 50 = 700 *features...*

o que traz a necessidade de:

SLQ}

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

- Por um lado, isso “achatou” as séries em 1 observação por empresa
- Por outro, gerou algo da ordem de 14 x 50 = 700 *features...*

o que traz a necessidade de:

Feature Selection:

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

- Por um lado, isso “achatou” as séries em 1 observação por empresa
- Por outro, gerou algo da ordem de 14 x 50 = 700 *features...*

o que traz a necessidade de:

Feature Selection:

- LASSO para lineares e semi-lineares

Machine Learning:

Feature Engineering:

14 transformações (e.g., desvio padrão, taxa de crescimento, etc) foram aplicadas a cada uma das variáveis:

- Por um lado, isso “achatou” as séries em 1 observação por empresa
- Por outro, gerou algo da ordem de 14 x 50 = 700 *features...*

o que traz a necessidade de:

Feature Selection:

- LASSO para lineares e semi-lineares
- Boruta para baseados-em-árvores (e.g, random forest e GBM)

Machine Learning:

Modelos testados:

- SVMs, LASSO, Ridge, elastic net (= LASSO + Ridge), RF, GBM

Machine Learning:

Modelos testados:

- SVMs, LASSO, Ridge, elastic net (= LASSO + Ridge), RF, GBM
- No final, ensembling (nos folds)

Machine Learning:

Modelos testados:

- SVMs, LASSO, Ridge, elastic net (= LASSO + Ridge), RF, GBM
- No final, ensembling (nos folds)
- Usamos OLS como benchmark

Machine Learning:

Modelos testados:

- SVMs, LASSO, Ridge, elastic net (= LASSO + Ridge), RF, GBM
- No final, ensembling (nos folds)
- Usamos OLS como benchmark

Obs:

- *Hipótese implícita: há comportamento semelhante de empresas abertas e fechadas*

Machine Learning:

Modelos testados:

- SVMs, LASSO, Ridge, elastic net (= LASSO + Ridge), RF, GBM
- No final, ensembling (nos folds)
- Usamos OLS como benchmark

Obs:

- *Hipótese implícita: há comportamento semelhante de empresas abertas e fechadas*
- ***Não** é forecasting; está mais para backcasting*

Machine Learning:

Modelos testados:

- SVMs, LASSO, Ridge, elastic net (= LASSO + Ridge), RF, GBM
- No final, ensembling (nos folds)
- Usamos OLS como benchmark

Obs:

- *Hipótese implícita: há comportamento semelhante de empresas abertas e fechadas*
- ***Não** é forecasting; está mais para backcasting*
- *Moralmente, combina "empresas comparáveis" com "controle sintético"*

Resultados

- Vencedor foi RF (random forest)

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

- RF (resp., OLS) teve seguintes R^2s:

  - vol:

  - ret:

  - corr:

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

- RF (resp., OLS) teve seguintes R^2s:

  - vol: 50%

  - ret:

  - corr:

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

- RF (resp., OLS) teve seguintes R^2s:

  - vol: 50% (15%)

  - ret:

  - corr:

SLQ}

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

- RF (resp., OLS) teve seguintes R^2s:

  - vol: 50% (15%)

  - ret: 40%

  - corr:

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

- RF (resp., OLS) teve seguintes R^2s:

  - vol: 50% (15%)

  - ret: 40% (5%)

  - corr:

SLQ}

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

- RF (resp., OLS) teve seguintes R^2s:

  - vol: 50% (15%)

  - ret: 40% (5%)

  - corr: 45%

SLQ}

Resultados

- Vencedor foi RF (random forest)

(A rigor foi ensemble de RF com GBM, mas pouco incremento de desempenho e muito aumento de custo computacional...)

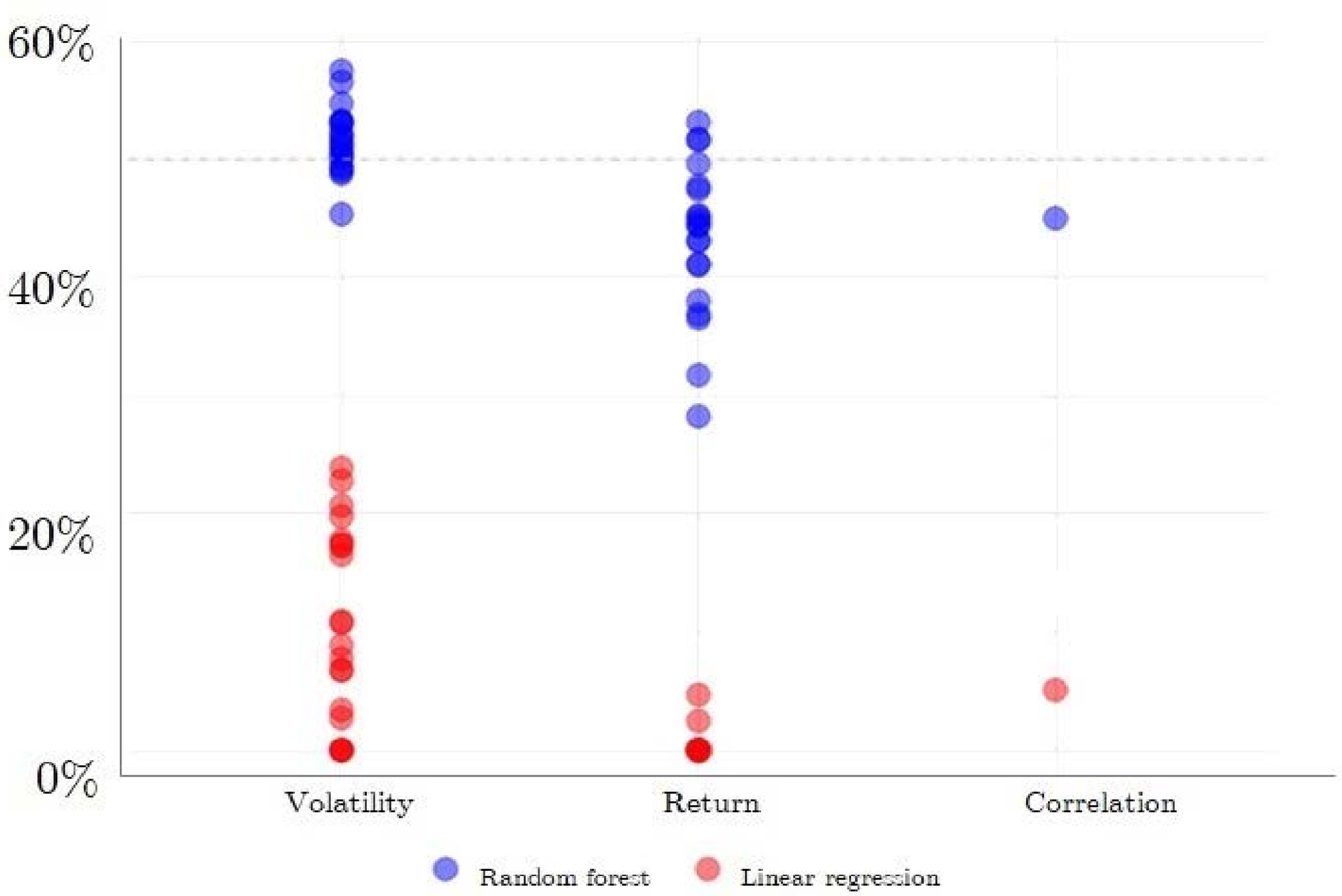
- RF (resp., OLS) teve seguintes R^2s:

  - vol: 50% (15%)

  - ret: 40% (5%)

  - corr: 45% (5%)

Resultados



SLQ}

Machine Learning:

Importância das features varia muito com o modelo:

Rank	Return		Volatility		Correlation	
	Variable	Importance	Variable	Importance	Variable	Importance
1	Country	100.00	Profitability.Dol.cummean	100.00	Sector	100.00
2	Exg_to_TotA.Dol.diff.cummean	95.86	Yield_end.Dol.cummean	89.87	WorkingCap.Dol.rollmean8	56.64
3	ExgvTt.Dol.diff.rollmean8	89.05	ProfitCOp.Dol.rollmean4	88.38	ROIC_middleio.Dol.cummean	40.42
4	Profitability.Dol.diff.cummean	86.39	Profit.Dol.rollmean4	85.66	Payout0.Dom.Delt.cummean	37.13
5	CapEmpl.Dol.Delt.rollmean8	81.91	Yield_middle.Dol.cummean	80.89	Revenue.Dol.Delt.rollmean8	31.13
6	PIC.Dol.Delt.rollmean8	76.60	ProfitCOp.Dol.rollmean8	62.15	TotAsset.Dol.Delt.rollmean8	30.75
7	TotalAsset.Dol.Delt.rollmean8	72.51	Yield_begin.Dol.cummean	60.80	Revenue.Dol.diff.rollmean8	29.85
8	Revenue.Dol	60.94	preTxprofit.Dol.rollmean4	58.72	MarginEbitida.Dol.rollmean8	28.77
9	Yield_middle.Dol.cummean	60.82	NetProfit.Dol.rollmean4	58.65	Yield_end.Dol.rollmean4	27.51
10	Equity.Dol	54.45	Profit.Dol.rollmean8	56.03	Country_USBR	27.48
11	Revenue.Dol.rollmean4	53.90	Payout0.Dol.rollmean4	55.83	LiqCor.Dol.cummean	27.13
12	CapEmpl.Dol	53.62	Payout0.Dol.rollmean8	55.22	FxA_to_Equity.Dol.rollmean8	27.13
13	Equity.Dol.rollmean4	51.38	EBITDA.Dol.rollmean4	53.84	Pais_US	27.05
14	Yield_begin.Dol.cummean	48.91	NetMargin.Dol.cummean	51.22	MarginEbitida.Dol	26.58
15	ExgvTt.Dol.Delt.rollmean8	48.09	preTxprofit.Dol.rollmean8	50.73	FxA_to_Equity.Dol.rollmean4	26.54
16	ExgvTt.Dol.rollmean8	47.45	NetProfit.Dol.rollmean8	48.30	CapEmpl.Dol.Delt.rollmean8	25.92
17	Profitability.Dol.diff.rollmean8	47.33	EBITDA.Dol.rollmean8	38.58	CapStrc.Dol.cummean	25.29
18	TolAsset.Dom.Delt.rollmean4	44.22	Payout0_suaav4.Dol	34.96	MarginEbitda.Dol.rollmean4	25.15
19	TolAsset.Dol	43.90	Profitability.Dol.rollmean8	34.72	MarginEbitida.Dol.cummean	25.07
20	InvestCap.Dol.Delt.rollmean8	42.74	EBIT.Dol.rollmean8	33.84	PIC.Dom.Delt.rollmean4	24.95

SLQ}

Aplicações

Aplicações

- Auxílio de tomada de decisões (de venda e aquisição de empresas fechadas)

Aplicações

- Auxílio de tomada de decisões (de venda e aquisição de empresas fechadas)
- Forecasting (ao invés de backcasting)?

Aplicações

- Auxílio de tomada de decisões (de venda e aquisição de empresas fechadas)

- Forecasting (ao invés de backcasting)?

- Private equity?

Aplicações

- Auxílio de tomada de decisões (de venda e aquisição de empresas fechadas)

- Forecasting (ao invés de backcasting)?

- Private equity? (neste caso variável-alvo natural seria valuation)