Department of Economics - FEA/USP

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

ELIAS CAVALCANTE-FILHO
FLAVIO ABDENUR
RODRIGO DE LOSSO

WORKING PAPER SERIES Nº 2018-23

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2018-23

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

Elias Cavalcante-Filho (e.cavalcante@usp.br)

Flavio Abdenur (flavio.abdenur@gmail.com)

Rodrigo De Losso (delosso@usp.br)

Abstract:

Constructing optimal Markowitz Mean-Variance portfolios of publicly-traded stock is a straighforward and well-known task. Doing the same for portfolios of privately-owned firms, given the lack of historical price data, is a challenge. We apply machine learning models to historical accounting variable data to estimate risk-return metrics – specifically, expected excess returns, price volatility and (pairwise) price correlation – of private companies, which should allow the construction of Mean-Variance optimized portfolios consisting of private companies. We attain out-of-sample R^2 s around 45%, while linear regressions yield R^2 s of only about 10%. This short paper is the result of a real-world consulting project on behalf of Votorantim S.A ("VSA"), a multinational holding company. To the authors' best knowledge this is a novel application of machine learning in the finance literature.

Keywords: assent pricing; Machine Learning; Portfolio Theory

JEL Codes: G12, G17.

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

Elias Cavalcante-Filho (*University of São Paulo*)
Flavio Abdenur (*SLQ Soluções Quantitativas*)
Rodrigo De-Losso (*University of São Paulo*)†

Abstract

Constructing optimal Markowitz Mean-Variance portfolios of *publicly-traded* stock is a straighforward and well-known task. Doing the same for portfolios of *privately-owned* firms, given the lack of historical price data, is a challenge. We apply machine learning models to historical accounting variable data to estimate risk-return metrics – specifically, expected excess returns, price volatility and (pairwise) price correlation – of private companies, which should allow the construction of Mean-Variance optimized portfolios consisting of private companies. We attain out-of-sample R^2s around 45%, while linear regressions yield R^2s of only about 10%.

This short paper is the result of a real-world consulting project on behalf of Votorantim S.A ("VSA"), a multinational holding company. To the authors' best knowledge this is a novel application of machine learning in the finance literature.

Keywords: Asset pricing, Machine Learning, Portfolio Theory.
JEL Codes: G12, G17.

1 Introduction

The Markowitz/Mean-Variance framework is traditionally used to measure and optimize the risk-return profiles of portfolios of publicly-traded assets. This approach, when applied to publicly traded equity stock, classically relies on historical price data, which is unavailable for portfolios of *privately* held (and so non-publicly traded) firms.

We circumvent this obstruction by applying supervised machine learning models to historical *accounting* variables to obtain the risk-return profiles of a large set of companies in the U.S. and Brazil. Once trained, these models can be applied to estimate the risk-return profiles of private companies, which in turn allows the construction of optimal portfolios of private companies in the sense of Markowitz Portfolio Theory.

We emphasize that this study is the by-product of a "real-world" consulting project undertaken on behalf of Votorantim S.A., a Brazil-based multinational holding company almost all of whose controlled companies are privately held. Once duly refined and adapted to specific sectors and regions, such models may potentially aid institutions that seek to construct optimized portfolios of private companies (as opposed to standard portfolios of stock-market equity).

*We thank the Votorantim team for very constructive suggestions and comments.

†E-mail addresses: e.cavalcante@usp.br (Cavalcante-Filho E.), flavio.abdenur@gmail.com (Abdenur F.), delosso@usp.br (De-Losso R.).

2 Economic background

The core of Modern Portfolio Theory is based on the work of Markowitz (1952), which gave rise to the class of Mean-Variance (MV) type models.

MV models analyse the behavior of a risk-averse investor over a finite time horizon. Starting from a given initial allocation the investor chooses a portfolio by selecting assets and their respective quantities from a set of N distinct available assets. This decision is based on the investor's knowledge regarding *(i)* the expected returns and *(ii)* the variances and covariances of the available assets. The investor builds the portfolio so as to maximize expected returns for a given level of risk.

Formally, the model supposes that the agent follows a utility function (U):

$$U(E(R_{p,t}), \sigma^2) = E(R_{pt}) - \gamma \times \sigma^2 \tag{1}$$

such that $E(R_{p,t})$ is the expected return of the portfolio p at time t, γ is the risk-averse parameter, σ^2 is the risk parameter measure as the variance of $E(R_{p,t})$. Since the portfolio p consists of a basket of assets, the parameters $E(R_{p,t})$ and σ^2 are defined as follows:

$$E(R_p) = \begin{bmatrix} w_1 & w_2 & \dots & w_n \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix} \tag{2}$$

$$\sigma_p^2 = \begin{bmatrix} w_1 & \dots & w_N \end{bmatrix} \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1N} \\ \sigma_{21} & \dots & \sigma_{2N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \dots & \sigma_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \tag{3}$$

where: w_i is the weight of the asset $i \in 1, 2, ...N$ in the portfolio p; $E(R_i)$ the expected returns the asset i; and σ_{ij} the covariance between assets i and j.

The goal is to maximize utility by selecting the optimal weights of the available assets to create the portfolio p; that is, to find the weights vector w that maximizes $U(E(R_{p,t}), \sigma^2)$.

While the basic insight underpinning the MV models is relatively simple, their implementation presents some practical problems, relying as it does on the estimation of $E(R_i)$ and σ_{ij}. When dealing with publicly traded assets the standard approach is to turn to the historical data of market returns to estimate risk and return parameters such as expected excess returns and price volatility. This approach, however, is obviously not feasible for non-traded assets such as private companies.

In light of this problem, considerable empirical research has been directed to the relationship between, on one hand, financial and accounting variables, and, on the other hand, market-based measures of risk. This literature indicates that some financial (i.e, accounting) variables are highly correlated with historical market-based measures of risk and are moreover useful for the prediction of future returns (Bowman, 1979; Laveren et al., 1997; Almisher and Kish, 2000; Amorim et al., 2012; Teixeira and do Valle, 2009; Brimble and Hodgson, 2007; Neto and Bruni, 2008).

There are essentially two distinct approaches to estimate market-based measures of risk and return of a non-publicly traded firm. The first one, called the accounting approach, relies on the estimation of risk and return using accounting variables and other non-market data. A second approach uses subjectively "similar" public companies as proxies for a given private firm's parameters. This second approach is commonly called the *comparable company approach*. This might involve relying on an industry beta rather than on a single firm or small set of firms.

This paper lays out a method for estimating the parameters $E(R_i)$ and σ_{ij} of private firms based only solely non-market data via machine learning models. In a sense this methodology fuses the accounting approach (insofar as it relies on acconting data) with the comparable company approach (since it uses models trained on publicly-traded firms to estimate the parameters privately-held firms, so that the first essentially play the role of proxies for the second).

3 Methodology

We first estimate the desired parameters (excess return and volatility) of *public* companies by finding functions which empirically model the relationship between the parameters and the values of a large set of accounting variables which are periodically disclosed by the companies. Namely, the following functions are estimated:

$$E(R_i) = f_R(AccountingVariables)) \quad (4)$$

$$\sigma_i = f_\sigma(AccountingVariables)) \quad (5)$$

$$\rho_{i,j} = f_\rho(AccountingVariables)) \quad (6)$$

where $E(R_i)$ is the expected return (in excess of the risk-free rate) of asset i; σ is the standard deviation (a.k.a, volatiity) of the returns of asset i; and ρ is the pairwise correlation between the returns of assets i and j.

The relationships linking the accounting variables with the parameters of interest were estimated using quarterly data from companies traded on the stock exchange; the resulting models are intended to be later applied to private companies. To estimate the functions for expected return, volatility, and correlation we used linear regression methods and tested several machine learning method, namely Elastic Net, Extreme Gradient Boosting, LASSO, Recursive Partitioning and Regression Trees, and Random Forest.[3] The linear regression model was used as benchmark; among the machine learning methods we chose the one with the a posteriori best performance, taking into account both error margins and computational cost.

We followed strict machine learning protocols. Training (about 70% of the sample) and testing (about 30% of the sample) sets were carefully separated. 5-fold cross validation was used for model tuning.

Performance was evaluated in terms of out-of-sample R^2 and RMSE. There were essentially two phases in the project. In the first phase, the testing set was chosen randomly. In the second that, once the predictive power of the trained models had been ascertained, the models were applied to a set of "peer" companies that had been hand-picked by the VSA team.

Remark: A third phase, where the models would be applied to VSA's own private companies, would involve proprietary data which cannot be published or otherwise disclosed.

[3]Kuhn and Johnson (2013); Abu-Mostafa et al. (2012); Breiman (2001); Chen et al. (2015).

4 Data

The raw variables used consist of price series of companies together with the series of their respective accounting variables (which were used as the predictive features). The raw prices were used to compute our ultimate target variables: excess returns, volatility and pairwise return correlations. In both the machine learning models and the linear regressions the accounting variables were used as the predictors, with excess returns, volatility and pairwise return correlations as the targets, so that three separate models were trained, one for each target. All of the dataset variables were gathered from the Economatica financial software service, except for the risk-free rates series, which came from Nefin[4] The data start in 01/01/2005 and end in 31/12/2016. The three target variables were computed using price data from the entire period.

Table 1 lists the raw variables used, and sets out brief descriptions of each one.

[Table 1 about here.]

Fourteen mathematical transformations, listed in Table 2, were applied to each of the raw accounting variables. (This constitutes this project's "feature engineering", in machine learning parlance.) Moreover, in order to deal with outliers that might bias the models, we excluded from the dataset the 10% least traded firms from both countries, as well as the smallest and largest 10% of companies by market cap. Companies that are exceptionally illiquid, small, or large are known to have unusual behavior. Moreover, the private companies that our client ultimately desired to model are largely mid-cap. The resulting dataset is summarized below in Tables 3 e 4.

[Tables 2, 3 and 4 about here.]

5 Results

This chapter lays out the results into three parts: *(i)* the lists of the relevant variables for each of the three models, with their respective measures of explanatory power; *(ii)* the overall results, in terms of R^2 and average error, of the Mean Return, Volatility and Correlation models, always measured out-of-sample (i.e., in the testing set), and *(iii)* the validation, in terms of Average Return and Volatility, of the 22 peer companies chosen by the VSA team.

5.1 Variable Importance

The original database, because of the application of the aforementioned transformations of the raw variables, contained a large number of potential explanatory variables: 541 explanatory variables, against 1,035 observations (i.e., firms).

To deal with this obstacle, we applied a variable selection algorithm, Boruta (Kursa et al., 2010), to this large set of features. This algorithm uses a logic specifically focused on random forests that examines one-by-one the variables of the database. Variables

[4]Brazilian Center for Research in Financial Economics of the University of São Paulo (NEFIN) – Núcleo de Pesquisas em Economia Financeira da Universidade de São Paulo <`http://nefin.com.br`>.

that satisfy a criterion of relevance (for RF) are maintained and the others are discarded. Once Boruta was applied, the following numbers of variables were left over from the 541 variables of the base:

- Volatility: 147 explanatory variables
- Expected Return: 134 explanatory variables
- Correlation: 151 explanatory variables

Although the decisions of random forest models are not easily interpreted, it is possible to extract relative importance "weights" for each of their variables. The following table shows the 20 most relevant explanatory variables of each model.

[Table 5 about here.]

5.2 Global Results

This section exhibits the "global" results - that is, the results in the entire test / out-of-sample set - for each of the models: Volatility, Expected Return and Correlation.

ML models may display some instability when calibrated. To ascertain robustness sensitivity in this respect, the Volatility and Average Return models were estimated 20 times each. What distinguished one estimate from the other was the subset of observations that were (randomly) placed in either the training or in the testing set. In the case of the model for Correlation, the estimation only performed once, since in this case the large number of observations (close to 500,000 pairs) made the training very cumbersome.

Figure 1 shows the R^2 of 20 iterations of the Volatility and Mean Return models, as well as the single iteration of the Correlation model. The dots in blue correspond to the R^2 of the ML models (specifically, the Random Forest), while the red dots correspond to the (much lower) performance of linear regressions applied to the same sets.

[Figure 1 about here.]

As can be seen in the figure, both the Volatility and Expected Return models are quite stable, ranging around 50% and 45% respectively. The corresponding linear regression models reach R^2 of the order of 10% for Volatility and 3% for Return.

The ML model of Correlation, calibrated only once because of its computational cost, reached an R^2 of about 45%, whereas the linear regression reached a R^2 of 15%. Due to the size of its training set (about 350,000 observations, corresponding to 70% of approximately 500,000 pairs) it is likely that the Correlation model is very stable, in the sense of consistently generating R^2 very close to this initial value.

Table 6 summarizes the results in terms of R^2 and RMSE of the three model types.

[Table 6 about here.]

5.3 Results for peer firms

The VSA team selected as peers, for the purpose of benchmarking the companies in its portfolio, a set of 22 open companies, each of which have characteristics in some (subjective) sense similar to one of the VSA companies. These are the so-called VSA *peer companies.*

Figures 2 and 3 illustrate the results obtained for return, volatility and correlation of the peer companies. The blue dots correspond to the values realized for each company and the green dots to the values predicted by the model. The average absolute error of the Volatility forecast among all peers is 3.13 p.p.. Once the three outliers are excluded, the mean absolute error goes down to 2.32 p.p.. The mean absolute error of the Return prediction among all peers is 1.2 p.p.. Once the two outliers have been removed, the mean absolute error becomes 0.97 p.p.. The results for Correlation follow below. Due to the large volume of observations for this model, there is no table with expected and observed values for all combinations two by two.

[Figures 2 and 3 about here.]

Remark 1: In addition to Random Forest, we experimented with several other machine learning models, such as LASSO, Elastic Net (glmnet), and Recursive Partitioning and Regression (RPart) Trees. Strictly speaking, the best performing model, as measured by RMSE and R^2, was actually a weighted ensemble of Extreme Gradient Boosting (xgbTree) and Random Forest. The xgbTree model, however, was much more computationally expensive than the RF, so due to time constraints the final model used was the Random Forest. The advantage in terms of error metrics of the xgbTree-Random Forest ensemble over the pure Random Forest model was very small and not worth the additional time expenditure. (Recall that this paper is the result of a paid consulting project, so time restrictions were important.)

Remark 2: All models were trained via the caret R package. The Random Forest version came from the "ranger" package and the Extreme Gradient Boosting model (see the following Section) came from the "xgboost" package. Linear regressions were implemented via the built-in "lm" model.

6 Conclusion

Machine learning models, in particular the Random Forest, consistently outperformed linear regression by a large margin in estimating the excess returns, volatilities, and pairwise return correlatons of companies. Roughly speaking, out-of-sample R^2 for estimating volatility and excess returns via Random Forests were respectively in the 40% and 50% ranges, while the corresponding R^2 s for linear regressions were around 10% and 3%. The models for pairwise price correlations yielded R^2 of 45% via Random Forest and 15% via linear regressions. These results are robust to resampling techniques and seem to work in very different regions, such as the U.S. and Brazilian economies. Duly adapted and refined, such models might guide institutions such as holding companies and private-equity funds in constructing optimal Mean-Variance private-firm portfolios. To our knowledge this is the first such application in the finance / machine learning literature. Further efforts in this direction tend to have both academic and financial value.

References

Abu-Mostafa, Y. S., M. Magdon-Ismail, and H.-T. Lin (2012). *Learning from data*, Volume 4. AMLBook New York, NY, USA:.

Almisher, M. A. and R. J. Kish (2000). Accounting betas - an ex anti proxy for risk within the IPO market. *Journal of Financial and Strategic Decisions 13*(3), 23–34.

Amorim, A. L. G. C., I. S. Lima, and F. D.-R. Murcia (2012). Análise da Relação entre as Informações Contábeis e o Risco Sistemático no Mercado Brasileiro/Analysis of the Relationship between Accounting Information and Systematic Risk in the Brazilian Market. *Revista Contabilidade \& Finanças 23*(60), 199.

Bowman, R. G. (1979). The theoretical relationship between systematic risk and financial (accounting) variables. *The Journal of Finance 34*(3), 617–630.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Brimble, M. and A. Hodgson (2007). Assessing the risk relevance of accounting variables in diverse economic conditions. *Managerial Finance 33*(8), 553–573.

Chen, T., T. He, M. Benesty, et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1–4.

Kuhn, M. and K. Johnson (2013). *Applied predictive modeling*, Volume 26. Springer.

Kursa, M. B., W. R. Rudnicki, et al. (2010). Feature selection with the boruta package. *Journal of Statistical Software 36*(11), 1–13.

Laveren, E., E. Durinck, M. de Ceuster, N. Lybaert, and Others (1997). Can accounting variables explain any beta? The empirical association between various betas and nine accounting variables in Belgian listed firms. Technical report.

Markowitz, H. (1952). Portfolio selection. *Journal of finance 7*(1), 77–91.

Neto, J. M. and A. L. Bruni (2008). Risco, retorno e equilíbrio: existe associação entre indicadores contábeis e os retornos das ações negociadas na Bovespa? *Gestão \& Planejamento-G\&P 1*(10).

Teixeira, S. C. and M. R. do Valle (2009). Associação entre beta contábil e beta de mercado: análise no mercado financeiro brasileiro. *RIC-Revista de Informação Contábil-ISSN: 1982-3967 2*(3), 1–21.

Table 1: Accounting Variables

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 Investiments)
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
ExgvTt	(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 - 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)

Table 2: Variable transformations

Label	Description
.Dol	dollar value
.Dom	values in domestic currency
.cummean	cumulative mean
.cumsd	cumulative standard deviation
.Delt	percentage variation
.Delt.cummean	cumulative mean of the past percentage variation
.Delt.cumsd	cumulative standard deviation of the past percentage variation
.Delt.rollmean4	mean of percentage variation of the past 4 quarters
.Delt.rollsd4	standard deviation of percentage variation of the past 4 quarters
.diff	difference from last quarter
.diff.cummean	Mean of difference from last quarter
.diff.cumsd	Standard Deviation of difference from last quarter
.diff.rollmean4	Mean of difference from last 4 quarters
.diff.rollsd4	Standard Deviation of difference from last 4 quarters
.rollmean4	Mean of last 4 quarters
.rollsd4	Standard Deviation of last 4 quarters

Table 3: Dataset Summary (Return and Volatility)

Country	N	Mean Return (% p. q.)	Volatility (% p. q.)	Capital Market (US$ millions)		
		Mean	Mean	Max	Mean	Min
Brazil	242	- 0,76	19,92	79.008,96	3.073,50	0,628
USA	793	3,00	14,45	617.588,49	20.533,09	30,434
Total	**1.035**	**2,12**	**15,73**	**617.588,49**	**16.450,074**	**0,628**

Table 4: Database Summary (Correlation)

N	Correlation (% p.q.)	Correlation (% p.q.)
	Mean	Standard Deviation
319.366	23,96	20,36

Figure 1: Stabiity of R^2 in the Testing Set

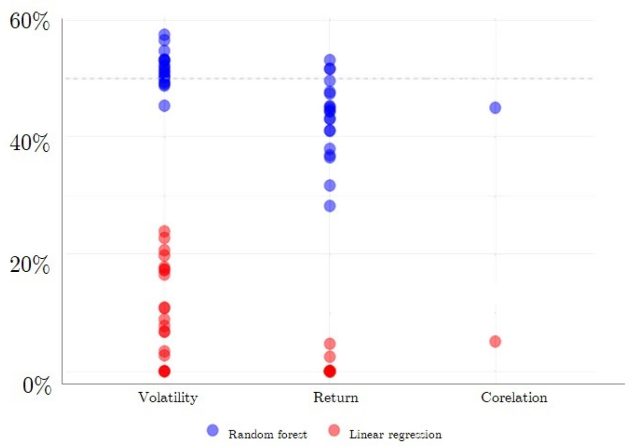

Table 5: Main Variables Used

Rank	Return Variable	Return Importance	Volatility Variable	Volatility Importance	Correlation Variable	Correlation 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

Table 6: Out of sample RMSE and R^2

Variable	Mean	Standard Deviation	RMSE	R^2
Volatility	15,73 p.p.	6,75 p.p.	~ 4,50 p.p,	~ 50%
Mean Return	2,12 p.p.	3,63 p.p.	~ 2,75 p.p.	~ 45%
Correlation	25,14 p.p.	20,99 p.p.	~ 0,12 p.p.	~ 45%

Figure 2: Predicted vs Realized Volatility (% p.q.)

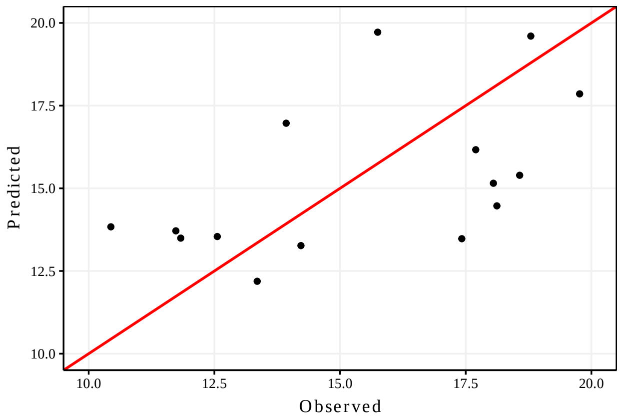

Figure 3: Predicted vs Realized Return (% p.q.)