AFP HABITAT

Private Equity Nowcasting

December 5, 2019

Executive Summary

The private equity market has been taking a more relevant role on many large portfolios. Normally private equity investors don't have access to the same level and quality of information compared to a public equity investor. We believe that gap can be reduced through the use of *Bayesian models* as well as other machine learning techniques including *advanced time series and features selection*. In this capstone project, we want to develop an effective decision-making framework via predictive models based on public information, in order to *understand the return of private assets and gain insights into private asset portfolios*.

Our aim is to boost the competitive analytics edge at AFP Habitat.

PART 01

Data Cleaning

PART 02

Feature Selection

PART 03

CAPM Model

PART 04

Multivariate LSTM Model

PART 05

CNN-Convolutional Neural Network Model

PART 06

Bayesian Model

PART 07

Key Takeaways

PART 08

Future Implication

Data Cleaning

Original Data

- **px_last_factors:** 37 variables
- **px_last_PE:** 20 variables
- **factors_data:** 380 variables

**missing values are imputed by previous and latest records*

Clean Data

- ***Time Period:*** *December 30th 2000 – December 30th 2018*
- ***Combined Variables:*** *435*
- ***Observations:*** *6,575 daily, 216 monthly*

Feature Selection

Private Equity

Real Estate

Buyout

Venture

Private Capital - Euro

Examine Variable Correlations

Run linear regression model to **explore statistical significance of regression coefficients** and variance inflating factor.

Select Highest Coefficients

Select variables with **p-value < 0.001**

Inspect Threat of Multicollinearity

Compute Variance Inflating Factor (VIF). VIF>10 indicates serious multicollinearity à **variables with VIF < 5 are chosen**

Test All Posissble Subsets

Subset Method: Hybrid Stepwise

Shrinking Method: Lasso

Regsubsets solution with **lowest cp**

Feature Selection Results

Macro Indices	Private Equity	Real Estate	Buyout	Venture	Private Capital-Europe
Argentina Price To Earningsn Ratio			x	x	
Atlanta Fed GDP Nowcast (2011)	x	(x)	x		x
Brazil LC Price To Earnings Ratio	(x)		(x)	x	
Europe Small Cap Price To Earnings Ratio		(x)			
Germany Price To Earnings Ratio		x			
India Total Return Index Gross Dividends					x
Japan Large Cap Price To Earnings Ratio				x	(x)
Japan Small Cap Price To Earnings Ratio		x			
Korea Price To Earnings Ratio	x		x		x
Private Equity Index		x			
Slope Chile	(x)				
USA Large Cap Blend DVD_SH_12M	x		x		
USA Large Cap Growth Closing Price				x	
USA Small Cap Growth Price To Earnings Ratio	x		x	x	
USA Small Cap Value DVD_SH_12M		x			

x : positive correlation
(x) : negative correlation

CAPM MODEL

Assumption

- Presume "Slope.CL" as the Risk-Free Rate
- Utilize variables based on the feature selection to avoid model overfitting

Results

Model	Data Usage	Language	TEST RMSE				
			Private Equity	Real Estate	Buyout	Venture	Private Capital-Europe
CAPM	Daily	R	22.4	29.01	33.54	8.78	41.62

While CAPM model is widely used, **the primary drawbacks are reflected in the model's inputs and assumptions**, including:

- Risk-Free Rate
- Return on the Market
- Ability to Borrow at a Risk-Free Rate

à **Inconsistency in testing results**

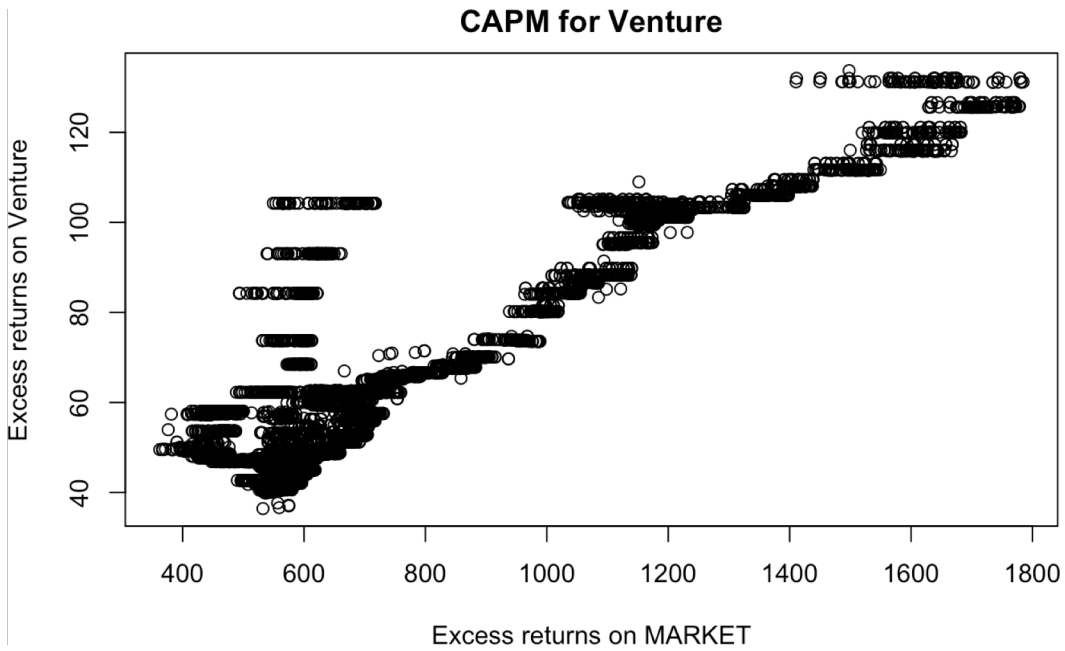
Example: Venture

Select Important Variables

```
[1] "VENTURE"                    "Slope.CL"                   "SGX.Index_PX_LAST"
[4] "RUO.Index_INDX_ADJ_PE"      "IBOV.Index_INDX_ADJ_PE"     "MERVAL.Index_INDX_ADJ_PE"
[7] "TPX.Index_INDX_ADJ_PE"
```

Results

- RMSE Train: 8.69
- RMSE Test: 8.78



Multivariate LSTM

Steps to Design a Fit Model and Network

Model Notation:

- Inclusion of 5 chosen assets
- Time series model
- Variable selection is not applicable

Split Into Train and Test Sets

- **5:1 ratio** yielded better results in training model

Split Into Input and Output

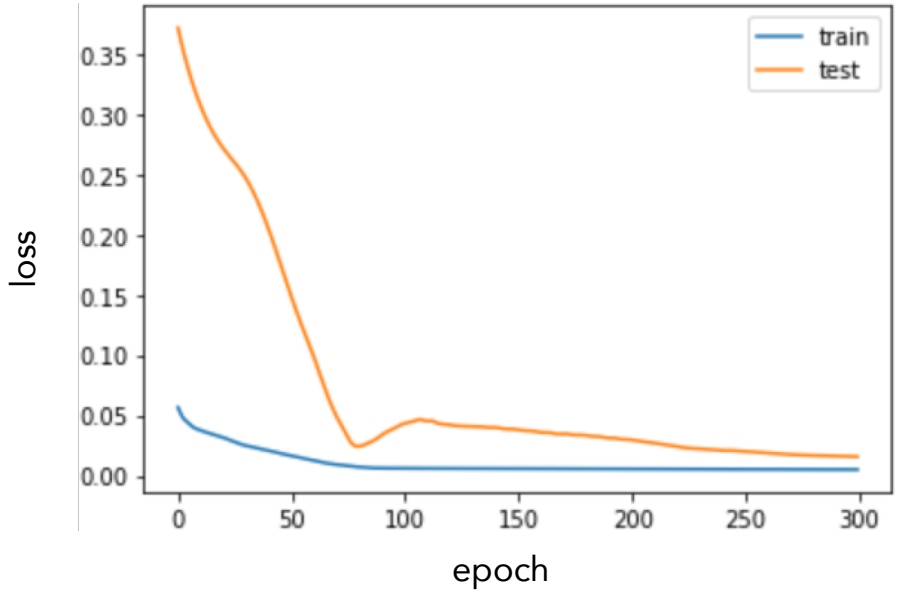
Reshape Input To Be 3D (Samples, Timesteps, Features)

Design Network

Fit Network

- **Epochs:** from 50 to 300 to create more hidden layers
- **Batch size:** from 1 to 1200

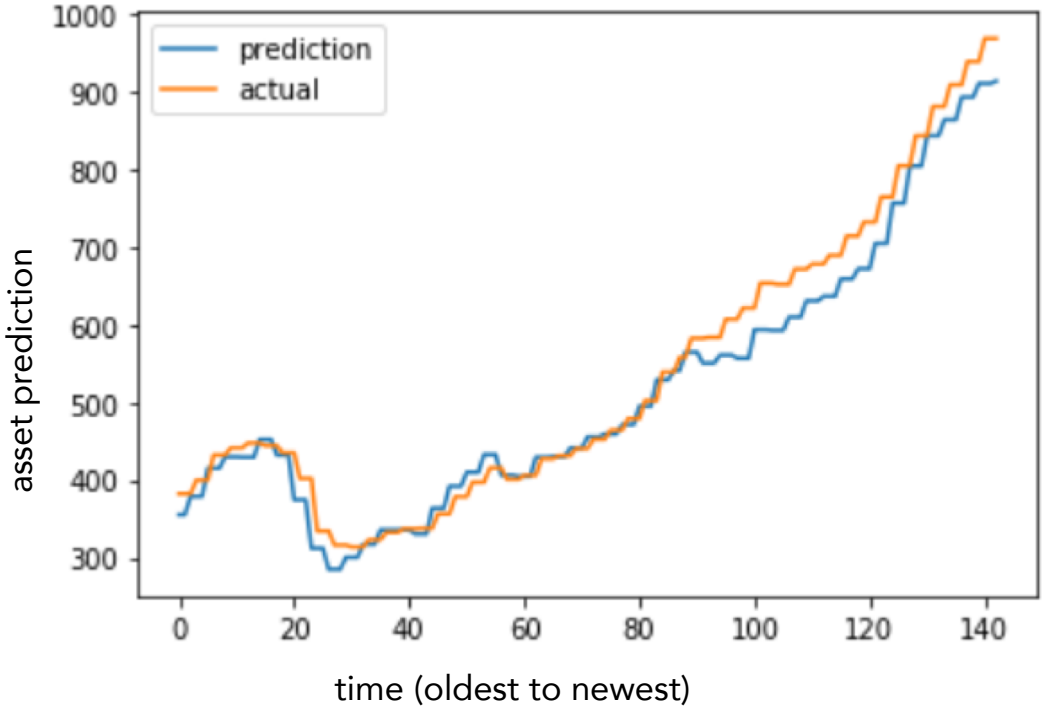
Plot History



Multivariate LSTM

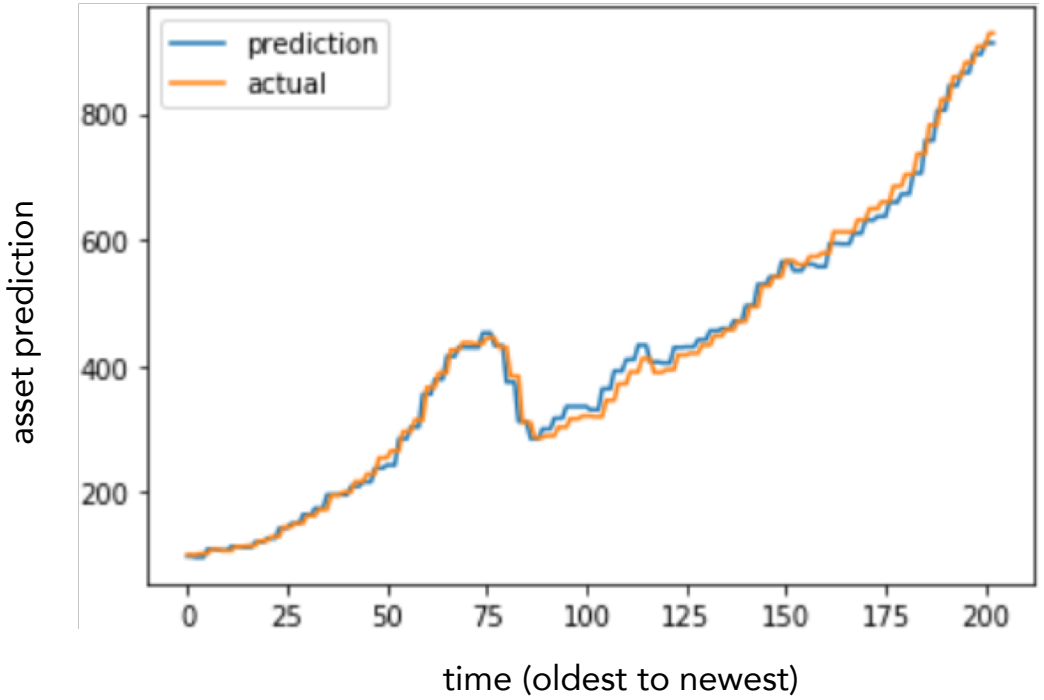
Results

Monthly Data



- Train RMSE: **61.99**
- Test RMSE: **32.18**

Quarterly Data



- Train RMSE: **5.91**
- Test RMSE: **16.79**

Quarterly dataset is recommended as the results **would be more accurate and aligned with AFP Habitat's reporting system. However the model seemed to be overfit due the big difference between train and test set**

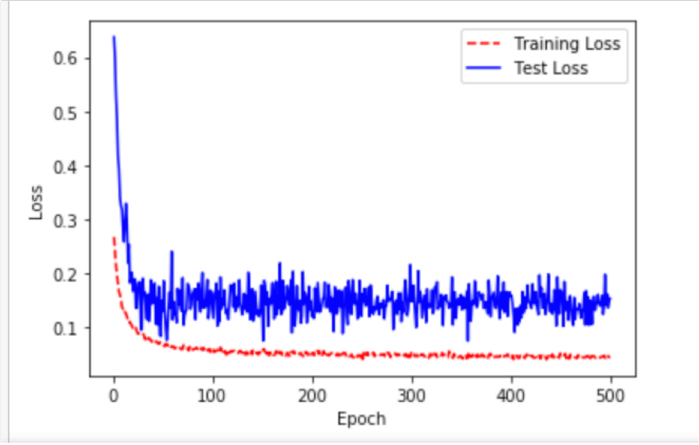
CNN-Convolutional Neural Network

Model Notation:

- Inclusion of 5 chosen assets
- Time series model
- Variable selection is not applicable

Results (Normal Training)

Test RMSE: 0.15



à 1D CNN is not ideal for this project due to its overfitting

Select Type of CNN

- 3D CNN - Action Recognition, Medical Imaging
- 2D CNN - Image Recognition, Computer Vision
- **1D CNN - Time Series, NLP**

Model Construction (Hidden Layers)

- Convolution Layers
- Pooling Layers
- Fully Connected Layers

Model Training Method

- Sliding windows
  - *test loss in sliding window methodology is upwarding*
  - *the reason might be that the dataset is too small*
- **Normal training –** train on the whole dataset directly

Prediction

- Prediction for 5 assets in Jan, Feb and March 2019
- Scale back to actual return

	Private Equity	Real Estate	Buyout	Venture	Private Capital-Euro
Jan	[[360.63818	365.20377	489.38266	113.44745	697.6881]
Feb	[362.99286	367.50415	492.92987	114.03789	701.13464]
Mar	[364.9497	368.6627	496.00967	114.597046	703.2142]]

Bayesian Model

Coefficients Distribution

As Bayesian model does not output the specific estimated coefficients, it gives out a distribution of estimated coefficients.

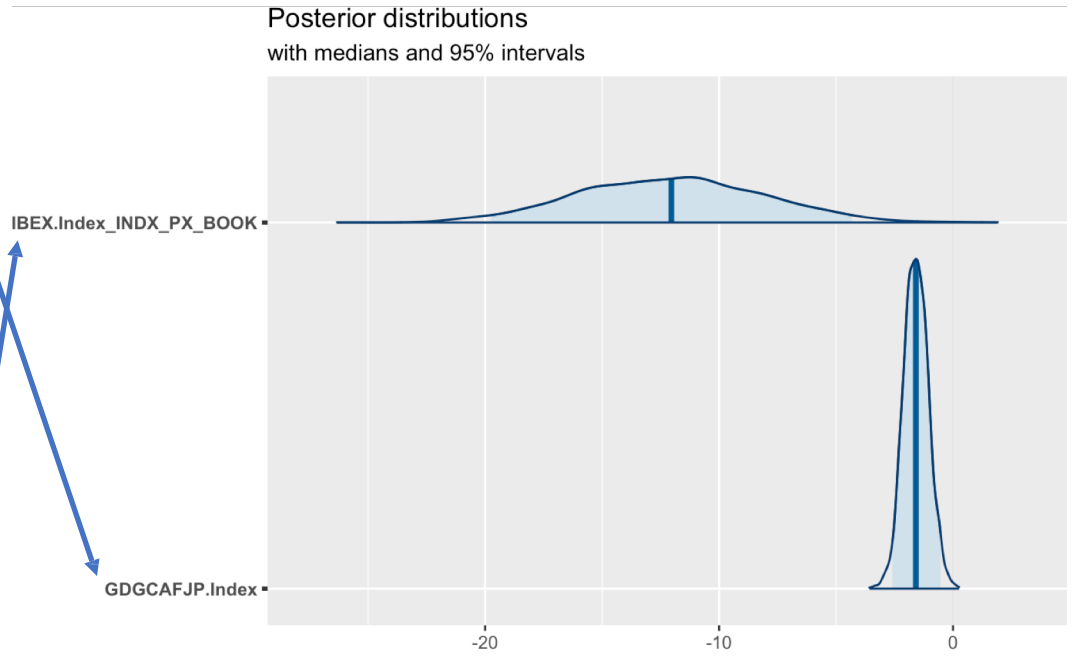
Below is the summary and visualization of parameters in the model to make it easier to compare the distribution and uncertainty.

Example:

Private Equity

```
 predictors:   83

Estimates:
                                    mean   sd     2.5%   97.5%
(Intercept)                         273.3  344.6 -382.4  963.0
SPLPEQTY.Index                        0.0    0.0    0.0    0.0
X0016639.Index                        1.3    2.0   -2.7    5.3
X1116639.Index                       -0.4    1.1   -2.5    1.6
X1706639.Index                        0.9    0.7   -0.5    2.3
X9246R004.Index                      -1.1    0.8   -2.6    0.4
CHBCIMCE.Index                        0.1    0.1   -0.1    0.4
EUICEMU.Index                         0.9    0.3    0.2    1.6
GDGCAFJP.Index                       -1.6    0.5   -2.6   -0.5
JFHIGLOB.Index                        0.0    0.0    0.0    0.0
NYFYPROB.Index                        0.2    0.1   -0.1    0.5
OECLKLAC.Index                       -1.1    1.3   -3.6    1.5
OECNKLAC.Index                        0.7    1.3   -1.6    3.2
OEOEKLAC.Index                        2.0    4.8   -7.9   11.3
OEOTKLAC.Index                       -4.9    4.3  -13.2    3.4
WGDPCHIL.Index                        0.6    0.2    0.2    0.9
WGDPCHIN.Index                        0.0    0.0    0.0    0.0
WGDPEURO.Index                        0.0    0.0    0.0    0.0
WGDPUS.Index                          0.0    0.0    0.0    0.0
WGDPWRLD.Index                        0.0    0.0    0.0    0.0
MEXBOL.Index_PX_LAST                  0.0    0.0    0.0    0.0
MXCLSC.Index_PX_LAST                  0.2    0.1    0.0    0.4
MXLA.Index_PX_LAST                    0.0    0.0    0.0    0.0
MXLASC.Index_PX_LAST                 -0.3    0.2   -0.6    0.0
NIFTY.Index_PX_LAST                   0.0    0.0    0.0    0.0
SX5T.Index_PX_LAST                    0.0    0.0    0.0    0.0
HSI.Index_INDX_PX_BOOK                2.7   11.1  -19.8   24.2
IBEX.Index_INDX_PX_BOOK             -12.1    4.0  -19.7   -4.3
IPSA.Index_INDX_PX_BOOK              -1.4    6.8  -15.2   11.8
```



Bayesian Model

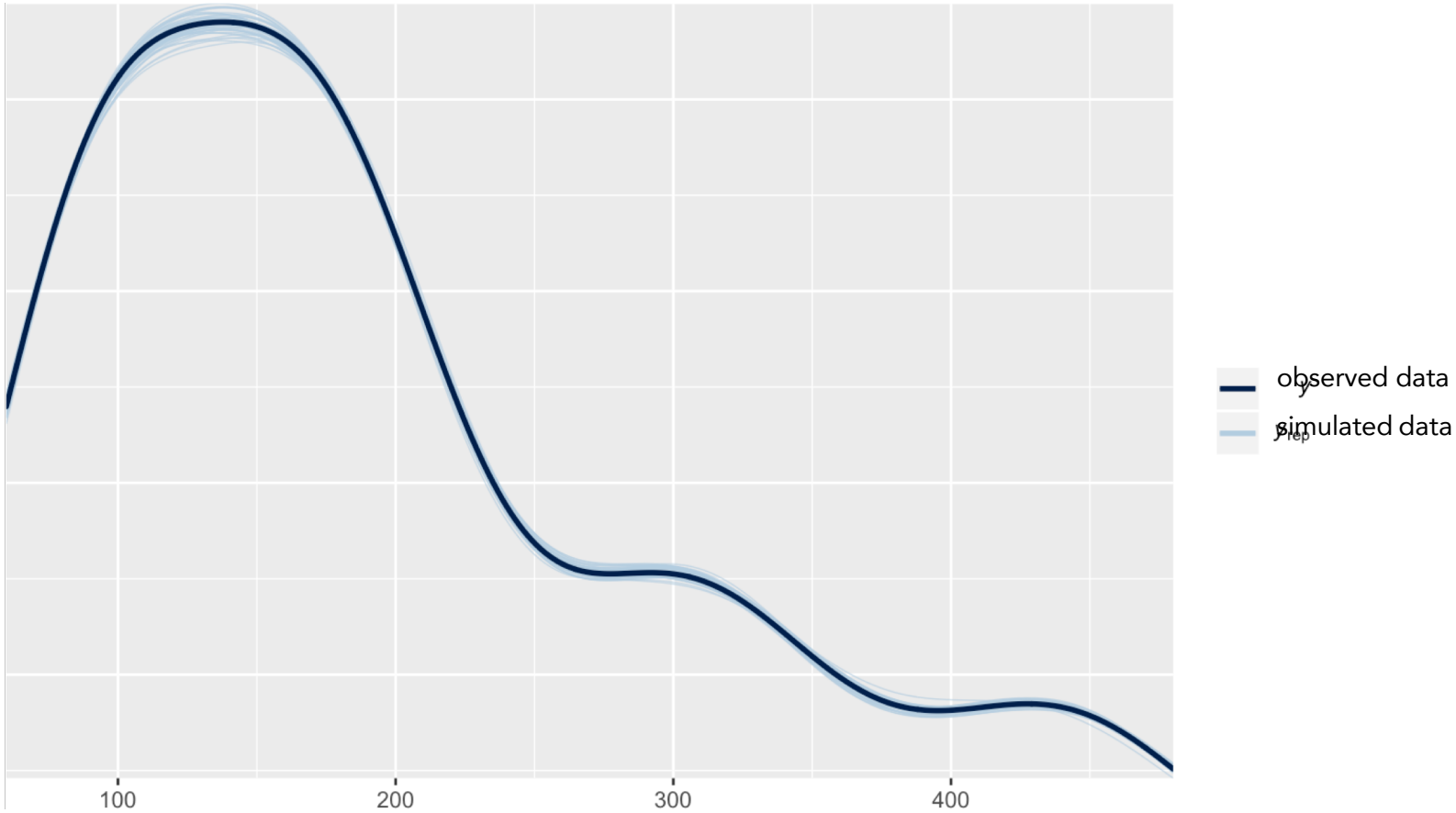
Model Performance

Visualize the posterior predictive distribution **to check the fit of the model.**

If a model is a good fit then we should be **able to use it to generate data that looks a lot like the data we observed.**

Example: Private Equity

Here is the simulation of data 50 times



Bayesian Model

Results

Model	Data Usage	Language	TEST RMSE				
			Private Equity	Real Estate	Buyout	Venture	Private Capital-Europe
Bayesian	Monthly	R	7.90	11.79	11.65	3.99	21.56
Number of Predictors			82	72	96	92	88

Benefit:

- **One of the major benefits of the Bayesian approach is the ability to incorporate prior information.**

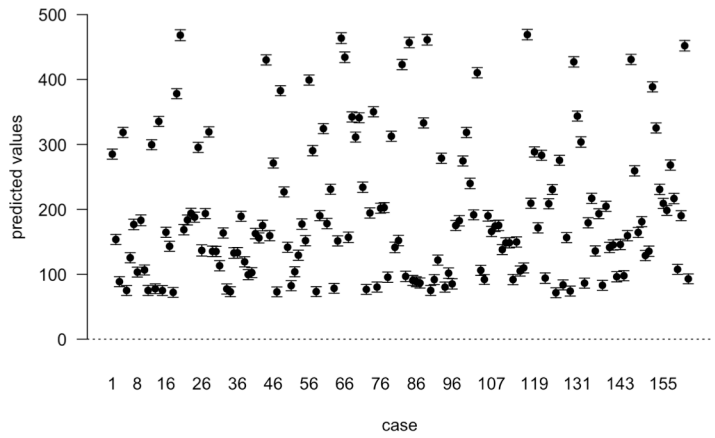
Drawback:

- **It is intensive computationally** to apply Bayes Theorem to complex models.

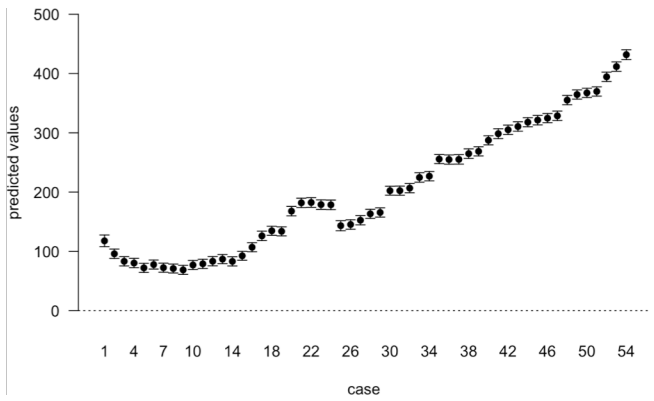
à **Consistent in testing results**

Example: Private Equity

These two plots show the 95% credible interval of predicted values with an input data of predictors from train and test.



Credible Interval For Train
Train RMSE: **7.91**



Credible Interval For Test
Test RMSE: **7.90**

Key Takeaways

Model	Model Behavior	Data Usage	Language	TEST RMSE				
				Private Equity	Real Estate	Buyout	Venture	Private Capital-Europe
CAPM	Regression	Daily	R	**22.40**	**29.01**	**33.54**	**8.78**	**41.62**
BAYESIAN	Regression	Monthly	R	**7.90**	**11.79**	**11.65**	**3.99**	**21.56**
LSTM	Time Series	Quarterly	Python	**16.79**				
CNN	Time Series	Monthly	Python	**0.1537**				

1

Say No To CAPM Model

Compare to all other model results, CAPM has its limitations in prediction. Hence, the framework produces unreliable outputs. We recommend to re-consider the usage of this model.

2

Say Yes To Bayesian Model

Bayesian model yielded the most consistent testing results compared to the other models. The ability to incorporate prior information proves its credibility and reliability.

MODEL EVALUATION

Consistency in Testing Results
RMSE Scores
Business Objectives

3

Consider Time-Series Deep Learning

LSTM and CNN model seemed to be overfit due to their structure and requirements. However, there is a huge potential for them to be powerful in predicting returns.

Future Implication

Framework Evaluation

Learn and test model outputs against actual results to identify which framework is the most meaningful for AFP Habitat's assets

Model Development

Continue the effort to tweak model features to fit with business objectives and goals

Analytics Advancement

Build a sophisticated network of analytics that can leverage AFP Habitat's strategic decisions with accurate returns prediction and beyond

THANK YOU