Machine Learning Applications for Finance

ML in Quantitative Finance demystified

Industry Lecturer

Jun Fan

junfan@bu.edu

About me

- B.S./M.S. in Computer Science from Shenzhen University/Columbia University
  - Statistical Pattern Recognition concentration
- A.B.D. from Ph.D. program
- M.B.A. from the Wharton School at UPenn
- Philips Research, FDO Partners, CargoMetrics Tech, Acadian Asset Management, State Street Global Advisor
- Started my career in Quantitative Finance in 2005
- 12 Years, Partner/Portfolio Manager in FDO Partners (a boutique hedge fund, HBS Professor Ken Froot and MIT Professor Rudi Dornbusch's firm)
  - Research, trading, data collection, infrastructure/coding, reconciliation, team managing, responsible for PnL
  - Managed a $150M AUM EMN hedge fund that traded $100M everyday, held 1000+ stocks, managed around $1B book
- Why I am here?
  - Interviewed and hired your alumni/ae
  - Know your strength and weaknesses
  - Here to share and help

Agenda

- Quantitative Investment intro
- ML in Quantitative Finance – industry perspective
- Introduce Foreign Exchange rate
- Discuss prediction model and tools
- Apply the theory to real world
  - OLS, GLS, AR, VAR
  - Support Vector Regression, Random Forest Regression (if time permits)
- Review and tips

What is quantitative investing

Quantitative investing

Approach that utilizes advanced math and stat models on large dataset leveraging computer technology supported by empirical evidences to calculate the optimal trades.

Fundamental investing

Approach that utilize in-depth analysis of a company's business, management team and market opportunities to determine a stock's intrinsic value and mispricing opportunities.

https://www.ssga.com/library-content/pdfs/insights/Quant-Investing-Comparing-and-Contrasting-Part-1-of-3.pdf

What is quantitative investing

Quantitative investing

- Systematic - usually unbiased

- Diversified alpha
- Breadth – thousands of positions
- Pre-defined rules
- Statistics can be misleading
- Limited to certain aspect of company valuation

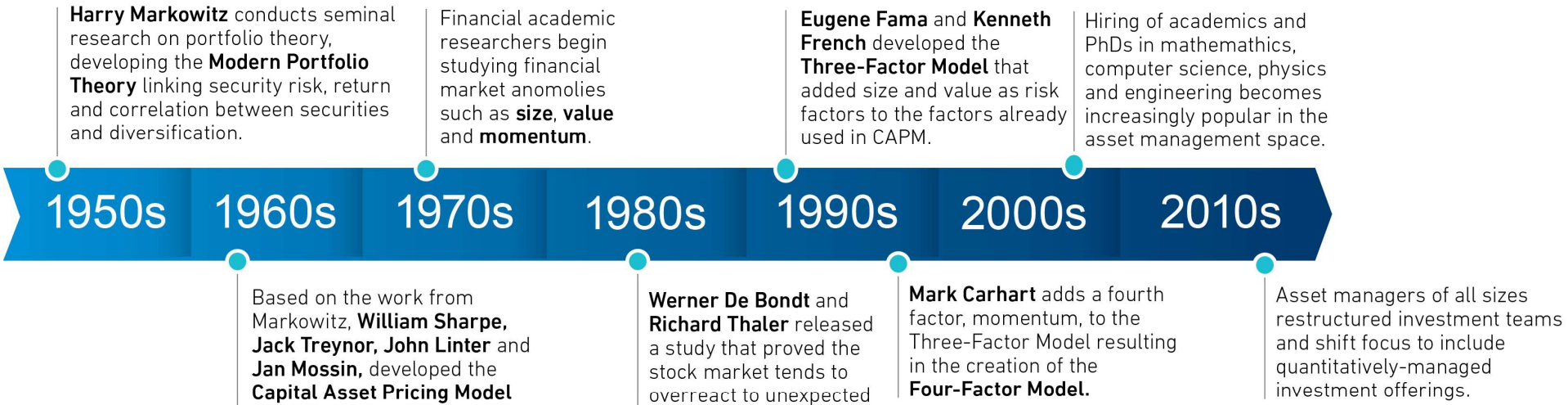
Fundamental investing

- Sometimes subjective

- Concentrated alpha
- Narrow – less than 100 positions
- Discretionary rules
- Not fully rely on data
- Deeper coverage to valuate company's intangible assets

https://www.ssga.com/library-content/pdfs/insights/Quant-Investing-Comparing-and-Contrasting-Part-1-of-3.pdf

What is quantitative investing



361capital.com

Top quantitative asset managements/hedge funds

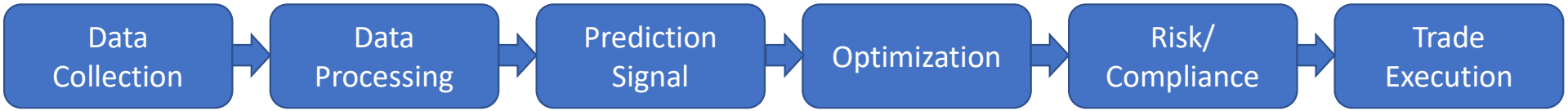
Asset Managers

- AQR
- Acadian
- Man Numeric
- PanAgora
- G.M.O.
- BlackRock (BGI)
- QMA

Hedge Funds

- Renaissance Technologies
- Two Sigma
- D. E. Shaw
- Point 72 (Cubiest)
- Millennium
- Jane Street
- WorldQuant

Typical Quantitative strategy process



- Aggregate multiple data sources

- Data Clean up
- Normalization
- Truncation

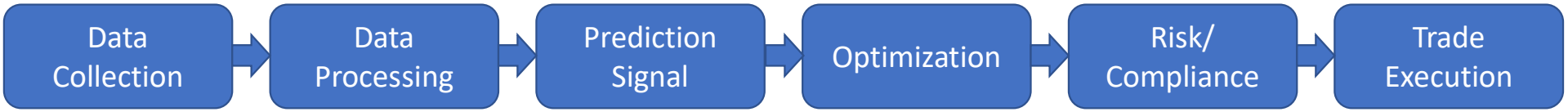
- Model Design
- Training Design
- Transformation
- FMB regression
- Goodness-of-fit

- Constraints
- Return and Risk
- Factors
- Covariance matrix

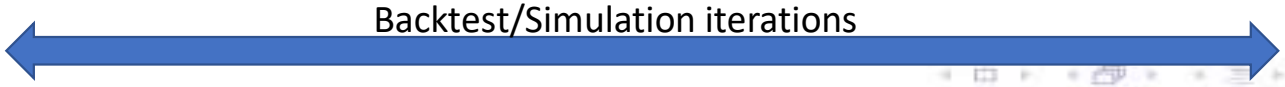
- Compliance
- Risk

- TCA
- Timing
- Liquidity

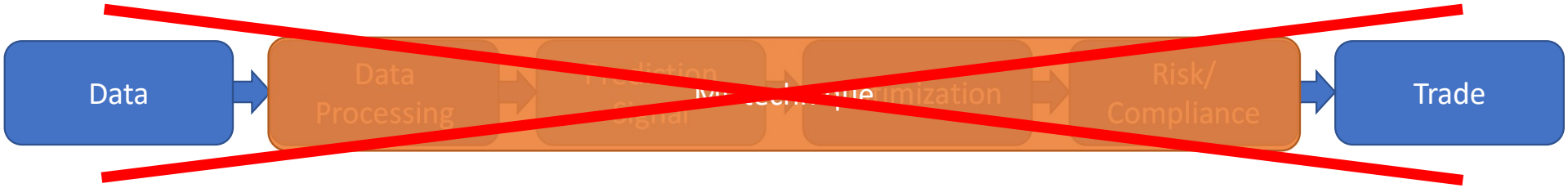
Typical Quantitative strategy process



		Goals			
Best Data available	Cleanest Data	Generate the best prediction	Find the best portfolio given our alpha and constraints	Meet risk and compliance requirement	Execute our orders at the lowest cost
		Measurement of Success			
Stable Efficient	Data quality	Goodness-of-fit	A set of statistics such as Transfer Coefficient, Active Risk, # of holdings, PnL etc	Risk and compliance target	Transaction Cost Analysis



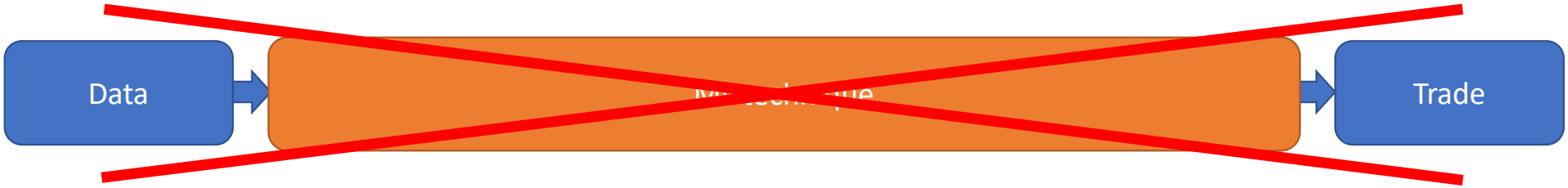
Where does ML fit in quantitative process?



- Not because we don't want ML, it's because finance is different
  - Too many moving variables
  - Too few relevant data
  - Heterogeneous
  - Too many goals

- We can't train the black box model and hope it will work without knowing why it works, or at the minimum, know when it will not work.

Where does ML fit in quantitative process?



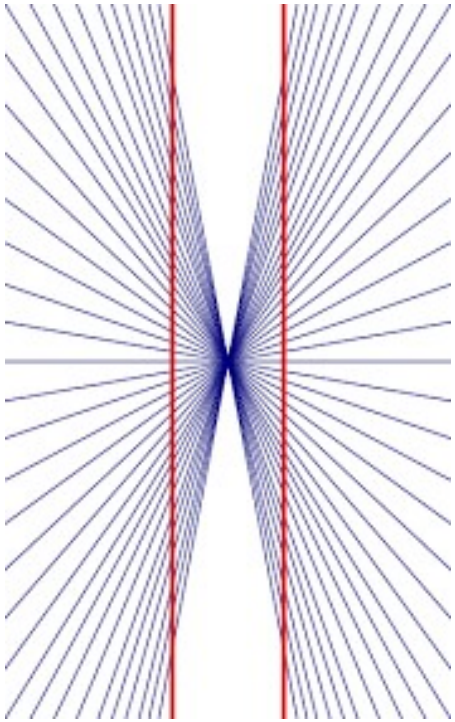
Typical Machine Learning – face recognition

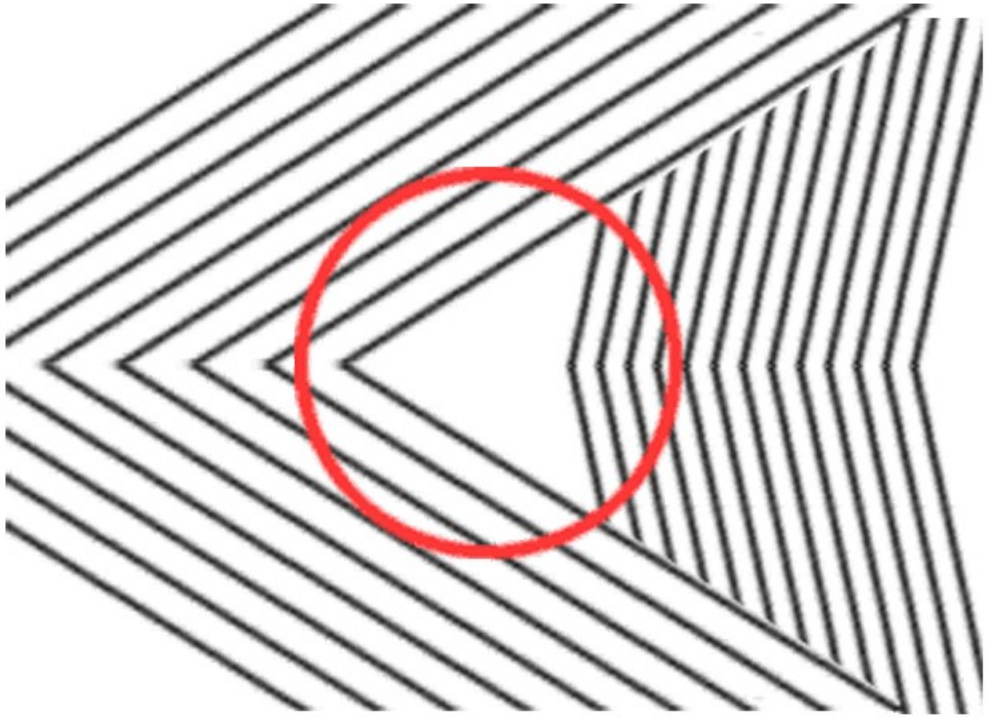
- Millions of faces as training data
- Most of faces have consistent features
  - Two eyes, one nose, one mouth, two ears
- Relatively homogenous
- Recognize faces

Typical Time-series - equity prediction

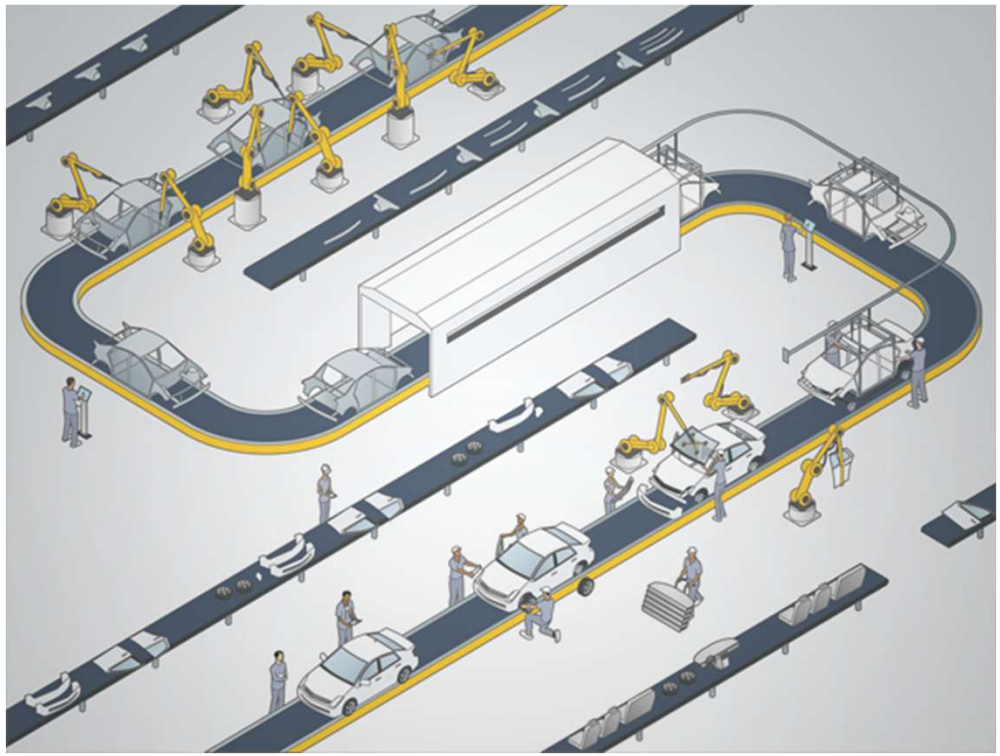
- 10 years data, 3 month forecast horizon. How many independent monthly data do we have?
- $(\overline{r}_P - \overline{r}_{Tbill}) = \lambda_0 + \lambda_1 \hat{b}_{p,1} + \ldots + \lambda_K \hat{b}_{p,K} + u_p$
- Generally Heterogenous - Large-cap, small-cap, Developed market, Emerging Markets, Sectors, Industries etc
- Find the best trade at lowest risk meeting the constraints that maximize the ex-post realized Sharpe ratio.

Even the best vision system sometimes fails

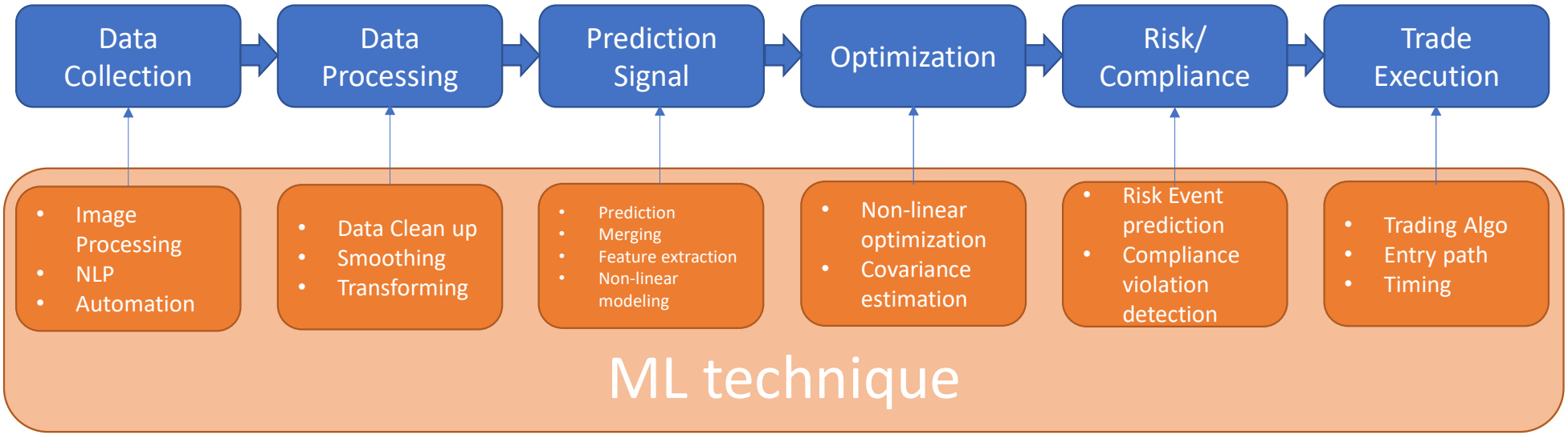




Automobile assembly line



Where does ML fit in quantitative process?



Quantitative Finance v.s. Data Science

Quant Finance

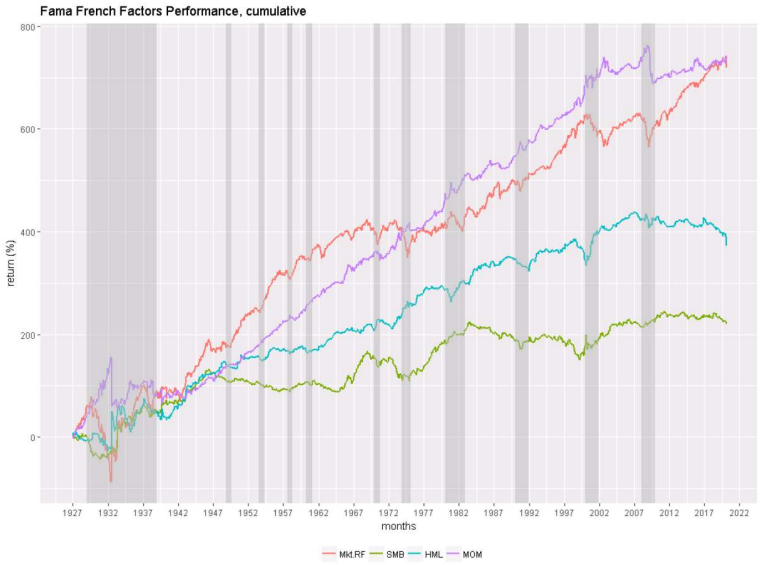
- Coding skills
- Stats/Math
- Social Science
- Finance
  - Time-series analysis
  - Equilibrium
  - Statistical arbitrage
- Examples
  - CAPM beta
  - Fama French 3 factors

Data Science

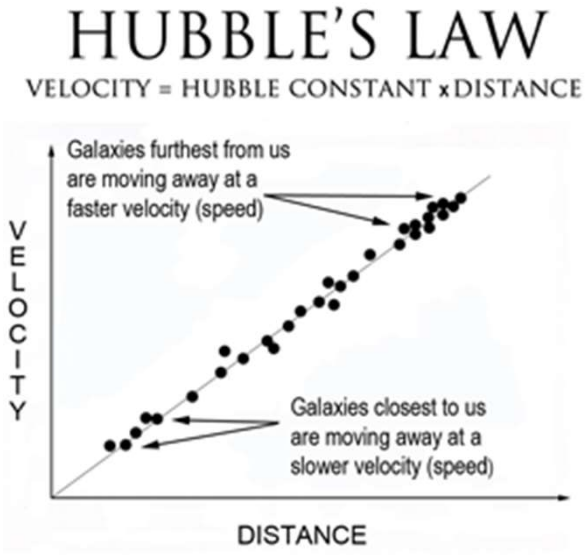
- Coding skills
- Stats/Math
- Data Science
- Biotech, Physics, Tech, Retail
  - Panel or time-invariant
  - Fundamental drivers
  - Empirical fitting
- Examples
  - Planck constant
  - Hubble constant

Quantitative Finance v.s. Data Science

Quant Finance

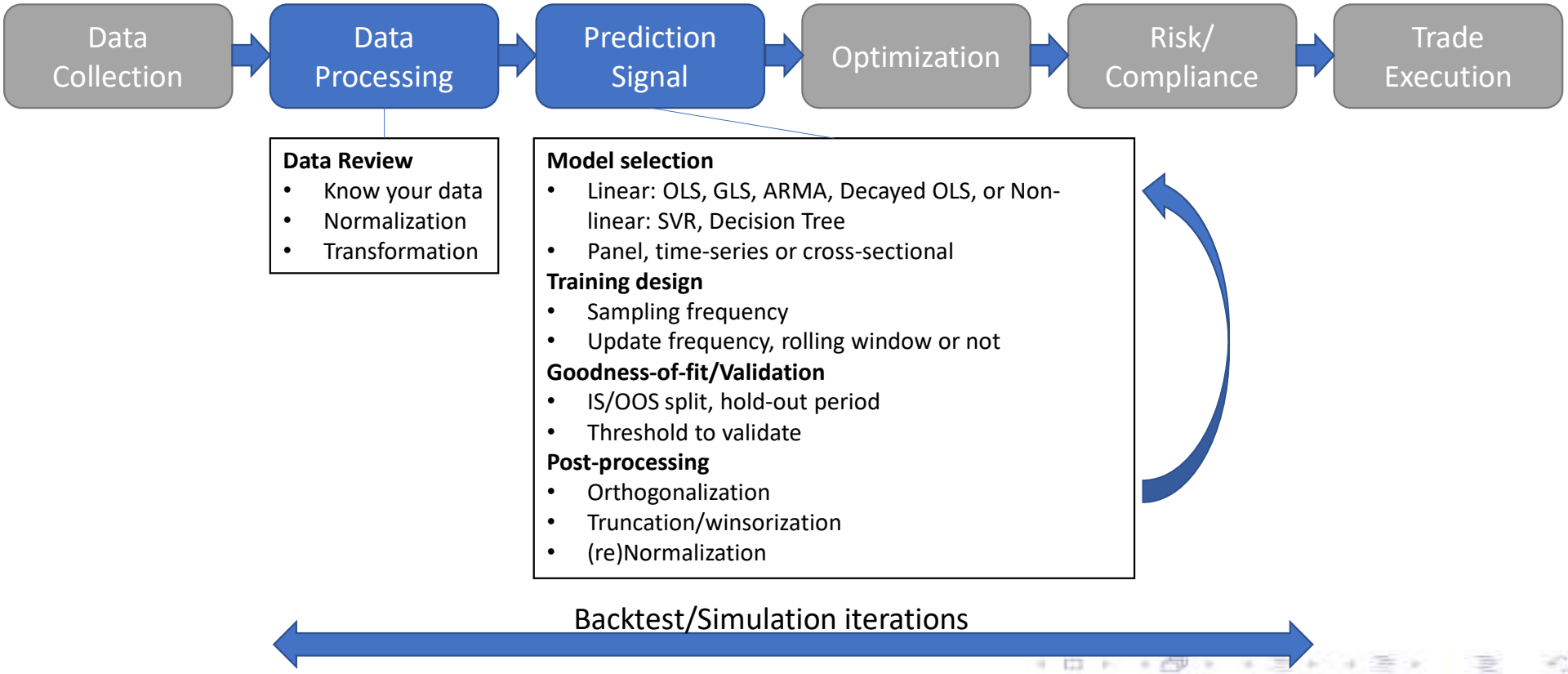


Data Science



astrobites.org

Typical alpha research practices

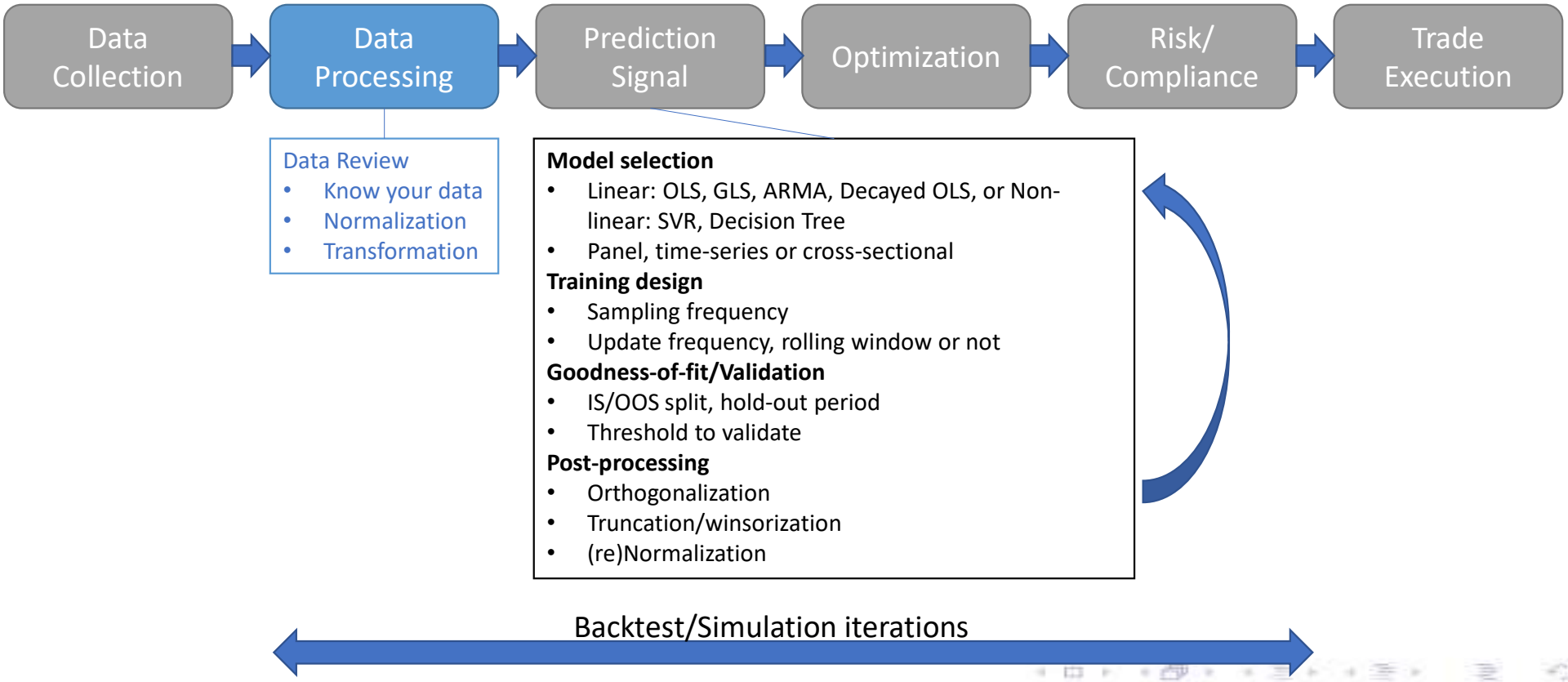


What to expect next

- Put all you've learned together in a real-world research
- Demonstrate an active FX trading strategy
  - Focuses on active alpha research in this 2 mini sections
  - Generate the **best alpha** for downstream process (optimization)
  - OLS, GLS, VAR, SVR
  - Practical rolling window application, MC simulation
- How to generate the best alpha quantitatively?
  - Grinold's rule $\alpha_s = \rho \cdot \sigma_s \cdot Score_s^{rank}$
  - Expected Return rule $\alpha_s = \beta \cdot Signal_s = \rho \frac{\sigma_y}{\sigma_X} \cdot Signal_s$

R.C. Grinold, Alpha is Volatility Times IC Times Score, Summer 1994

Typical alpha research practices



Data review

- Time-series plot
  - Time-varying features
- Histogram
  - Distribution and outliers
- Scatter plot
  - Relationship
- Covariance/Correlation between independent variables
  - Multicollinearity

Data - Foreign Currencies (FX)

- What's FX?
- What's unique about FX?
  - Priced in pairs
  - No centralized exchange
    - No holiday schedules
    - No official NBBO
  - Smaller cross-section (comparing to equity market)
  - Sensitive to geopolitics, macro, real inflation etc

DATE	USDJPY	GBPUSD	USDCAD	AUDUSD	NZDUSD	USDCHF	USDNOK	USDSEK	USDDKK	EURUSD
1/1/1971	358.0200	2.4058	1.0118	1.1181	1.1194	4.3053	7.1411	5.1639	7.4846	NaN
2/1/1971	357.5450	2.4178	1.0075	1.1238	1.1250	4.2981	7.1425	5.1726	7.4854	NaN
3/1/1971	357.5187	2.4187	1.0064	1.1243	1.1254	4.3003	7.1377	5.1628	7.4808	NaN
4/1/1971	357.5032	2.4179	1.0077	1.1238	1.1250	4.2987	7.1287	5.1630	7.4887	NaN
5/1/1971	357.4130	2.4187	1.0087	1.1243	1.1254	4.1242	7.1145	5.1660	7.4998	NaN
...	...	...	...	...	...	...	...	...	...	...
10/1/2020	105.2095	1.2980	1.3218	0.7121	0.6638	0.9124	9.2956	8.8346	6.3243	1.1768
11/1/2020	104.4061	1.3198	1.3073	0.7270	0.6856	0.9110	9.0999	8.6569	6.2968	1.1826
12/1/2020	103.7952	1.3434	1.2809	0.7532	0.7093	0.8884	8.7071	8.3631	6.1154	1.2168
1/1/2021	103.7883	1.3641	1.2725	0.7726	0.7200	0.8865	8.5096	8.2867	6.1082	1.2178
2/1/2021	105.3774	1.3867	1.2696	0.7753	0.7245	0.8979	8.5083	8.3437	6.1489	1.2094

Data - Foreign Currencies (FX)

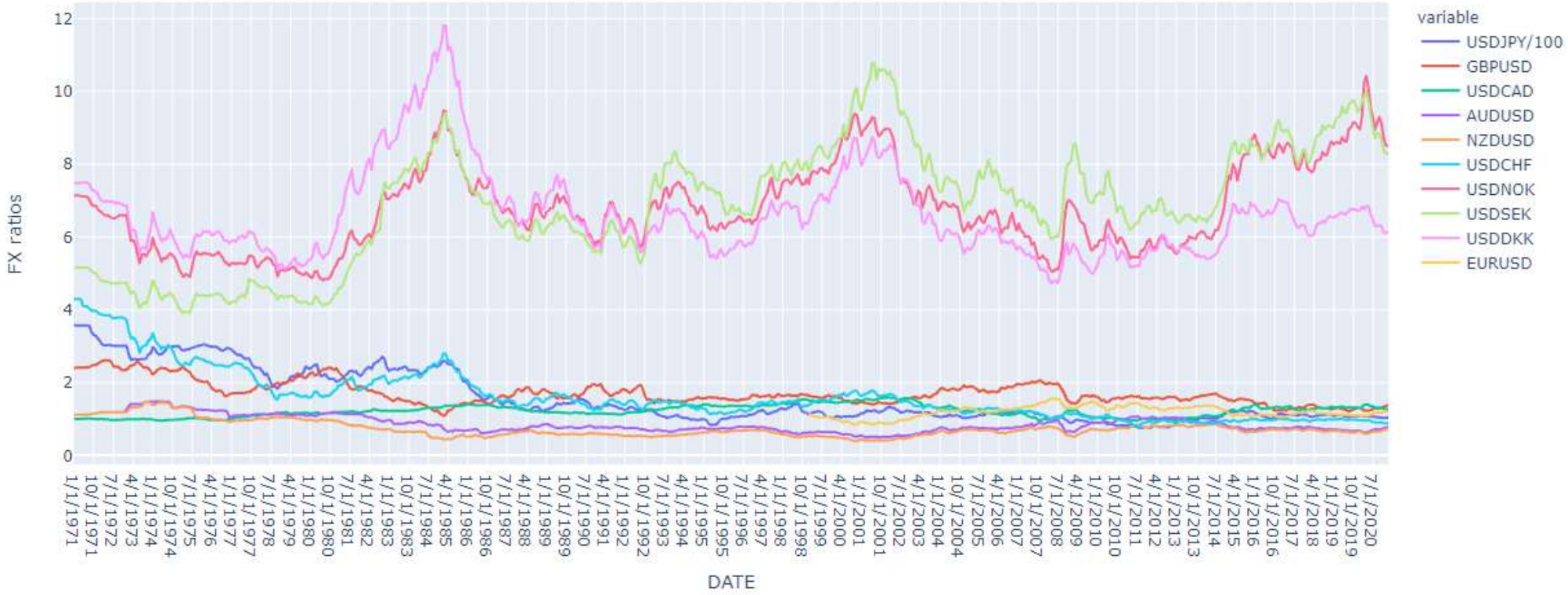
90) FX Markets Overview | 91) FX Forwards | 92) FX Options and Volatility | 93) Economics

Base Currency USD

30) FX Rate vs USD | DMMV » Deposit Rates | Bond Yields | Equity Indices

	Spot	% Chg	3M	Spread	10Y	Spread	Index	% Chg
USD	1.0000	0.00	0.25		1.72		4019.87	+1.18
EUR	1.1759	-0.15	-0.56	-81 bp	-0.33	-205 bp	3945.96	+0.68
JPY	110.69	+0.07	-0.10	-35 bp	0.13	-160 bp	29854.00	+1.58
GBP	1.3832	-0.01	0.10	-15 bp	0.79	-93 bp	6737.30	+0.35
CAD	1.2578	+0.24	0.22	-3 bp	1.51	-21 bp	18990.32	+1.55
AUD	.7610	-0.10	-0.03	-28 bp	1.84	+12 bp	6828.69	+0.56
NZD	.7032	+0.18	0.30	+5 bp	1.82	+9 bp	12488.31	-0.58
CHF	.9423	+0.02	-0.65	-90 bp	-0.31	-203 bp	11118.03	+0.64
NOK	8.5335	+0.16	0.25	0 bp	1.53	-19 bp	0.00	–
SEK	8.7267	+0.15	-0.10	-35 bp	0.38	-134 bp	2202.61	–

Data - Foreign Currencies (FX) monthly exchange rate

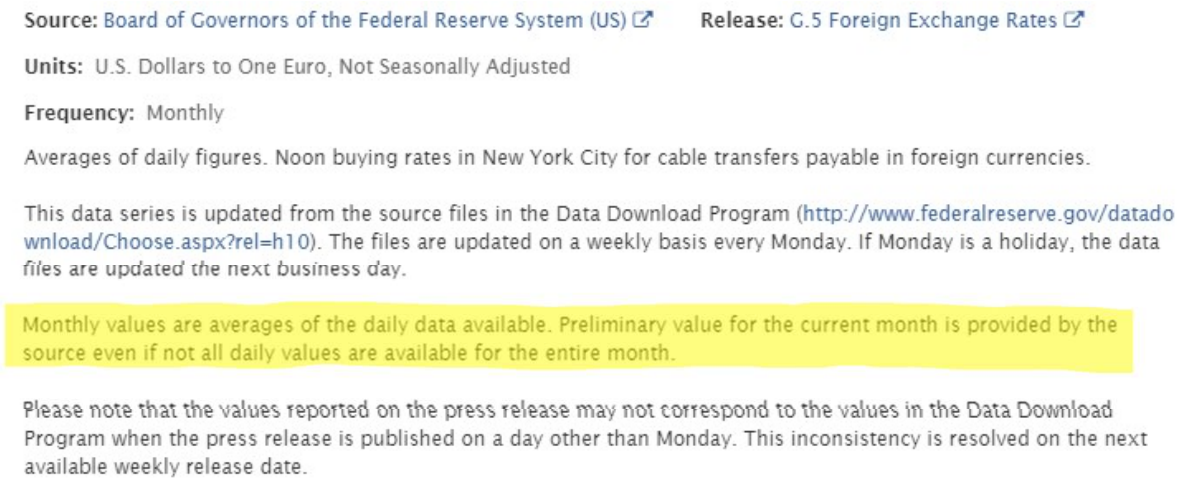


Data - Foreign Currencies (FX) monthly return

Data - Foreign Currencies (FX) daily return

Data - Foreign Currencies (FX)

- What's unique about FX?
  - Priced in pairs
  - No centralized exchange
    - No holiday schedules
    - No official NBBO
  - Smaller cross-section (comparing to equity market)
  - Sensitive to geopolitics, macro, real inflation etc



Source: Board of Governors of the Federal Reserve System (US) **Release:** G.5 Foreign Exchange Rates

Units: U.S. Dollars to One Euro, Not Seasonally Adjusted

Frequency: Monthly

Averages of daily figures. Noon buying rates in New York City for cable transfers payable in foreign currencies.

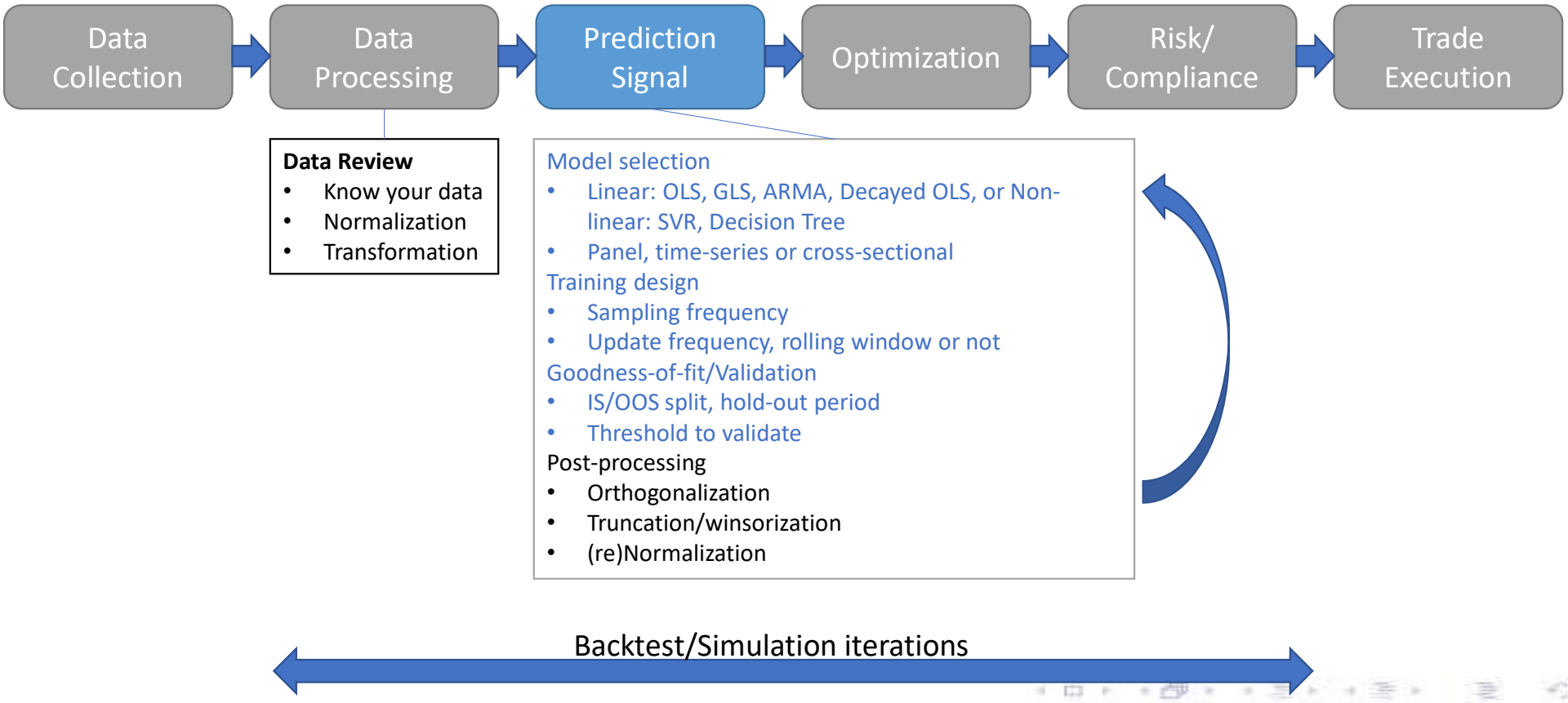
This data series is updated from the source files in the Data Download Program (http://www.federalreserve.gov/datadownload/Choose.aspx?rel=h10). The files are updated on a weekly basis every Monday. If Monday is a holiday, the data files are updated the next business day.

Monthly values are averages of the daily data available. Preliminary value for the current month is provided by the source even if not all daily values are available for the entire month.

Please note that the values reported on the press release may not correspond to the values in the Data Download Program when the press release is published on a day other than Monday. This inconsistency is resolved on the next available weekly release date.

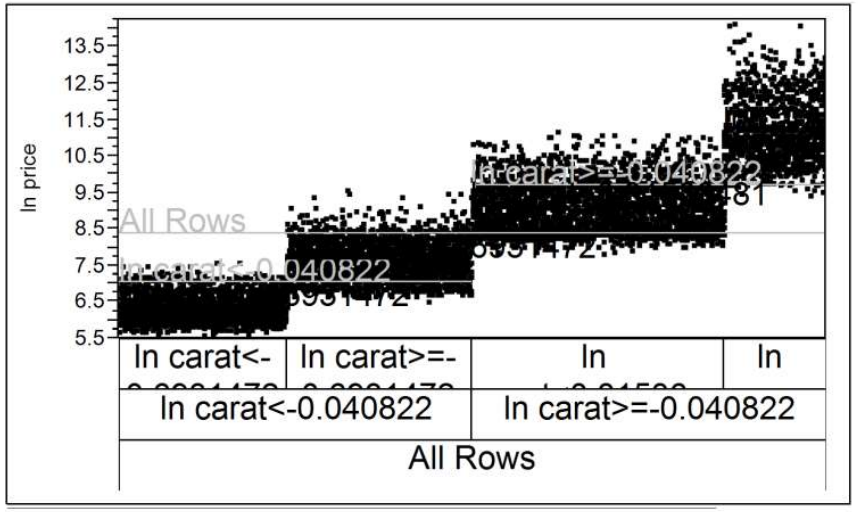
https://fred.stlouisfed.org/series/EXUSEU

Typical alpha research practices



Model Selection - OLS

- Why study OLS in ML class?
  - Is OLS a ML method? How about Ridge and Lasso?
  - Don't underestimate the effectiveness of OLS in finance prediction
- Usually practitioners use OLS as a starter to capture the basic linear pattern between predicter (independent, RHS) variables and predicted (dependent, LHS) variables.
  - transform the data if data doesn't exhibit linear relationship to dependent variable, i.e. LOG(mktcap)
- If we find meaningful linear pattern, we seek more fine-tuned models to enhance the model efficacy
  a) More fine-tuned linear relationship, or
  b) Some non-linear relationship on top of linear relationship
- Or, if we truly believe the relationship is non-linear, go straight to ML algorithms
  - But cross-validate the fitting and don't reply on model output as the only criteria for successfulness



Model Selection - OLS refresh

- Common Regression Assumptions
  - $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)\ and\ i.i.d.$
    - No autocorrelation
    - No heteroskedasticity
  - Model is linear between independent and dependent variables
    - Independent variables are not correlated (no multicollinearity)
  - $Cov(X_t, \varepsilon_t) = 0$

$$y = \alpha + \beta X + \varepsilon$$

$$\beta = X'X^{-1} \cdot X'y$$

$$se(\beta) = \frac{\sigma_\varepsilon}{\sqrt{n}} \cdot \frac{1}{\sigma_X}$$

$$\beta = \frac{\sigma_{X,y}}{\sigma_X^2} = \frac{\sigma_{X,y}}{\sigma_X \sigma_y} \frac{\sigma_y}{\sigma_X} = \rho \frac{\sigma_y}{\sigma_X}$$

$$R^2 \equiv \rho^2$$

$$Sharpe \equiv \rho \cdot \sqrt{T \cdot N}$$

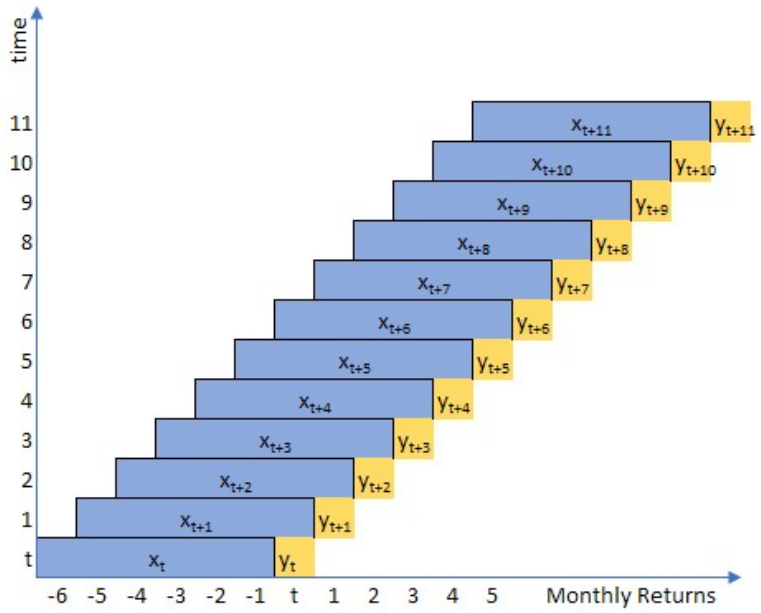
Training design principle

- Train all samples
  - Maximize data samples
  - Capture long term trends if exist, otherwise, converge to zero
  - Lower IS performance, but may exhibit strongest OOS performance
  - Limited OOS hide-out
- rolling windows/expanding window
  - Smaller data samples
  - Capture short term trends if exist, otherwise, just noise
  - Higher IS performance, but may have lower OOS performance
  - More OOS hide-out periods

		In-Sample Performance				Out-of-Sample Performance			
		Reg1	Reg2	Reg3	Reg4	Reg1	Reg2	Reg3	Reg4
5 year rolling window	Ret	4.12	3.43	7.31	6.80	4.76	4.20	4.51	4.18
	Vol	4.84	3.89	6.72	5.72	5.51	5.41	8.15	6.88
	IR	0.85	0.88	1.09	1.19	0.86	0.77	0.55	0.61
	maxDD	-19.90	-12.74	-12.69	-10.85	-22.35	-24.52	-37.73	-27.16
10 year rolling window	Ret	4.05	3.50	6.32	5.89	4.76	3.91	5.68	5.30
	Vol	4.90	4.47	6.46	5.71	5.54	5.64	8.15	7.29
	IR	0.83	0.78	0.98	1.03	0.86	0.69	0.70	0.73
	maxDD	-20.48	-14.75	-13.07	-11.15	-22.26	-29.86	-32.37	-25.03
30 year rolling window	Ret	3.97	3.41	5.85	5.33	4.76	4.01	5.97	5.65
	Vol	5.00	4.76	7.31	6.11	5.57	5.71	8.11	7.12
	IR	0.79	0.72	0.80	0.87	0.85	0.70	0.74	0.79
	maxDD	-21.58	-20.14	-32.02	-21.79	-25.39	-35.66	-36.55	-27.08

Training design - A typical time-series prediction model

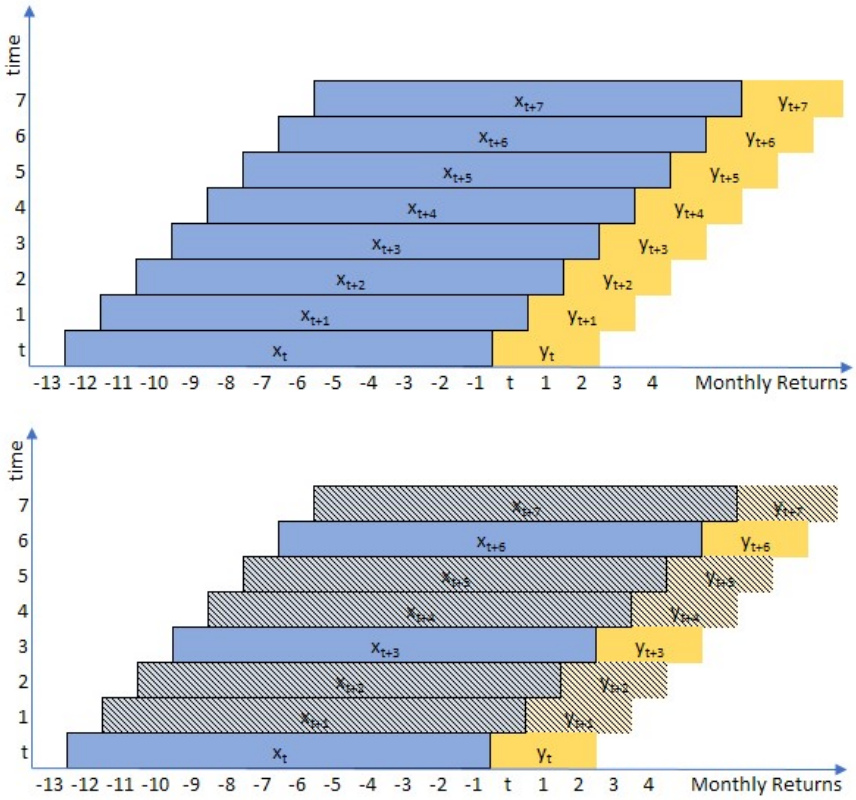
- The future return is a function of accumulative past returns
- At time t, look back m months data and use the sum of m months' data to predict next month (t+1) return
  - rolling moving average window approach
- Pooled or Time-series or Cross-sectional training?
- Challenges



Training design, cont.

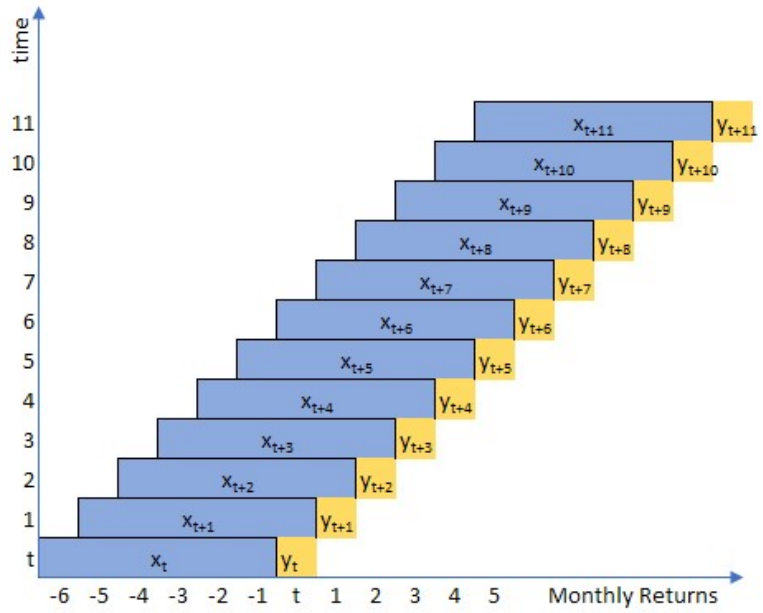
Challenges in OLS

- Autocorrelation -> t-stat will be biased
  - Newey-west adjustment
  - Adjust by sqrt(T) – back-of-the-envelope calculation
- Multicollinearity -> coefficients are hard to interpret but prediction can still be unbiased (as long as the Multicollinearity is stationary)
  - Take residual (piecewise)
  - Remove variables (parsimony)
- Biased to outliers, or the size of cross-sectional
  - GLS
  - Root-cap weighting



Model Selection – iterative process

- Iteration 1, no regression or assume beta of 1
- Iteration 2, OLS regression
- Iteration 3, Multi-variate regression if needed
- Iteration 4, GLS regression if needed
- Iteration 5, SVR regression if needed
- Iteration 6, Decision tree or Random Forest



To be continued