



NUMERAI MEETUP JAPAN 2021

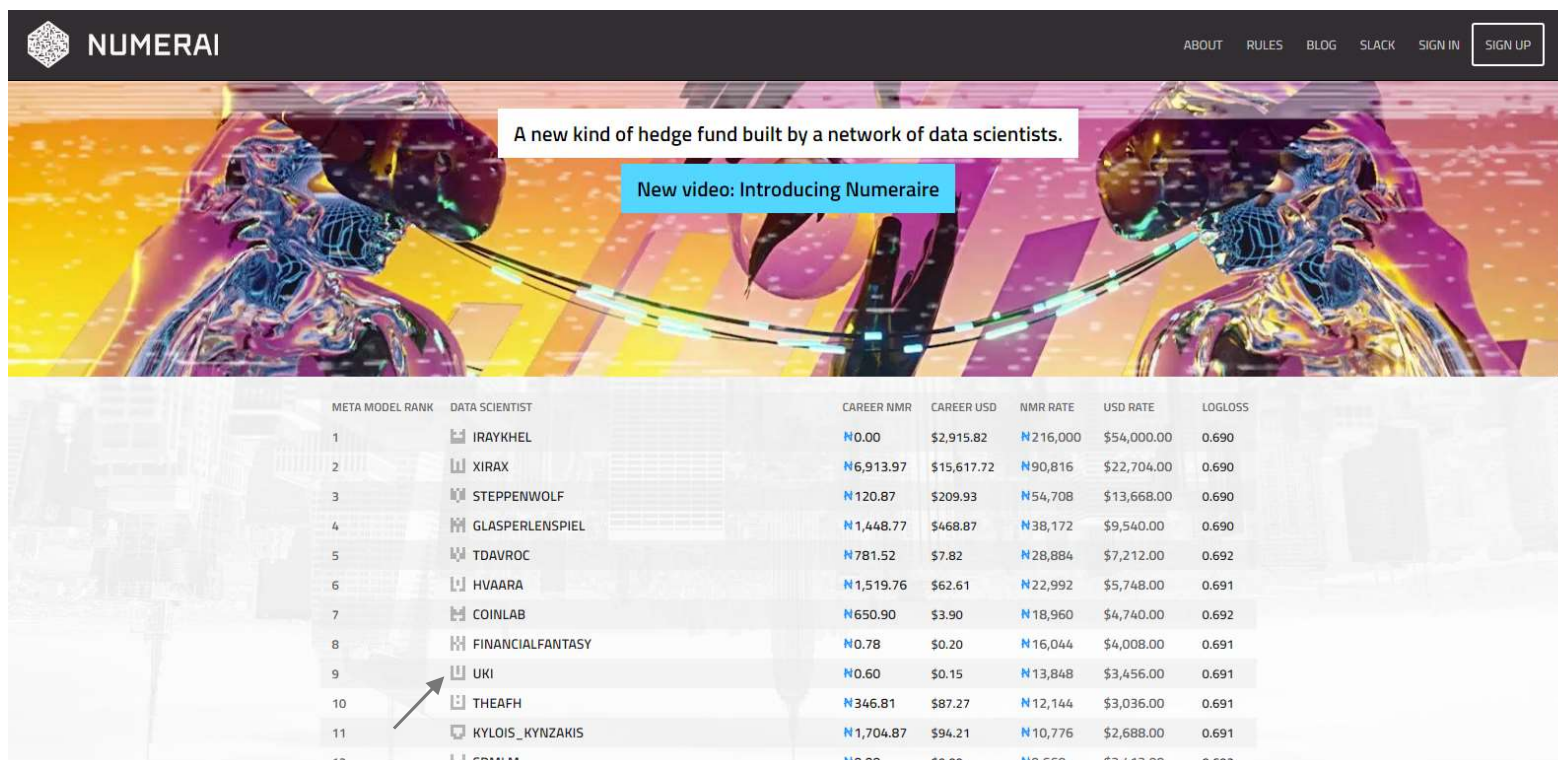
Machine Learning for Advancing Traditional Quantitative Approach



Dec. 18, 2021

UKI@blog_uki

自己紹介：Numeraiの経歴



META MODEL RANK	DATA SCIENTIST	CAREER NMR	CAREER USD	NMR RATE	USD RATE	LOGLOSS
1	IRAYKHEL	N0.00	\$2,915.82	N216,000	\$54,000.00	0.690
2	XIRAX	N6,913.97	\$15,617.72	N90,816	\$22,704.00	0.690
3	STEPPENWOLF	N120.87	\$209.93	N54,708	\$13,668.00	0.690
4	GLASPERLENSPIEL	N1,448.77	\$468.87	N38,172	\$9,540.00	0.690
5	TDVROC	N781.52	\$7.82	N28,884	\$7,212.00	0.692
6	HVAARA	N1,519.76	\$62.61	N22,992	\$5,748.00	0.691
7	COINLAB	N650.90	\$3.90	N18,960	\$4,740.00	0.692
8	FINANCIALFANTASY	N0.78	\$0.20	N16,044	\$4,008.00	0.691
9	UKI	N0.60	\$0.15	N13,848	\$3,456.00	0.691
10	THEAFH	N346.81	\$87.27	N12,144	\$3,036.00	0.691
11	KYLOIS_KYNZAKIS	N1,704.87	\$94.21	N10,776	\$2,688.00	0.691
12	SDMLM	N0.00	\$0.00	N9,660	\$2,412.00	0.692

Numerai Website in 2017

- 2017年3月 Numeraiへ初参加（ROUND67）
- 2017年6月 トーナメント報酬として初回エアドロップを受領（26.42NMR）
- ...
- 2020年4月 再参加（ROUND208～）
- トーナメントの最高順位20位（2021年11月）
- Signalsはベータ版に参加、2021年11月より本格的に参入

本日のプレゼン内容

対象：

主に中級者向け

内容：

クオンツに機械学習を適用すると、どのような収益機会が得られるのか
従来の理論体系に則った上で独自理論へ展開する。

- 株式市場がどのような動きをするか
- その中で伝統的なクオンツはどのようなことをしているのか
- 機械学習がどのように市場を捉えるのか
- 機械学習によって従来の理論体系の収益機会はどう拡張できるのか
- 伝統的なクオンツを発展させるための機械学習モデリング
- Signalsを推奨する理由

それでは始めます

株式投資やってますか??

時間軸	戦略の例	キーワード
<div>短期売買</div> <div>↑</div> <div>↓</div> <div>長期保有</div>	デイトレ	出来高ランキング 板読み 連れ高/連れ安
	イベント	好決算、上方修正 悪材料、不祥事 M&A、TOB
	テーマ株	コロナ（マスク・リモート） 半導体 メタバース
	配当投資	高配当 株主優待
	積み立て	米国株、インデックス ETF、投資信託 ロボアド

<株式投資の悪いイメージ>

両津 @ryoooootu

もっとも最高値の上昇率が高かった銘柄
2020年「テラ」
2021年「グローバルウェイ」

ハメコミ銘柄が毎年No.1上昇銘柄となる日本株市場。
これは衰退国の株式市場の特徴なのでしょう。
まともな投資家がアメリカ株に移る気持ちも分かります。

午後8:53 · 2021年12月9日 · Twitter Web App



株式市場の一日がどのように動いているか??

<一般的な視点>

- ① ランキング（上昇率・下落率・出来高）
＝仕手株やイベント、テーマ株

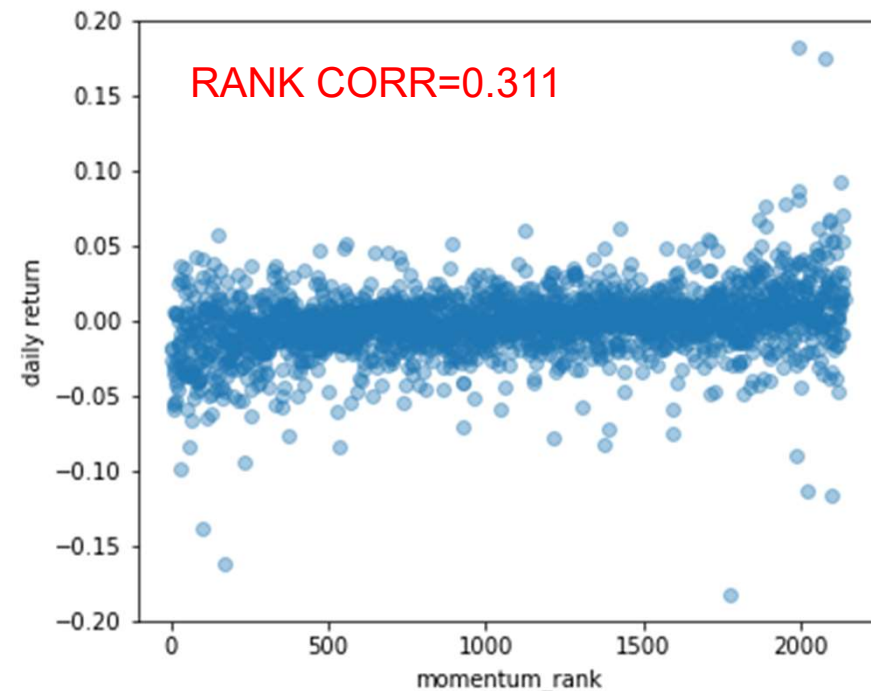
順位		銘柄	市場	現在値	日	前日比	前日比(%)
1	2195	アマタHD	JQ グ	13,150	C	+3,000	+29.56%
2	7063	B i r d m a n	東証マザ	1,463	C	+300	+25.80%
3	3189	ANAP	JQ ス	402	C	+80	+24.84%
4	3691	リアルワールド	東証マザ	968	C	+150	+18.34%
5	4194	ビジョナル	東証マザ	10,180	C	+1,500	+17.28%
6	6730	アクセル	東証1部	955	C	+136	+16.61%
7	7826	フルヤ金属	JQ ス	10,700	C	+1,140	+11.92%
8	6161	エスティック	東証2部	6,480	C	+640	+10.96%
9	4772	ストリームメディ	JQ グ	196	C	+16	+8.89%
10	6047	Gunosy	東証1部	840	C	+67	+8.67%

- ② 業種別ランキング
＝景気やマクロ情勢を反映

順位	業種	現在値	前日比	前日比(%)
1	金属製品	1,339.67	+1.58	+0.12%
2	電気・ガス業	328.97	-0.02	-0.01%
3	保険業	1,105.34	-1.34	-0.12%
4	機械	2,308.39	-3.35	-0.14%
5	卸売業	1,887.75	-5.44	-0.29%
6	鉱業	285.53	-0.88	-0.31%
7	その他製品	3,531.14	-11.28	-0.32%
8	電気機器	4,015.65	-15.08	-0.37%
9	ガラス・土石	1,195.95	-4.55	-0.38%
10	鉄鋼	429.52	-1.68	-0.39%

<クオンツ視点>

ファクターの動きを観察する。

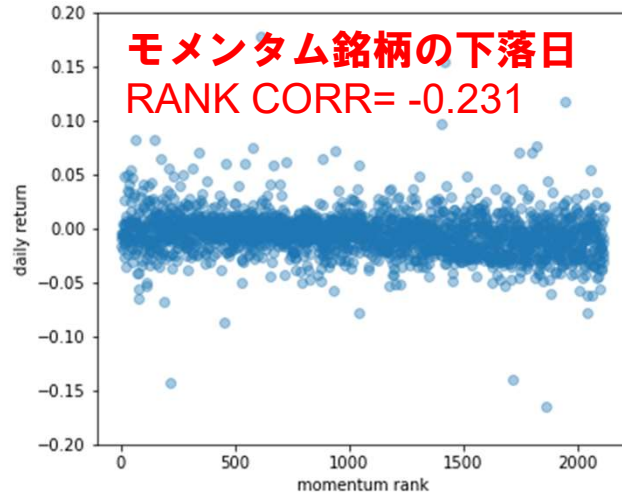
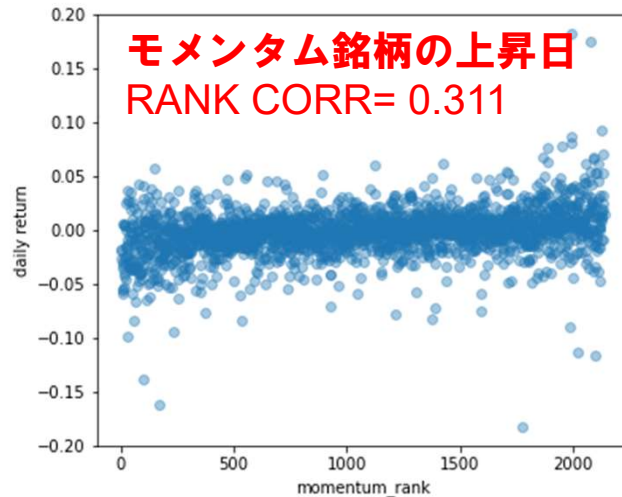


Price Move in a Stock Market on a Certain Day

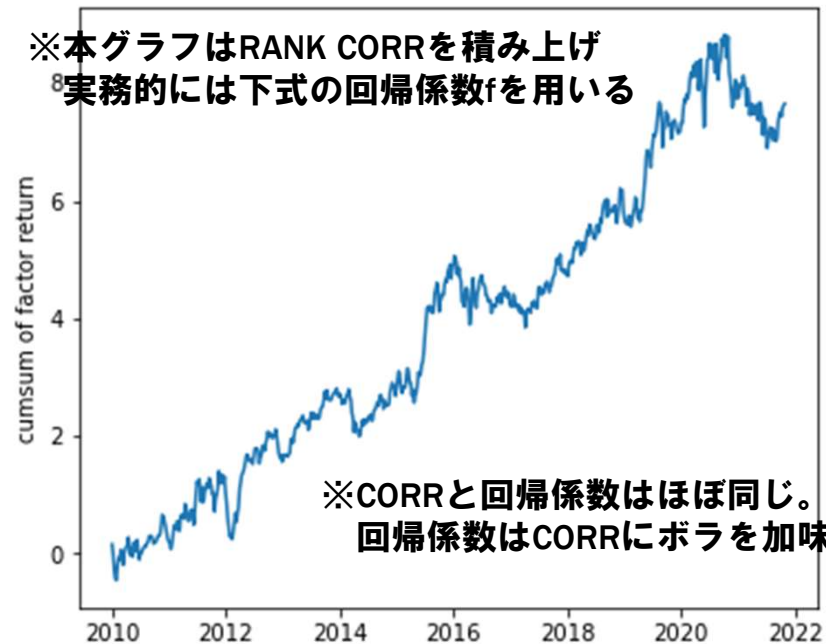
一般的な視点からは気付かないが、
市場には大きな力が働いている。
(ファクターリターンと呼ぶ)

クオンツモデル

ファクターに対する市場の動きは一定ではない → 時系列で積み上げて推移を表現する。



時系列で
積み上げ



Cumsum of Factor Return of Momentum Factor

<回帰式>

$$r_{i,t} = f_t \cdot x_{i,t} + u_{i,t}$$

- (注釈)
- ・ r はリターン、 u はスペシフィックリターン
 - ・ f はファクターリターン、 x はファクターエクスポージャー
 - ・ i は各銘柄を示す ($i=0 \sim 2000$ 程度)
 - ・ $t=0 \sim T$ まで各々を計算し、 f_t を $t=0 \sim T$ まで積み上げる
(=タイムシリーズではなく加セクション回帰)

Momentum-Trend Day and Momentum-Reversal Day

これを重回帰に拡張するとクオンツになる。

伝統的なクオンツと機械学習の対応

手法	伝統的なクオンツ	機械学習
目的	既知のファクターを使い、 市場の構造を記述する	様々な特徴量を使い、 将来のリターンを予測する
ラグ	ラグのないモデル (前頁の式に t と $t-1$ は混在しない)	ラグのあるモデル
可読性	リスクの可視化が重要 (アクティブリスクを制限し、 説明責任を全うする必要性)	可読性は不要 (リターンのみ追求する一方で FNCのような評価指標も存在する)
系の状態	静的なシステム (市場構造が将来的に継続すると いう絶対的仮定に基づく)	動的な予測が望ましい (参考: DLM)
アウトカム	ポートフォリオウェイト	予測 (Prediction)

是非はともかく、今後パラダイムシフトが起こる可能性を秘める。

リターンの分解と 各所への機械学習の効用

リターンの分解と機械学習の適用可能性

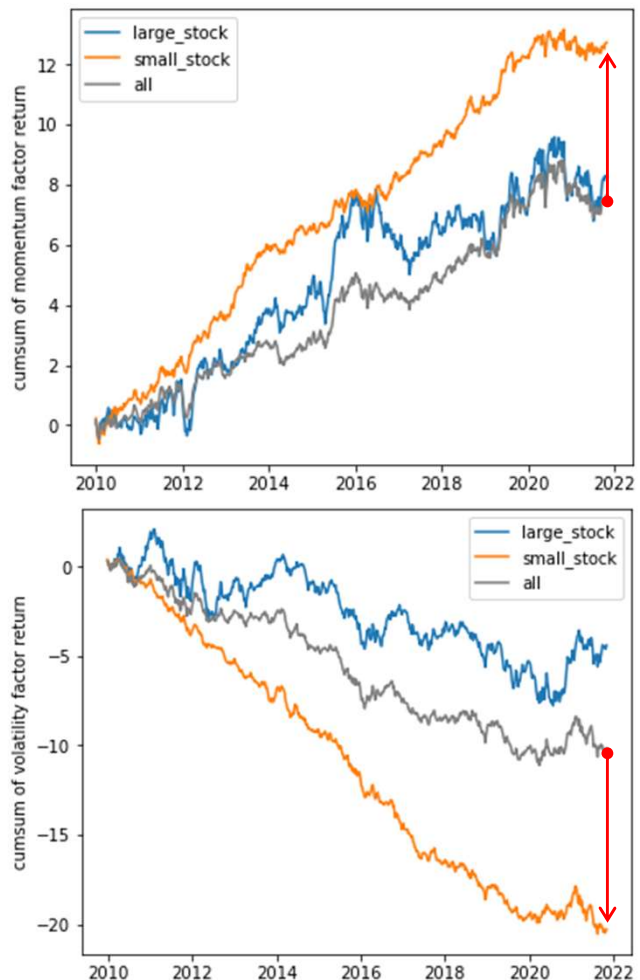
従来のクオンツでは
収益源としない部分

パフォーマンス分析のためのリターンの分類			説 明	MLの適用可能性
アクティブ リターン	アクティブ システムティック リターン ベータに起因する部分	期待 アクティブベータ リターン	いわゆるアクティブベータ。期待ABRはベンチマークの期待リターン（およそ6%）に起因する部分、サプライズはそこからの乖離に依存した部分。	
		アクティブベータ サプライズリターン		
	アルファに起因する部分	アクティブ ベンチマーク タイミングリターン	市場動向に応じてポートフォリオ・エクスポージャーを増減したことによるリターン。一般的に収益化は難しいとされている。	最終段のモデルの入力に対し、ファクタータイミングを予測する特徴量を導入する。
		レジデュアル リターン		
ベンチマーク リターン	ファクター リターン	ファクター タイミングリターン	業種配分による超過リターン	従来の業種に囚われない、数値に依拠したクラスタリング（次元圧縮＋分類器）。
		業種 ファクターリターン	リスクファクターへのティルトによる超過リターン	交互作用によるファクターRTの性能向上、MLによる新しいファクターRTの創出。
		リスクインデックス ファクターリターン	上記のいずれにも該当しない、固有（残差）成分	MLによるクラスタやMLによるファクターの控除による説明力向上。
		スペシフィック リターン		

以降、それぞれについて説明する

まずはお手柔らかに①交互作用

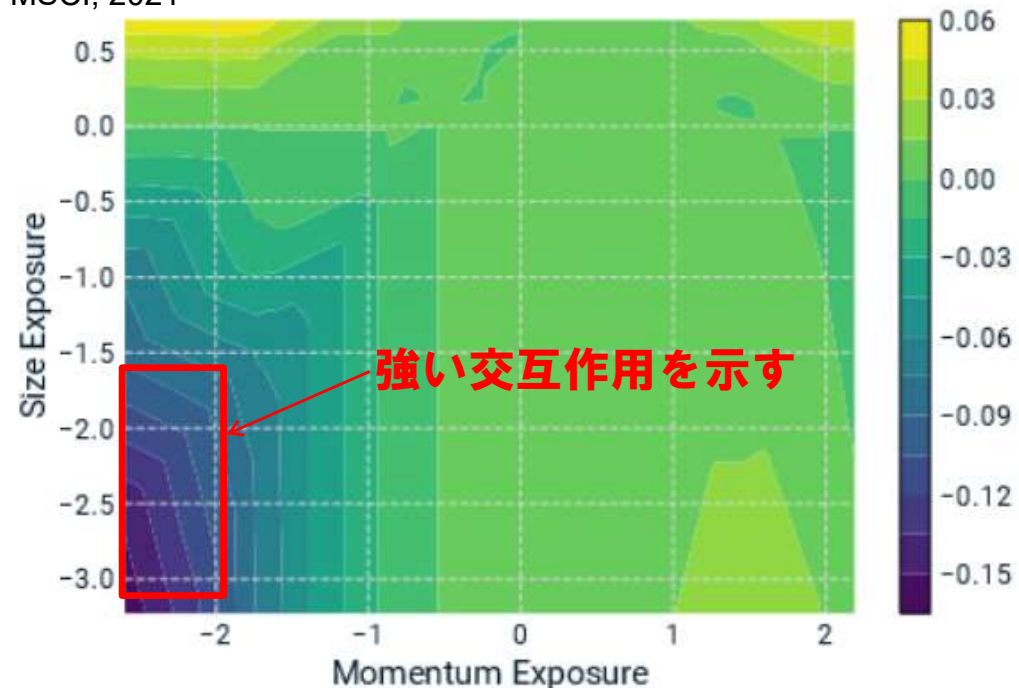
交互作用とは：2つの因子が組み合わさることによって現れる相乗効果



Factor Returns enhanced by the Interaction

ユニバースを限定することで性能が大きく向上する。
(時価総額、リージョン、業種、グロース...)
(クオンツではNon-Linearityと表現することが多い)

“Machine Learning Factors Capturing Non Linearities in Linear Factor Models”,
MSCI, 2021



Momentum-Size Interaction Strength

既存のML手法の殆どは明示的に与えなくとも交互作用を自動的に考慮してくれる

収益化が難しい？②ファクタータイミング

ファクタータイミング：アウトパフォームするファクターを見極める動的戦略

もともとはレジームの意味合いが強い

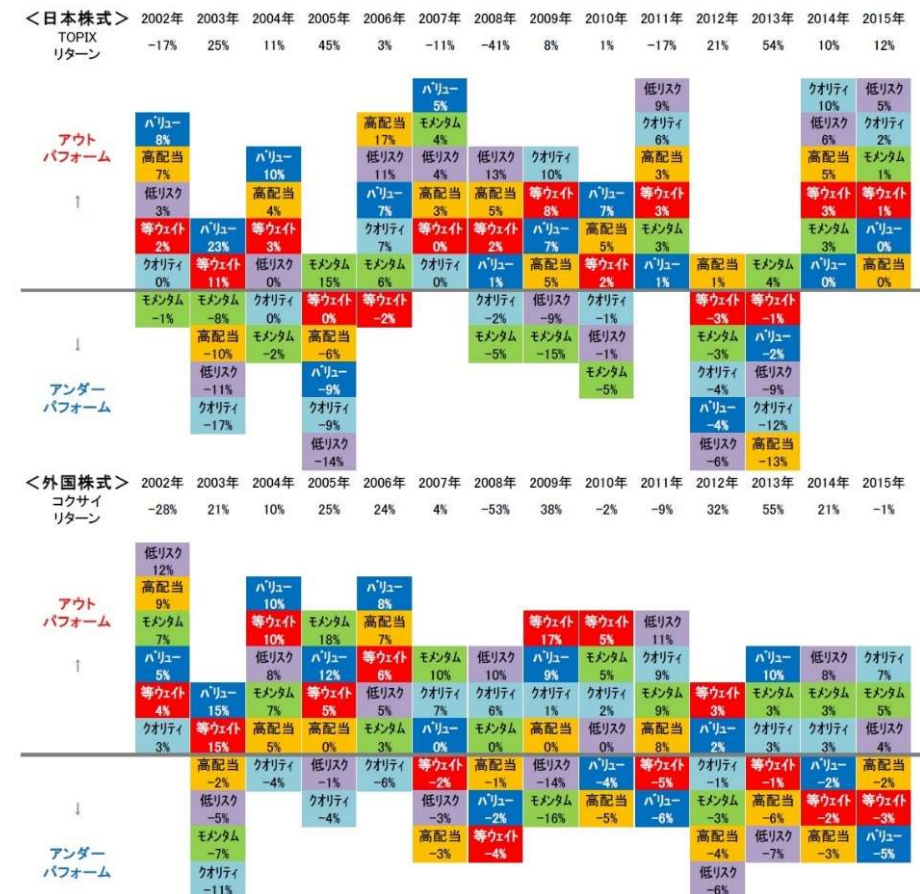
①マクロ指標と関連付け
景気回復期はバリュー、後退期はクオリティ
(Gupta et al [2014])

②市場動向と関連付け
高ボラ期はバリュー、低ボラ期はモメンタム
(Anmman[2009])

③モメンタムベット
ファクターの正の自己相関を利用
(Flogel et al [2019])

“ファクタータイミング戦略の有用性（リサーチ）”，
@hippoasset, 2019 より引用

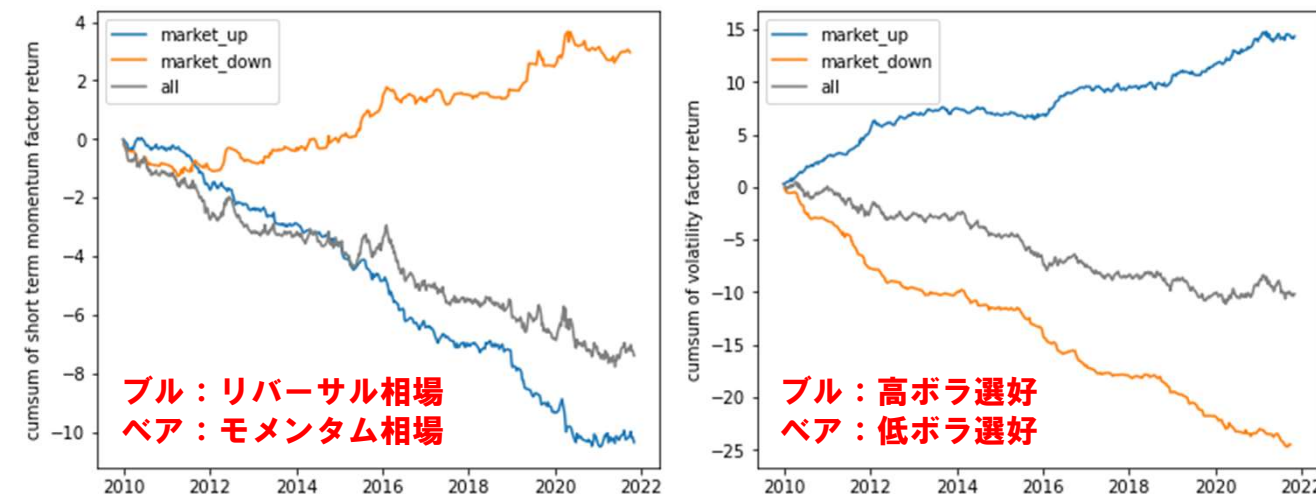
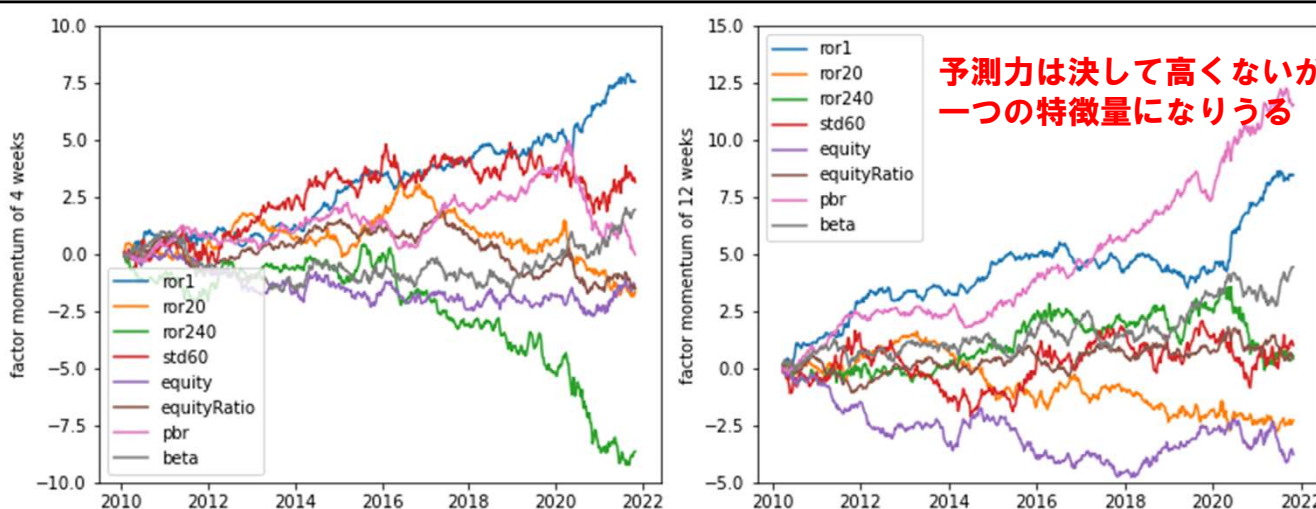
“リスク・パリティによるスマートベータの合成”，前山, 2016



Excess Return of Smart Beta

ファクタータイミングはコスト控除後のエッジは出にくいとされるが、
単一の戦略ではなく特徴量として取り込むことで十分にモデル改善が可能。

もっと短期的なファクタータイミングの予測可能性

手 法	概 要	事 例
マクロ市況 による 予測可能性	マーケットの 状況でファク ターリターン の挙動は大き く変わる	 <p>ブル：リバーサル相場 ベア：モメンタム相場</p> <p>ブル：高ボラ選好 ベア：低ボラ選好</p>
ファクター モメンタム	ファクターリ ターンには自 己相関が発生 する場合があ る	 <p>予測力は決して高くないが 一つの特徴量になりうる</p>

モデルに必ず加えるべき（予測可能性よりも賭けの対象が増えることが重要）

とりあえずクラスタリング③業種

業種クラスタリング：価格データを次元圧縮して分類する

そもそも企業の主力事業は変化するし、各業種の産業的な位置付けも変わる。
このため既存の業種分類は不適当である場合が多い。

“株価時系列に基づく企業クラスタリング”，松井，2020

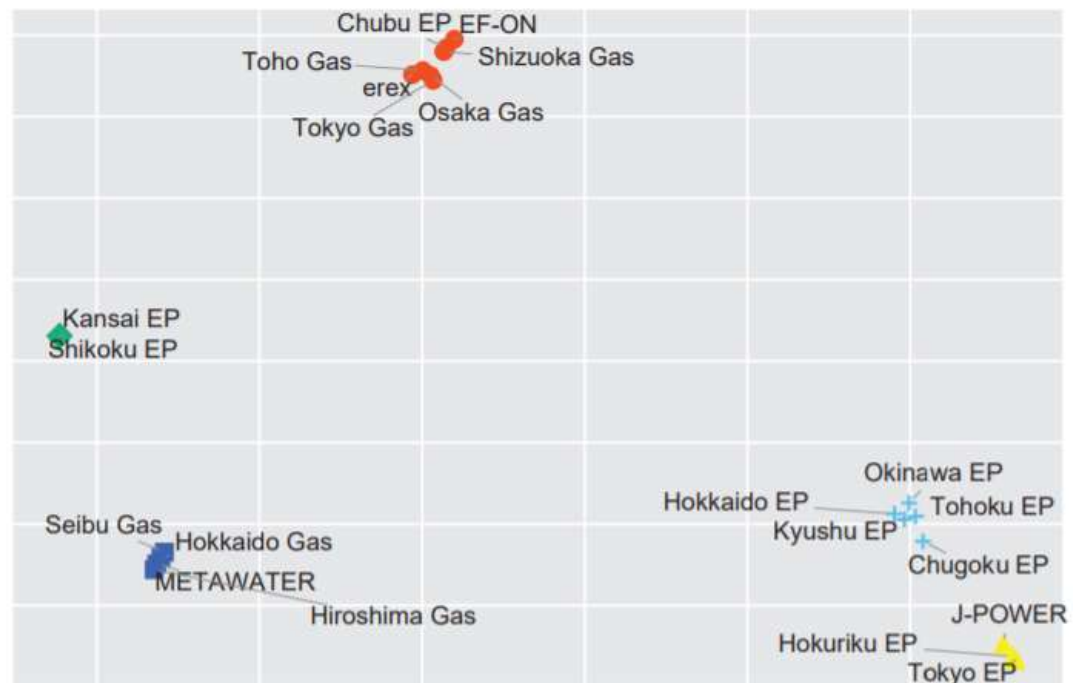
＜次元圧縮手法について＞

最近はUMAPが人気。

- PCA
- t-SNE
- UMAP

次元圧縮の対象は、
普通に株価データでよい
ファンダは難しい（実体験）

テーマ株の分類も定量的に
捉えることができる



Clustering Result for Electricity and Gas Industry

交互作用による既存のファクターリターンの性能向上を期待するが、
別に改善しなくてもよい。情報控除の手段が増えることが大事（後述）。

機械学習が作り出す④新しいファクター

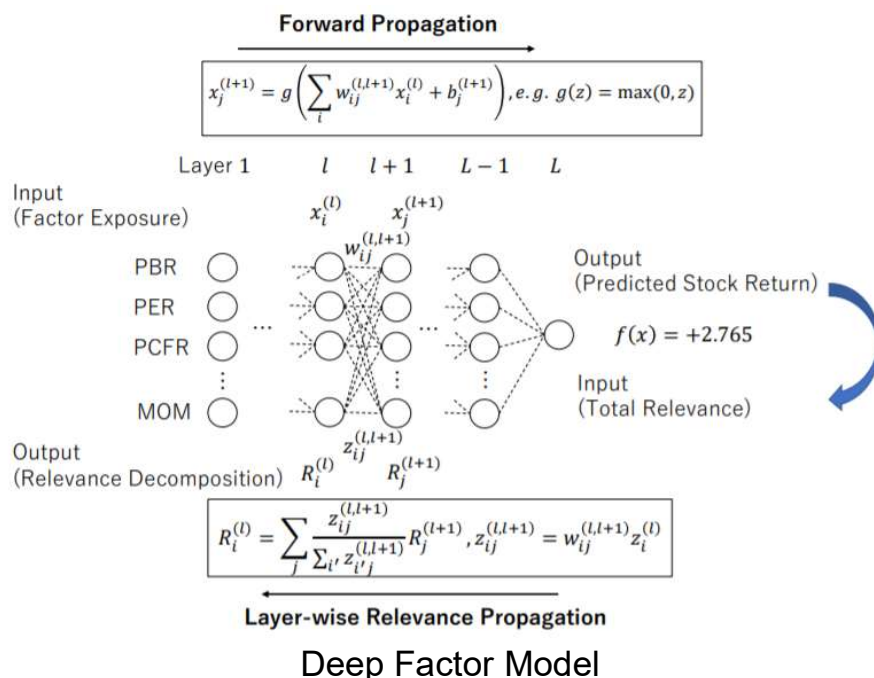
アプローチ：既存のファクターから合成・抽出する

エッジの根源はファクターの持つ非線形性や交互作用と考えられており、このことは様々な文献で言及される。

<事例1>

ファクターに深層学習を適用。

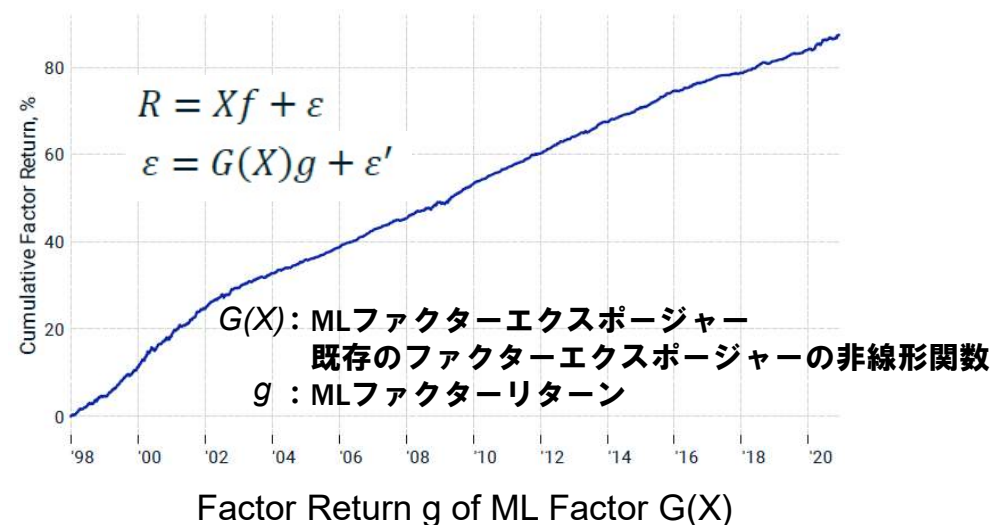
“Deep Factor Models”, Nakagawa, 2018



<事例2>

従来のBARRAファクターの残差分を説明する非線形関数Gを導入する。

“Machine Learning Factors Capturing Non Linearities in Linear Factor Models”, MSCI, 2021

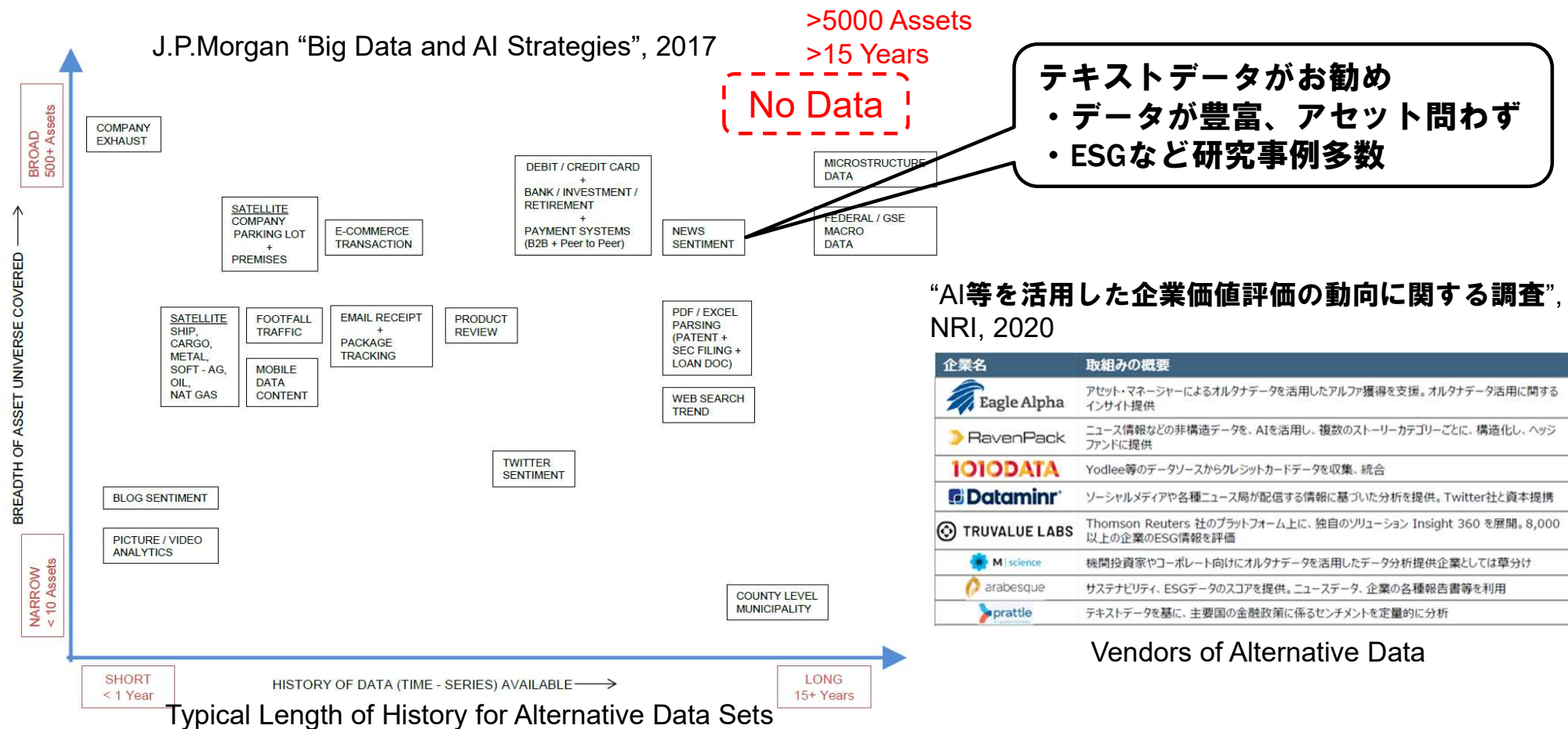


非線形モデルや関数の出力を直接Predictionとしてもよいが、後段に学習器を備えることを推奨する（スタッキング）

機械学習が作り出す④新しいファクター

アプローチ：全く新しいソースを用いる（オルタナティブデータ）

エッジの根源は収集コスト。定性的には「カバーしている人が少ない指標ほど効果的」。



ヒストリカルでユニバース全体をカバーするのは難しい。
もしも使うなら、スペシフィックリターンとの関係を観察すること。

最後に絞り取れ⑤スペシフィックリターン

手持ちのファクター（特徴量）全てを用いてリターンを直交化する

スペシフィックリターンのおさらい

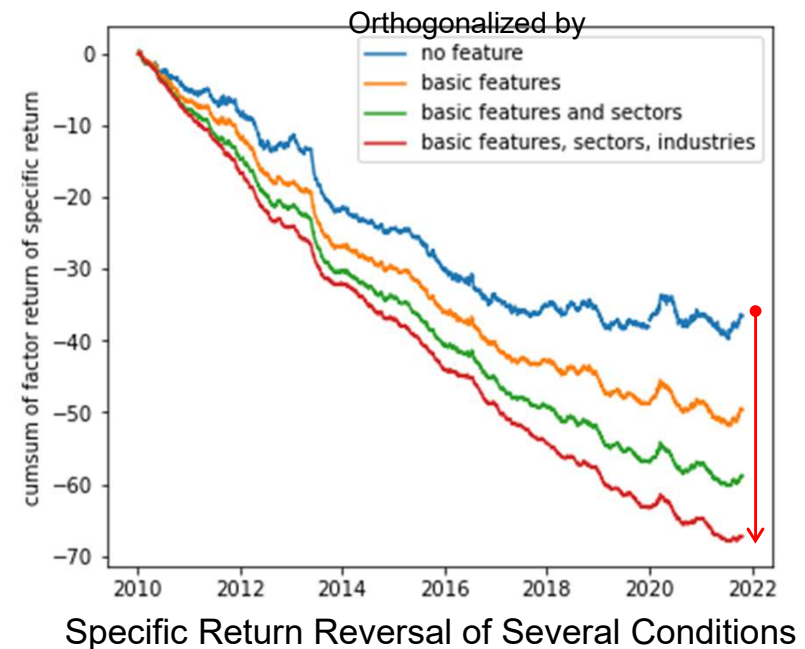
手持ちのファクターを
全て控除（直交化）した残差分

$$r_i = \sum x_{i,j} \cdot f_j + u_i$$

Numeraiの直交化と同じ
複雑な計算式は不要

`u=rt-LinearRegression().fit(x, rt)`

控除する情報を
増やすとどうなるか



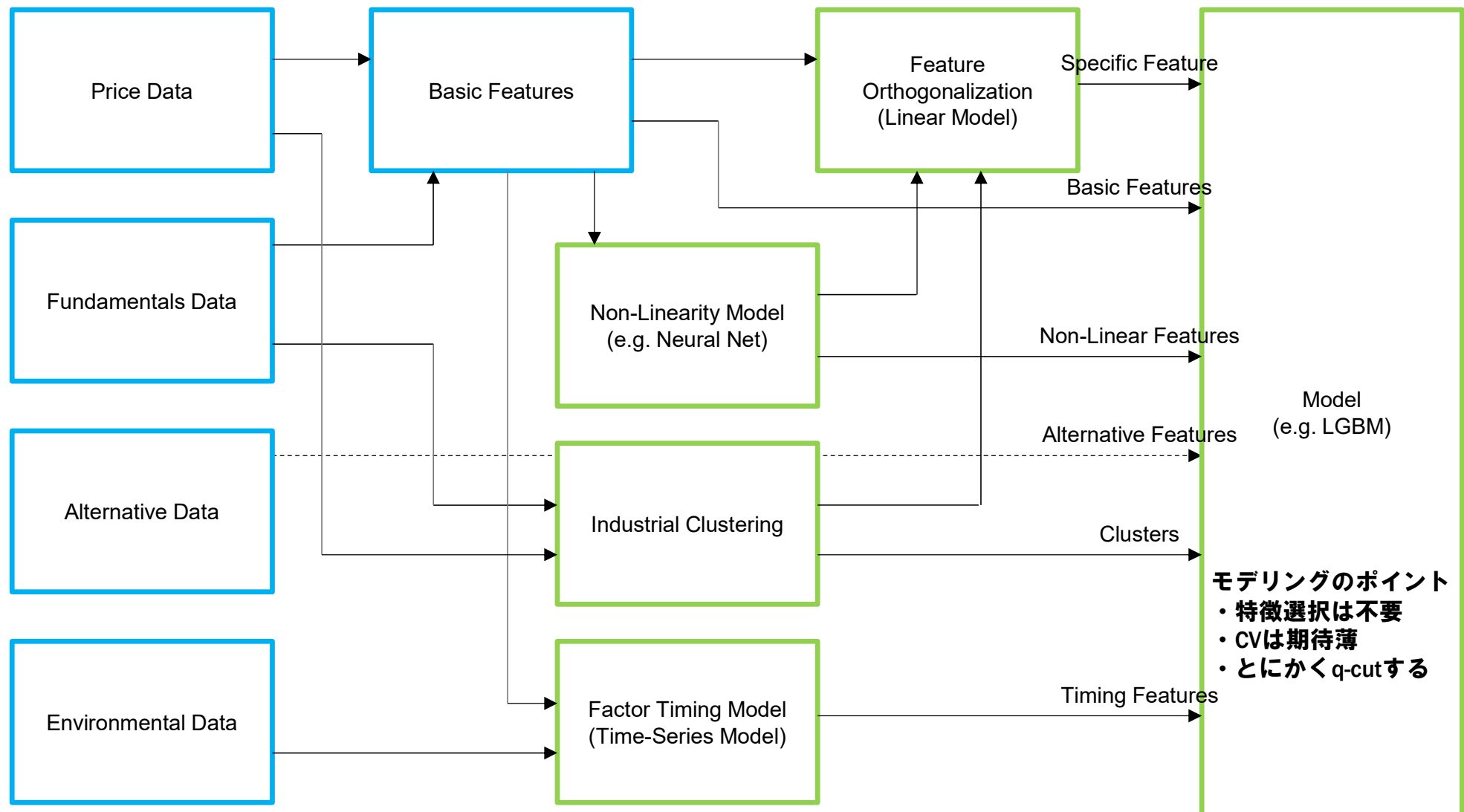
- ・スペシフィックリターンは強いリバーサル特性を示す
- ・（原理的に）情報控除するほど性能向上する（はず）
- ・市場によって特徴の強さが異なる。
- ・特に短いタイムフレームで効果がある。

【参考】最強の指標
スペシフィック・ギャップ

特徴量が出揃ったらオリジナルのスペシフィックリターンを算出すること。

まとめ

まとめ：伝統的クオンツを発展させるための機械学習モデリング



上記のようなモデリングで従来未考慮であった収益機会を獲得することを目指す

Signalsの奨め

思い：

エンジニアの方に自身のスキルの収益化を実現してほしい

正しい努力と正しいリスクテイクができれば自身の力で資産形成できる

<Signalsのメリット>

- ・資金の制約がない（通常の資金では組めないPF）
- ・資金効率が高い（マーケットニュートラルの資金効率を劇的に改善）
- ・売買コストが掛からない（ゼロコストでPF構築）
- ・知的財産が保護される（コードの提出不要）

<Signalsのデメリット>

- ・仮想通貨の売買・保有に伴う諸所のリスクが発生します。

本資料の内容に加え、データサイエンスのスキルがあれば始めることができる。
年末年始どう過ごしますか。

ご清聴ありがとうございました