



# Macroeconomic Predictions – a Machine Learning Approach

ODSC

15th June 2023

# Machine Learning and Econometrics

The two fields use different tools and mindset, but they can benefit from each other



Econometrics	Machine Learning
<p>Econometric models are based on economic theory, generally expressed through parametric models.</p> <p>Traditional statistical inference methods (such as maximum likelihood and the method of moments) are thus used to estimate the values of a vector of parameters <math>\theta</math>, in a parametric model <math>m_{\theta}(\cdot)</math></p>	<p>In ML, by contrast, non-parametric models are often built based almost exclusively on data (i.e. no distribution hypothesis), very often Big Data</p>

# Machine Learning and Econometrics

The two fields use different tools and mindset, but they can benefit from each other



Econometrics	Machine Learning
Econometrics focusses on studying the asymptotic properties of $\theta^*$ (viewed as a random variable, thanks to the underlying stochastic representation)	ML focuses to a greater extent on the properties of the optimal $m^*(\cdot)$ based on a criterion that has to be defined, or even simply $m^*(x_i)$ for observations $i$ deemed to be of interest for example in a test population
Econometrics uses mostly in-sample goodness-of-fit measures e.g. AIC, BIC	In ML the meta-parameters used (tree-depth, penalty parameter, etc.) are optimised by cross-validation.

# Machine Learning and Econometrics

The two fields use different tools and mindset, but they can benefit from each other



Econometrics	Machine Learning
<p>Econometric models are simple to interpret built with a significant amount of human input.</p> <p>Econometrics is a fundamentally a process of human learning. It is fundamentally a search for knowledge and understanding, and not just an ability to make better predictions.</p>	<p>ML focuses on prediction robustness while sometimes sacrificing transparency.</p>

# Inflation is a Hot Topic



Article | 7 October 2021 | Read in 6 minutes

The inflation scare in financial markets is here to stay

POLITICS

## Investors are worried about inflation data, but economists see it as a sign of temporary growth

PUBLISHED WED, MAY 12 2021-12:02 PM EDT | UPDATED WED, MAY 12 2021-2:54 PM EDT

The editorial board DECEMBER 1 2021

Opinion **The FT View**

## Transitioning away from 'transitory' inflation

Federal Reserve chair Jay Powell is right to retire word that has caused confusion

## Inflation is traumatic for the Fed and existential for the eurozone

2 December 2021 • 6:00am

The European Central Bank can no longer deny the reality of inflation nor hide behind the Fed

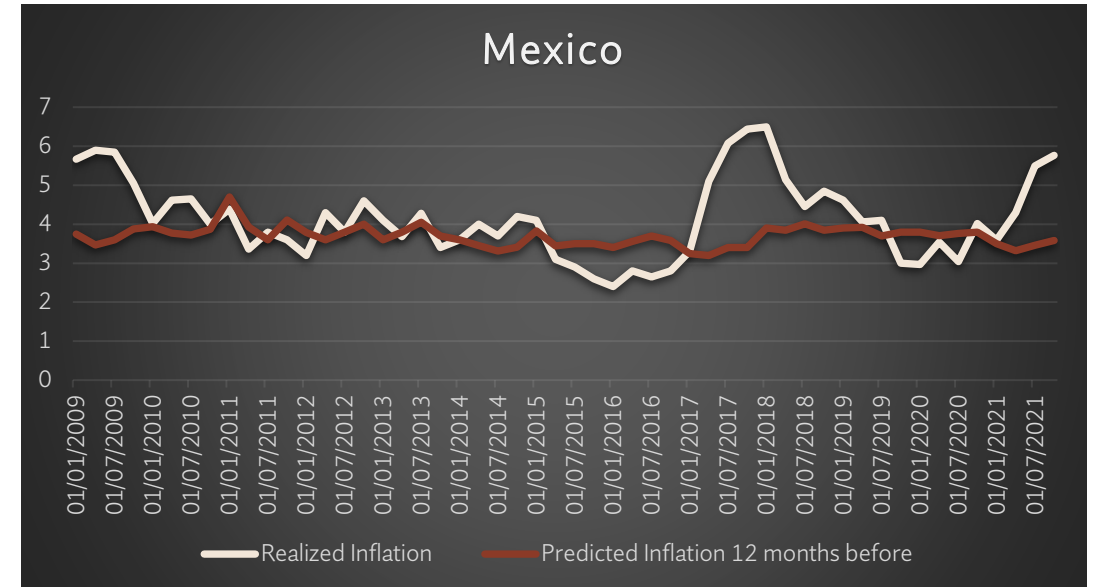
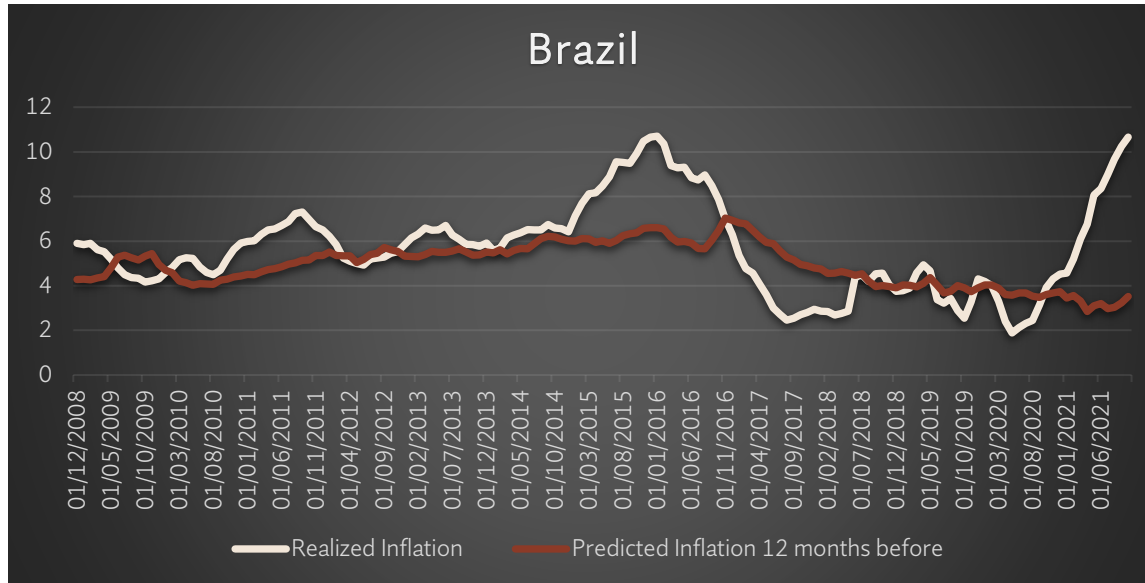
**S&P Global**  
Ratings

## Emerging Markets Will Inflation Be The Next Pandemic?

# An Issue Illustrated



Consensus estimates and traditional econometric models are commonly centred around the long-term inflation average with little variability. Bigger fluctuations in inflation are rarely predicted



# Can We Predict Better– a Big Data Approach

## How forecasting was done: Pre-Big Data age

### Models

- Approach based on descriptive models with **limited parameters**
- Models contain many assumptions about the relationships between variables

### Data

- Data is mainly used to **estimate model parameters**
- Limited data types used - only **few** exogenous and endogenous variables of similar nature

## How we do it: Big Data age

- Models with **many more parameters** that contain very limited pre-conceived assumptions about the relationship between variables
- Models exploit latest advances in **Machine Learning**

- Data is used to **estimate model parameters** and to **enhance predictive ability** (~300 variables per country)
- **Significantly more data** used to fit models with **multiple data types**: numerical, text, images ...

# How We Do It – a Big Data Approach (cont'd)

We use a variety of drivers which include traditional macroeconomic and alternative data variables, in total  
~300 per country

## Traditional Data Sources

### Output, Production, Business Activity

- Total Industrial Production
- Preliminary GDP
- Factory Orders, etc.

### Retail

- Consumer Prices
- Personal Consumption Expenditure
- Retail Trade, etc.

### Labour

- Wage Changes
- Unemployment
- Job Openings & Labor Turnover, etc.

### Sentiment & Expectations

- Manufacturing Confidence
- Consumer Confidence
- Import & Export Forecasts, etc.

### Money, Credit & Interest Rates

- Producer Prices
- Capital Account Statistics
- National Balance Sheets
- Money Supply
- Debt Statistics
- Government Balance Statistics
- CPI, etc.

Etc.

## Alternative Data Sources

### Output, Production, Business Activity

- Air Pollutants & Air Quality & Rainfall
- Electricity Consumptions & Shortages
- Factory Capacity Utilisation
- Oil Rig Counts
- Natural Resources Stock levels (incl. injections & withdrawals), etc.

### Retail

- Airport Arrivals and Departures
- Restaurant visits & reservations
- Ground, Sea & Air Traffic
- Entertainment ticket sales
- Overnight stays and expenditures
- Vehicle Registrations, etc.

### Geopolitical

- Geopolitical Risk News
- World Trade Uncertainty, etc.

### Labour

- Online Job Adverts
- Slack labor statistics, etc.

### Sentiment & Expectations

- News Sentiment
  - Food
  - Energy
  - Communications
  - Entertainment
  - Housing, etc.



# Algorithms



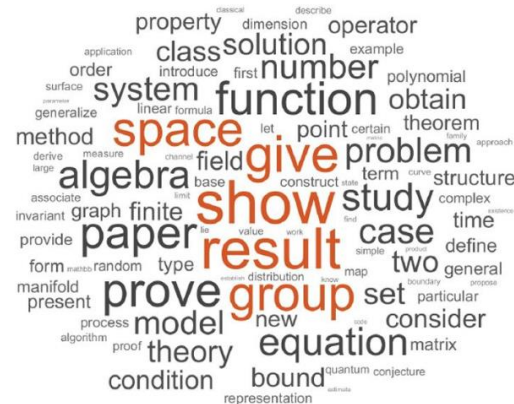
ML algorithms are needed in alternative data world to extract features from a variety of unstructured data sources

- **Features extraction** – unstructured data usually requires ML algorithms

Image processing (e.g., CNNs)



Text/sound processing (e.g., RNNs)



Geofencing for spatial datasets



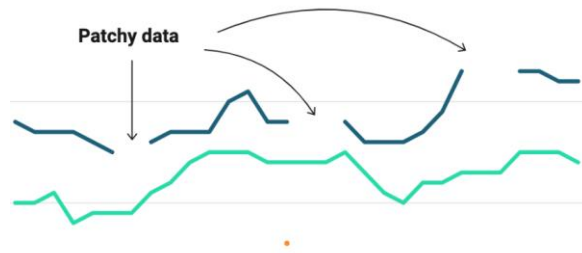
# Algorithms

Data needs to be curated to ensure data cleanliness, reliability and that it is fit for purpose

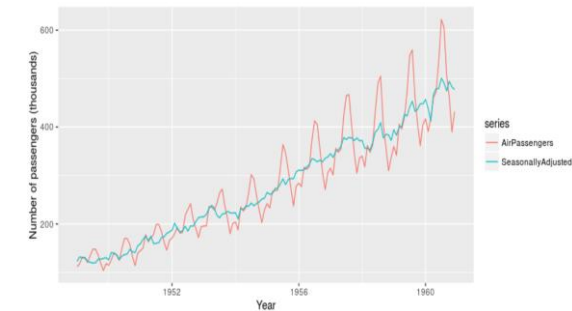


- Data Transformations

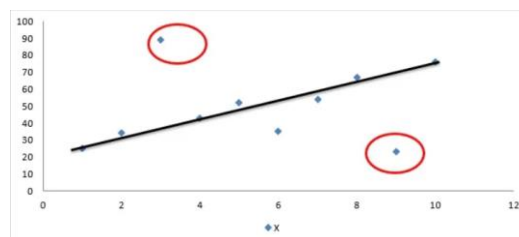
Missing Data (e.g., EM algorithm)



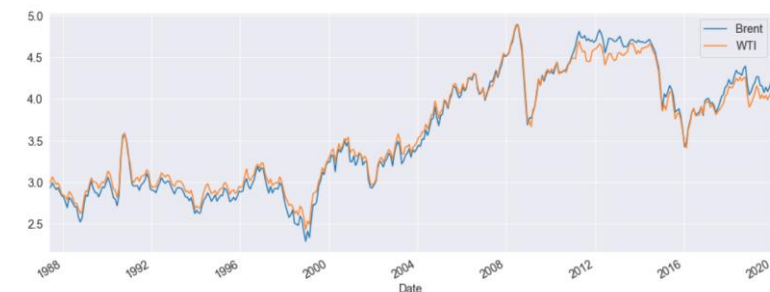
Making variables de-seasonalized (e.g., ARIMA X-12)



Outlier removal (e.g., Mahalanobis distance)



Search for cointegrating variables



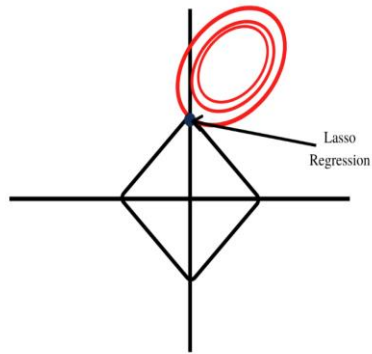
# Algorithms



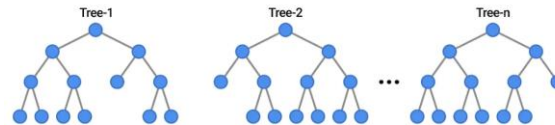
Novel algorithms are needed in alternative data world to deal with the variety of potential explanatory variables as well as their (big) number, potential non-linearities and covariate shifts

- **Forecasting** – since we deal many more variables when we augment our predictive models with alternative data, we must use algorithms that allow sparsity to avoid overfitting

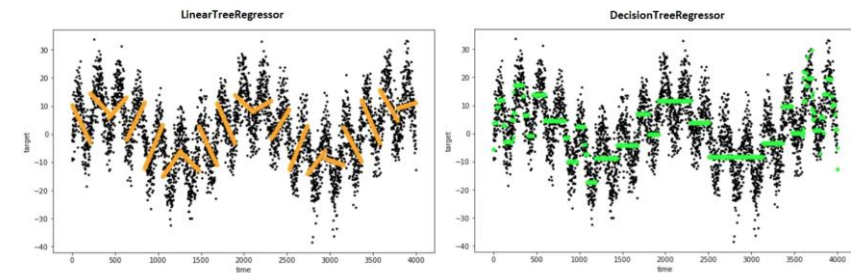
Linear (e.g., LASSO, Ridge, Elastic Nets)  
able to extrapolate, bad with non-linearities



Random Forest (RF)  
unable to extrapolate, good with non-linearities



Linear Random Forest (LRF)  
able to extrapolate and deal with non-linearities

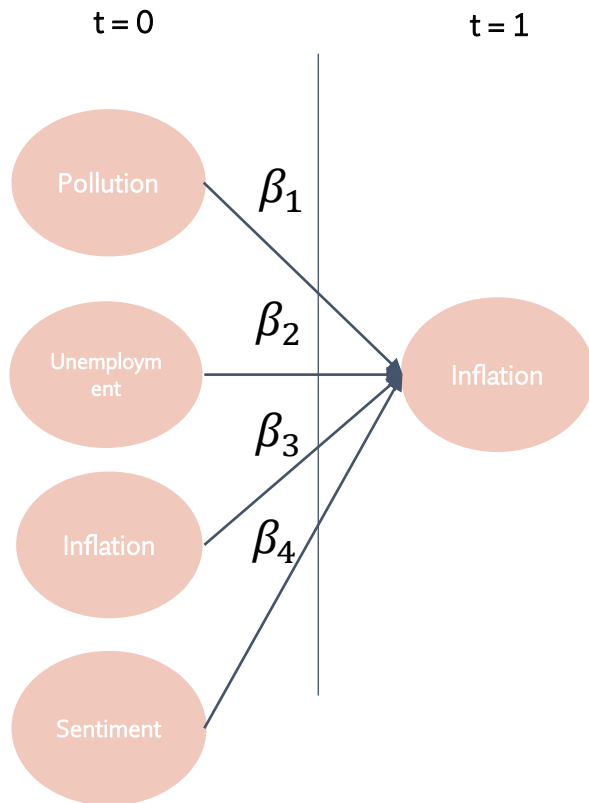


# Forecasting & Bias

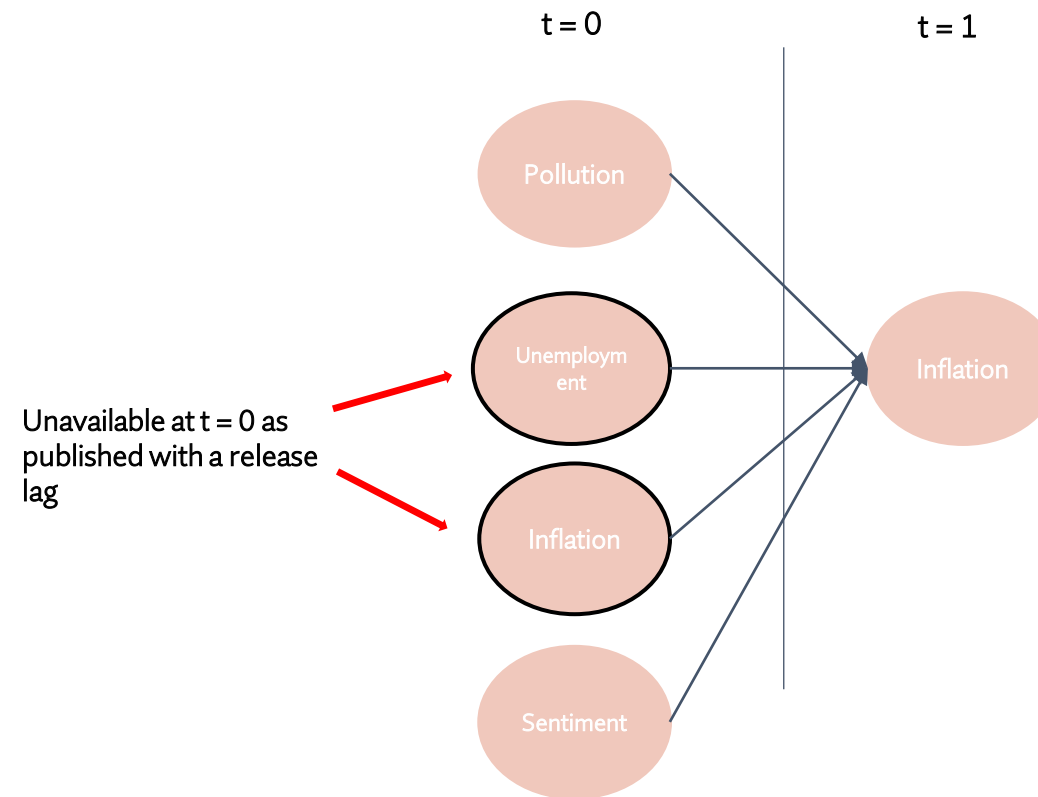


Since we aim to maximize predictive power sometimes by using a big number of variables, we can get biased coefficients that hide the true causal impact, but this does not impact predictive performance.

If this was the true causal model...



...it cannot be calibrated as many of the variables are non observed at time  $t=0$



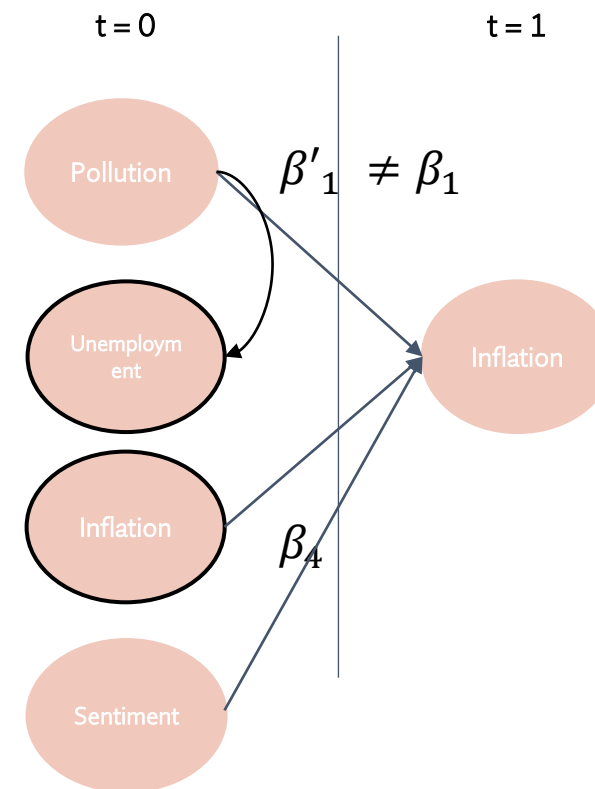
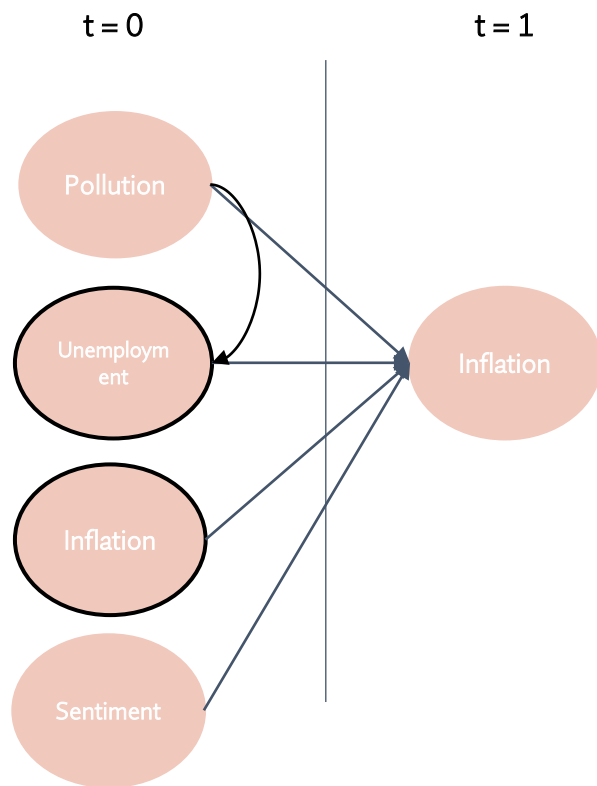
# Forecasting & Bias



Since we aim to maximize predictive power sometimes by using a big number of variables, we can get biased sensitivities that hide the true causal impact, but this does not hinder predictive performance.

The independent variables can be also correlated (or better sometimes - causally related)...

...which biases the coefficients estimations in the case of omitted variables which are now being proxied by other variables



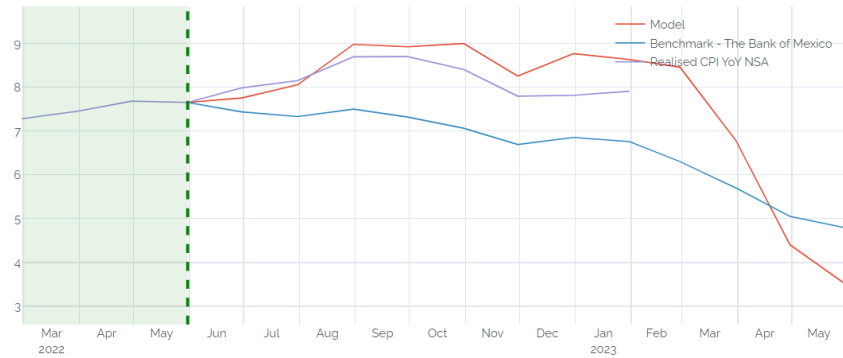
# What has been predicted so far

Live performance since May 2022

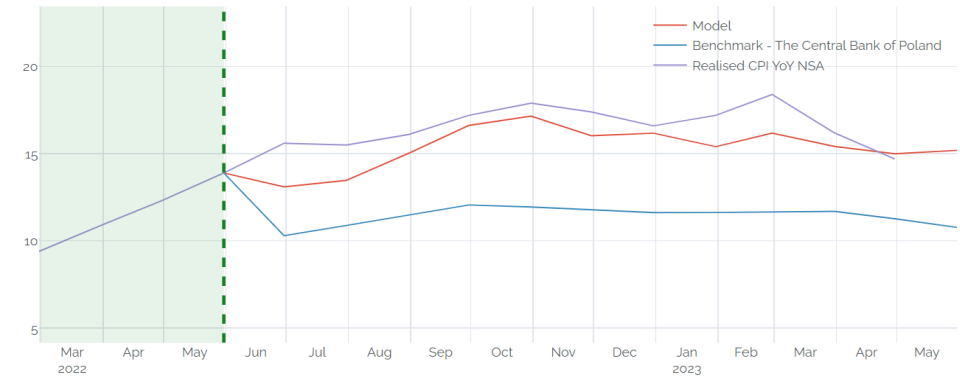


## Forecast Plot

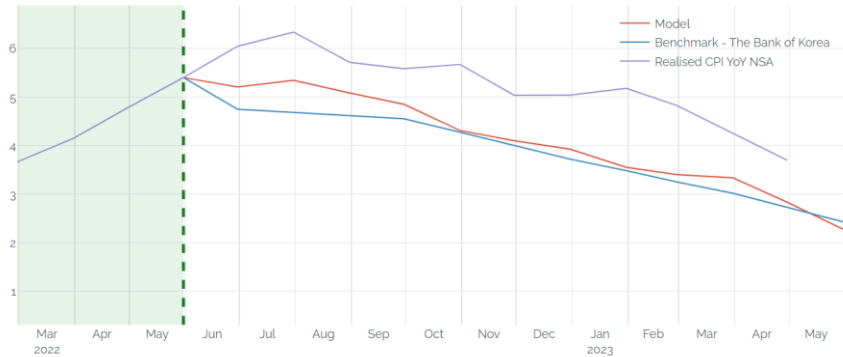
Turnleaf Analytics Model vs. Benchmark Consensus CPI YoY NSA for Mexico, May 2022



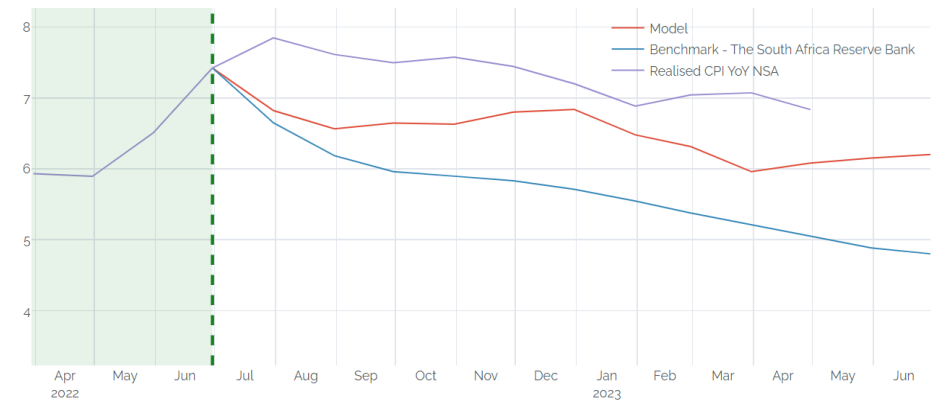
Turnleaf Analytics Model vs. Benchmark Consensus CPI YoY NSA for Poland, May 2022



Turnleaf Analytics Model vs. Benchmark Consensus CPI YoY NSA for South Korea, May 2022



Turnleaf Analytics Model vs. Benchmark Consensus CPI YoY NSA for South Africa, Jun 2022

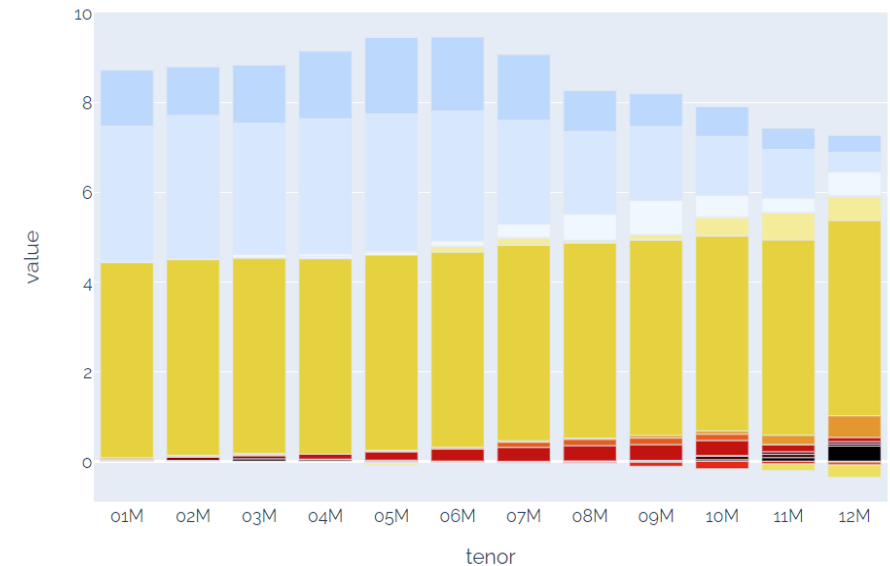




- If models are non-interpretable (e.g., Random Forests) interpretability methods can be used, such as:

- ## Contribution Plot

- variable
- consumer prices
- consensus
- FX, equities & commodities
- government
- global macro
- geopolitical stability
- long term average
- labor
- money, credit & interest rates
- output, production & business activity
- producer prices
- retail
- sentiment & expectations
- trade & investment

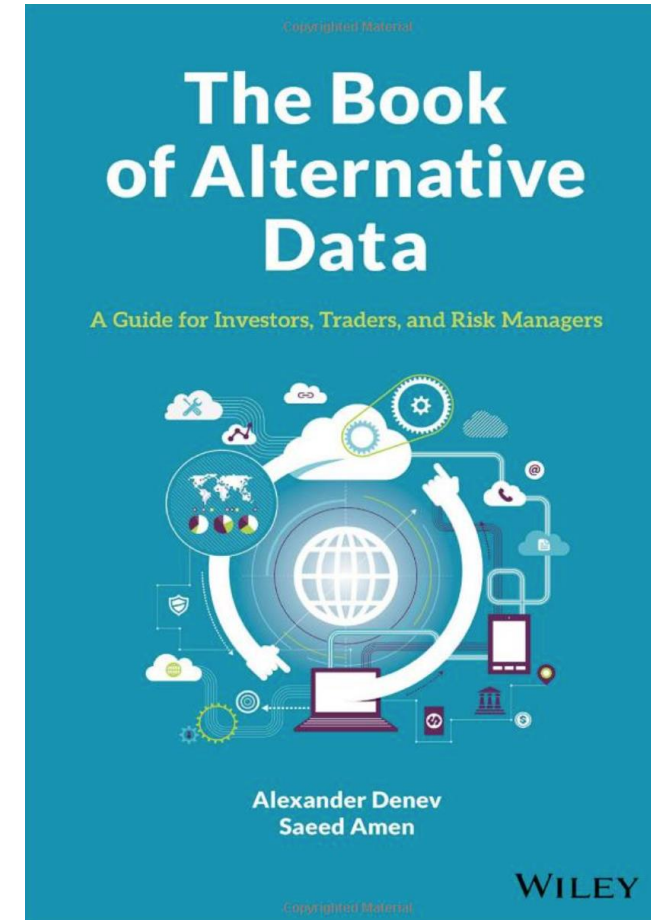


# Other Challenges

Practical implementations are fraught with challenges that need to be overcome with the right processes' setup



- Complexity of managing diverse data sources (>500)
  - Data onboarding
  - Legal agreements
  - Data vendors risk management
  - Large software/technology stack
  - Internal processes setup
- Data variability
  - Accuracy
  - Technical outliers
  - Interruption of service
  - Rebasing and revisions of data
- Client Communication
  - Attribution Plots
  - Mailing lists (AI generated)
  - Product Support processes





# Applications of inflation forecasts

Can be used for trading and investing, as well as forward planning by corporates

- Macro instruments directly tied to inflation
  - Inflation swaps
  - TIPS/linkers – inflation linked bonds
- Macro instruments indirectly tied to inflation
  - Fixed Income
  - FX
- Forward planning by corporates
  - Pricing of their product
  - Understanding wages

# Monetary policy & FX

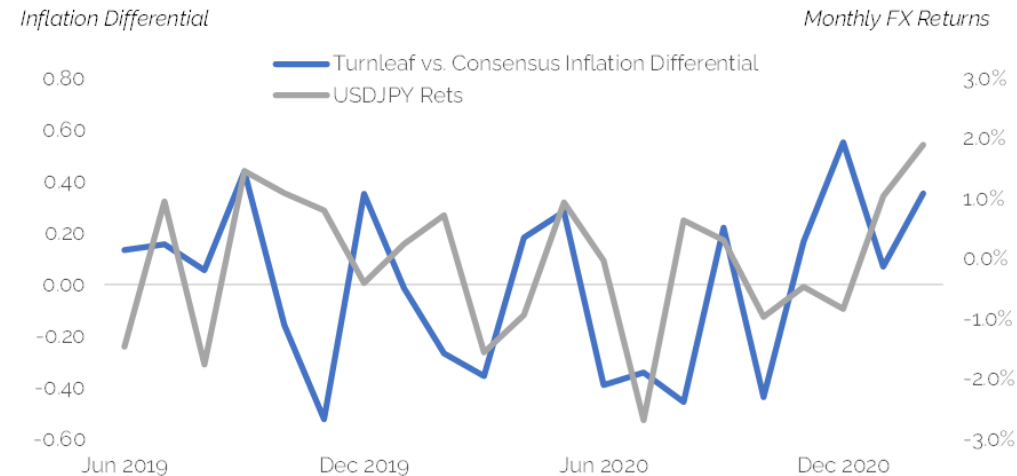
Inflation impacts relatively monetary policy & hence FX

- In developed markets (& an expectation of credible central banks)
  - High inflation is accompanied by increasing yields as market expects hikes from hawkish central banks
  - Low inflation is accompanied by falling yields as markets expects cuts from dovish central banks
- Not necessarily the case in some of EM...
  - Can have the dynamic of investors dumping local currency and bonds at the same time
- Indeed, hyperinflation is not a reason to buy a currency!
- In summary... inflation => monetary policy expectations => FX
- For FX, it is the relative monetary policy expectations which matter
  - There is a difference between looking at
  - Absolute yield differential (carry)
  - Relative monetary policy (growth)

# Relative inflation differentials for FX

Instead of relative yields to understand FX, look at relative inflation differentials?

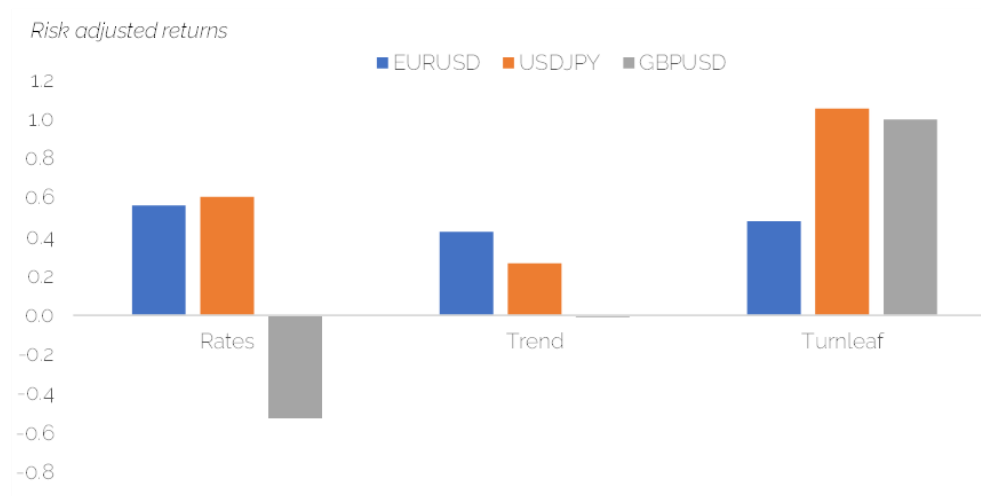
- Instead of looking at relative yields (which market participants tend to look at)
  - As an input for trading FX
- What about...
  - Inflation differential forecast vs. benchmark inflation differential forecast
  - How much more/less inflation differential are we pricing for inflation on a relative basis compared to the market?



# Comparing inflation FX trading rule

Against other common strategies in FX

- Focus on the most liquid FX pairs
  - Historical sample is from 2018 onwards
- Compare against trend and rates strategies for FX
  - Important to use a benchmark

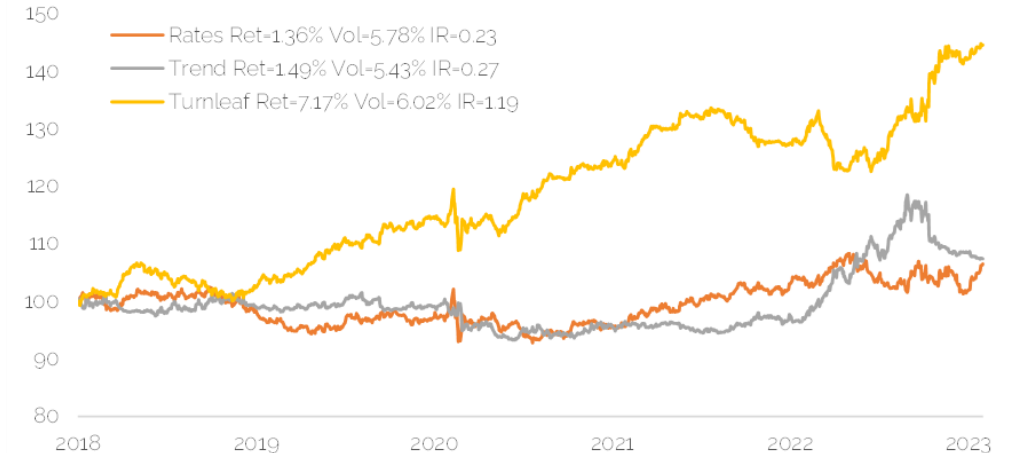


# Creating an FX inflation basket

Correlation against other strategies

- On a historical basis the inflation strategy for FX outperformed trend and rates
- Strategy is decorrelated to trend
- Has some correlation to rates, but outperforms

	Rates	Trend	Turnleaf
Rates	100%	-11%	28%
Trend	-11%	100%	-29%
Turnleaf	28%	-29%	100%



The background of the slide is composed of two main parts. On the left and right sides, there are close-up, vertical photographs of green, elongated leaves, possibly from a plant like a banana or a similar tropical species. These leaves are slightly out of focus, creating a soft, naturalistic texture. In the center, there is a vertical rectangular area with a light pink or peach-colored background. Overlaid on this pink area is a subtle, abstract geometric pattern consisting of several thin, intersecting lines in a slightly darker shade of pink, creating a series of soft, triangular and quadrilateral shapes.

Thank you