
The Makridakis Open Forecasting Center (MOFC)
Advancing the Theory and Practice of Forecasting
THE M COMPETITIONS
THE M CONFERENCES
THE M PUBLICATIONS
THE M COURSES
UNIVERSITY of NICOSIA

Forecasting and Uncertainty: Four Distinct Types with Different Risk Implications

Forecasting, Uncertainty and Risk

	KNOWN	UNKNOWN
KNOWN	**I. Known/Knowns** (Normal conditions, Law of large numbers, independent events, wisdom of the crowds) Forecasting: Accuracy measurable Uncertainty: Thin tailed and measurable Risks: Manageable (e.g. having inventories) **Only works in Usual/normal times**	
UNKNOWN		

Forecasting and Uncertainty: Four Distinct Types with Different Risk Implications

Forecasting, Uncertainty and Risk

	KNOWN	UNKNOWN
KNOWN	**I. Known/Knowns** (Normal conditions, Law of large numbers, independent events, wisdom of the crowds) Forecasting: Accuracy measurable Uncertainty: Thin tailed and measurable Risks: Manageable (e.g. having inventories)	**III. Unknown/Knowns** (Cognitive biases, Strategic actions, self-fulfilling and self-defeating prophecies, game theory) Forecasting: Purely Judgmental Uncertainty: Extensive/hard to measure Risks: Depends on biases, strategic actions/reactions
UNKNOWN	**II. Known/Unknowns** (Special settings, effects of the next recession on economy/firms, madness of crowds) Forecasting: Inaccuracy can vary considerably Uncertainty: Fat tailed, hard to measure Risks: Can be substantial, tough to manage	**IV. Unknown/Unknowns (Black Swans)** (Black Swans: Low probability high impact events, e.g. implications of the total collapse of global trade) Forecasting: Impossible Uncertainty: Infinite Risks: Unmanageable, need for antifragile strategies

We now operate in a usual/special period

A White Swan

WORLD ECONOMIC OUTLOOK REPORTS

World Economic Outlook Update, June 2020

June 2020

العربية español français 日本語 русский 中文

A Crisis Like No Other, An Uncertain Recovery

Read full report PDF Download the Data

Global growth is projected at –4.9 percent in 2020, 1.9 percentage points below the April 2020 World Economic Outlook (WEO) forecast. The COVID-19 pandemic has had a more negative impact on activity in the first half of 2020 than anticipated, and the recovery is projected to be more gradual than previously forecast. In 2021 global growth is projected at 5.4 percent. Overall, this would leave 2021 GDP some 6½ percentage points lower than in the pre-COVID-19 projections of January 2020. The adverse impact on low-income households is particularly acute, imperiling the significant progress made in reducing extreme poverty in the world since the 1990s.

IMF sees less severe global recession in 2020

Published date: 13 October 2020

Share:

The global economy is likely to contract less severely this year as a result of the Covid-19 pandemic than previously expected, the IMF said today.

But nearly every major economy next year will still be below 2019 levels, the IMF said in its latest *World Economic Outlook* report, which projects that the global economy will shrink by 4.4pc this year.

The IMF in June was anticipating a sharper contraction of 4.9pc this year, but it has since changed the metrics it uses to evaluate economic activity. Under the new metrics, the revised projection is an upgrade of 0.8 percentage points from its previous forecast. The IMF forecasts the global economy to grow by 5.2pc in 2021.

The less downbeat forecast reflects a lower than forecast contraction in the US economy and the eurozone during the second quarter, which the IMF attributes to trillion-dollar stimulus packages that helped consumer demand rebound following relaxation in travel and economic activity imposed to contain the pandemic.

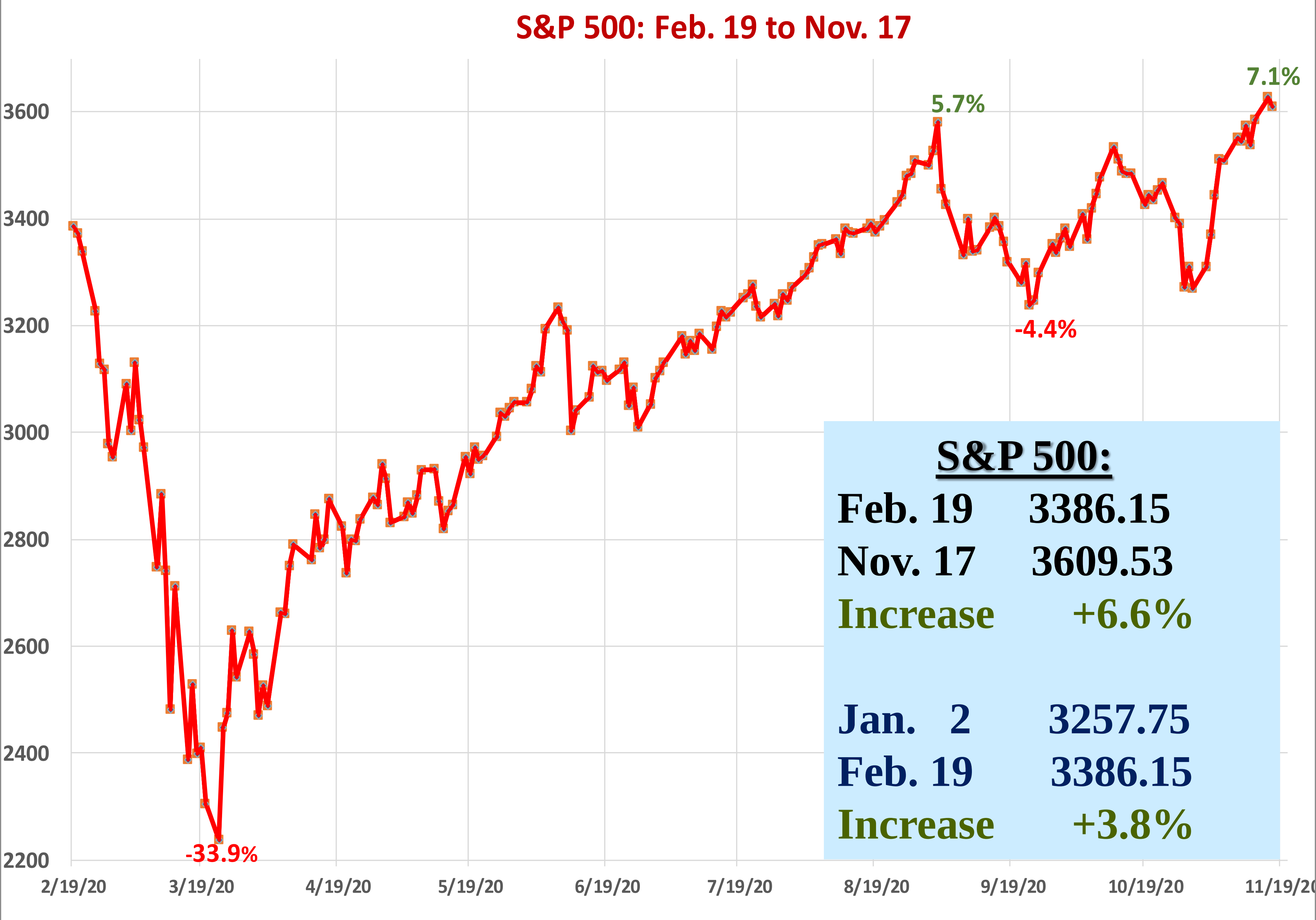
China's emergence from the economic downturn also proved stronger than expected. The report projects China's economic growth at 1.9pc this year and

IMF 2020 GNP Forecasts

Jan.	3.3%
April	-3.0%
June	-4.9%
Oct.	-4.4%

The Great Recession

GDP 2007Q4 - 2009Q2: -4.3%



S&P 500: Feb. 19 to Nov. 17
3600
3400
3200
3000
2800
2600
2400
2200
2/19/20
3/19/20
4/19/20
5/19/20
6/19/20
7/19/20
8/19/20
9/19/20
10/19/20
11/19/20
-33.9%
5.7%
-4.4%
7.1%
S&P 500:
Feb. 19 3386.15
Nov. 17 3609.53
Increase +6.6%
Jan. 2 3257.75
Feb. 19 3386.15
Increase +3.8%

Applied Forecasting Course: The M and Other Competitions

- Before the M Competitions
- Competitions: The equivalent of experimentation
- M1: Simple Methods and Combining Forecasts
- M2: The Role of Judgment in Forecasting
- M3: Theta the Most Accurate Method
- M4: The Forecasting Spring
- Other Forecasting Competitions
- M5: Major Findings

Professor Spyros Makridakis,
Director, Institute For the Future (IFF)
and the MOFC, University of Nicosia

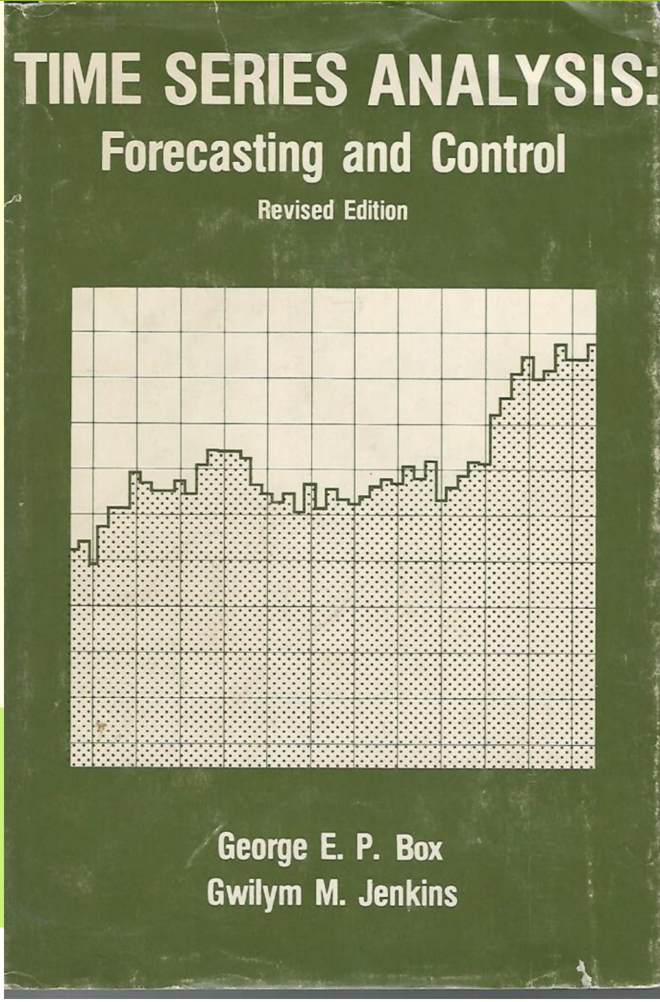
Before the M Competitions: No distinction between model fitting and forecasting

- There is an Appropriate, Best Model for every Single Time Series
- Identify such a Model judgmentally by studying each Time Series
- Estimate its Parameters and Check its Adequacy by making sure that its residuals are random
- If adequate Use it to Forecast; Otherwise, Identify Another Model

The single objective: Fitting the Most Accurate Model to Past Data

- No Worry about Overfitting

The Bible of the time



Forecasting Competitions

- The equivalent of experimentation in physical sciences
- How the M1 competition started

Percentage of time that Naive 2 is better than other methods listed (*111 series*)

	Forecasting Method	Model Fitting	1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1	82.7	50.0	49.1	56.4	60.9	59.1	57.3	65.5	67.3
	2 Single moving average	78.2	51.8	53.6	53.6	60.9	62.7	60.9	65.5	67.3
	3 Single exponential smoothing	79.1	50.9	51.8	56.4	62.7	61.8	62.7	66.4	70.9
	4 Adaptive response rate exponential smoothing	94.5	52.7	56.4	58.2	67.3	69.1	69.1	73.6	74.5
	5 Linear moving average	85.5	60.9	68.2	65.5	70.0	75.5	71.8	72.7	71.8
	6 Brown's linear exponential smoothing	80.9	47.3	54.5	58.2	63.6	67.3	63.6	73.6	78.2
	7 Holt's (2 parameters) linear exp. smoothing	83.6	45.5	53.6	54.5	62.7	64.5	60.9	68.2	70.0
	8 Brown's quadratic exponential smoothing	81.8	52.7	55.5	56.4	60.0	65.5	61.8	67.3	74.5
	9 Linear trend (regression fit)	87.3	65.5	61.8	64.5	68.2	72.7	70.0	72.7	75.5
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	54.5	63.6	70.9	71.8	69.1	70.0	70.9	71.8	70.0
	11 Winters' Linear and Seasonal Exp. Smoothing	67.3	49.1	57.3	54.5	51.8	53.6	55.5	57.3	57.3
	12 Adaptive Filtering	71.8	47.3	60.0	60.0	61.8	62.7	60.9	61.8	60.0
	13 Autoregressive moving average (Box-Jenkins)	59.1	46.4	53.6	55.5	57.3	51.8	58.2	61.8	60.9
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	5.5	33.6	39.1	40.0	39.1	43.6	42.7	42.7	37.3
	15 Single moving average	4.5	40.9	44.5	50.0	47.3	47.3	48.2	48.2	48.2
	16 Single exponential smoothing	40.0	42.7	59.1	63.6	63.6	64.5	67.3	62.7	60.0
	17 Adaptive response rate exponential smoothing	32.7	53.6	58.2	60.0	61.8	61.8	63.6	62.7	66.4
	18 Linear moving average	17.3	45.5	52.7	55.5	55.5	56.4	56.4	58.2	59.1
	19 Brown's linear exponential smoothing	18.2	43.6	47.3	49.1	50.0	51.8	54.5	51.8	50.9
	20 Holt's (2 parameters) linear exp. smoothing	20.0	46.4	54.5	55.5	56.4	55.5	53.6	61.8	61.8
	21 Brown's quadratic exponential smoothing	49.1	61.8	66.4	65.5	61.8	63.6	61.8	60.9	59.1
	22 Linear trend (regression fit)									

The Three Major Findings of the M1 Competition

1. Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones.

MAPE: Comparisons Deseas. Single Exponential Smoothing and Box-Jenkins								
Forecasting Period								Average
	1	2	3	6	8	12	18	1 - 18
Deseas. SES	7.8	10.8	13.1	17.2	16.5	13.6	30.1	16.8
Box-Jenkings	10.3	10.7	11.1	17.1	18.9	16.4	34.2	18.0
Difference	**2.50**	**0.10**	**2.00**	**0.10**	**2.40**	**2.80**	**4.10**	**1.20**

The Three Major Findings of the M1 Competition

1. Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones.
2. The Accuracy of various methods being combined outperforms, on average, the accuracy of the individual methods involved.

Big, Heated Arguments with the statisticians of the time about these two findings

The Three Major Findings of the M1 Competition

1. Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones.
2. The Accuracy of various methods being combined outperforms, on average, the accuracy of the individual methods involved.
3. The Uncertainty of the assemble is considerably smaller than that of the individual methods, meaning that the prediction intervals around the point forecasts are tighter.

The Major Finding of the M2 Real Time Competition

The Accuracy of Five Human Forecasters Was Lower than that of Simple Statistical Methods

Real Time Budget Forecasting for Two Years

MAPE all series (period Oct 1987–Dec 1989)

Methods	Forecasting horizons									Average year	Overall average
	1	2	3	4	5	6	8	12	15		
Naive 1	8.3	16.4	20.7	22.2	35.6	34.6	33.3	15.4	28.1	23.6	21.9
Naive 2	9.1	7.6	14.0	14.5	19.2	12.9	15.0	15.4	16.4	14.0	13.3
Method O/S	4.2	8.7	9.1	14.1	20.0	16.6	12.0	14.3	15.1	13.9	12.6
Single O/S	4.8	8.1	10.0	13.5	18.5	12.3	10.8	16.3	16.3	13.0	11.9
Holt O/S	4.6	8.9	10.1	13.9	20.7	14.4	10.2	18.0	14.4	14.1	12.8
Dampen O/S	4.1	8.1	10.9	13.8	19.2	14.3	10.8	17.5	15.1	14.1	12.8
Single	6.5	7.7	10.9	13.4	19.2	13.9	11.9	15.3	14.6	12.8	11.9
Holt	5.0	9.1	9.0	14.7	20.1	17.3	11.9	19.3	15.3	14.6	13.2
Dampen	4.6	8.0	9.9	14.1	19.1	14.4	10.6	16.3	14.4	12.6	11.6
Long	6.4	17.9	18.1	21.0	39.6	42.1	29.3	16.7	18.7	23.9	22.0
Box–Jenkins	9.8	11.0	14.6	16.7	20.9	20.6	16.8	19.3	20.6	17.0	16.0
Forecaster A	6.3	9.6	10.4	15.9	28.4	14.9	13.5	17.6	17.4	14.9	13.7
Forecaster B	5.1	11.6	8.9	19.2	28.6	23.6	15.9	24.6	26.1	22.3	19.5
Forecaster C	3.2	12.2	13.8	22.5	28.6	22.2	17.7	20.7	18.3	18.2	16.5
Forecaster D	5.9	14.2	14.9	19.5	26.0	22.0	21.5	23.7	22.9	21.2	19.3
Forecaster E	7.6	9.7	10.3	15.4	21.5	15.9	13.0	19.6	17.2	14.9	13.7
Comb exp sm	4.7	8.2	9.9	13.6	19.4	14.4	10.5	16.4	13.8	12.8	11.7
Comb forec	4.1	8.7	9.1	15.5	25.1	18.0	13.4	16.9	14.2	14.9	13.4

What we Know about Judgmental Forecasts and Judgmental Adjustments

- In a study of eight supply-chain companies making over 300,000 real-life forecasts, it was found that 52% were less accurate than the no-change, Naïve 1 benchmark.
- judgmentally adjusting the statistical forecasts is often necessary, but to improve accuracy it must be done objectively and systematically
  - Concentrating on large changes
  - Avoiding upward adjustments influenced by positivism, wishful thinking and need to achieve desired goals
  - Justifying the reasons for the adjustments, preferably written and anonymously
  - Keep track of the adjustments to evaluate their contribution

The 3003 Series of the M3 Competition and its Major Finding

The classification of the 3003 time series used in the M3-Competition

Time interval between successive observations	Types of time series data						
	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57		756
Monthly	474	334	312	145	111	52	1428
Other	4			29		141	174
Total	828	519	731	308	413	204	3003

A new method Theta, a variation of exponential smoothing, prove to be the most accurate of all methods

Comparison of various methods with Naïve2 as the benchmark

	Forecasting horizon(s)				
	1	Average: 1–4	Average: 1–6	Average: 1–12	Average: 1–18
Theta	2.1%	2.2%	2.1%	2.3%	2.5%
ForecastPro	1.9%	2.0%	1.9%	2.1%	2.3%
ForecastX	1.8%	1.8%	1.8%	2.0%	2.0%
Comb S-H-D	1.6%	1.5%	1.5%	1.8%	2.0%
Dampen	1.7%	1.6%	1.5%	1.8%	1.8%
RBF	0.6%	1.1%	1.3%	1.5%	1.7%
ARARMA	0.8%	0.8%	0.7%	0.7%	0.7%

The Five Major/Consistent Findings of Forecasting Competitions

- Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones.
- The relative ranking of the performance of the various methods varied according to the accuracy measure being used.
- The accuracy of various methods being combined outperforms, on average, the individual methods involved and did well in comparison to alternatives.
- The variance of combinations is lower than the methods
- The accuracy of the various methods depended on the length of the forecasting horizon involved.

The 100,000 M4 Competition Series

Number of M4 series per data frequency and domain.

Time interval between successive observations	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	6,538	3,716	3,903	6,519	1,088	1,236	**23,000**
Quarterly	6,020	4,637	5,315	5,305	1,858	865	**24,000**
Monthly	10,975	10,017	10,016	10,987	5,728	277	**48,000**
Weekly	112	6	41	164	24	12	**359**
Daily	1,476	422	127	1,559	10	633	**4,227**
Hourly	0	0	0	0	0	414	**414**
Total	**25,121**	**18,798**	**19,402**	**24,534**	**8,708**	**3,437**	100,000

Common Benchmarks: Simple and Combined

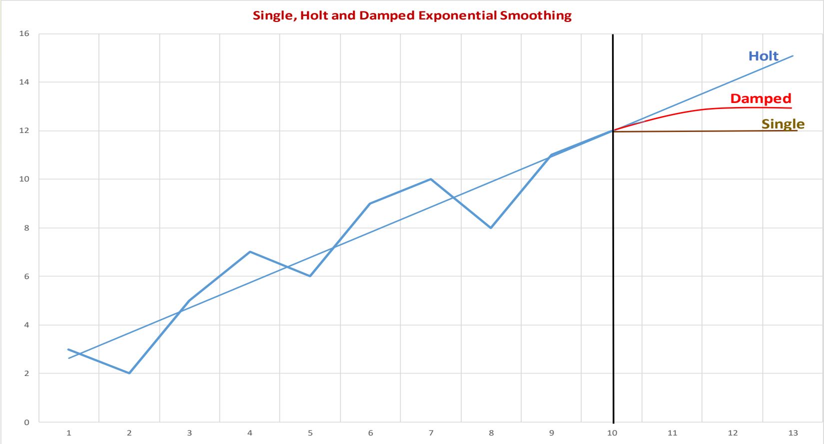
- Naïve 1: $X_{t+i} = X_t$ (The forecasts of tomorrow and the days after will be the same as that of today, also called a Random Walk Model)
- Naïve 2: $X'_{t+i} = X'_t$ (The forecasts of tomorrow and the days after will be the same day last week)
- Comb: The simple numerical average of Single, Holt and Damped Exponential Smoothing:

<u>**The Extrapolation of the Trend**</u>

Single: Horizontal extrapolation

Holt: Linear extrapolation

Damped: Damped extrapolation



The M4 Competition and its 7 Findings

Finding 1: ***The improved numerical accuracy of combining:*** The most important finding of the M4 Competition was that all the top performing methods, both in terms of PFs and PIs, were combinations of mostly statistical ones, with such combinations being numerically more accurate than pure statistical and ML methods. More specifically, there was only **one pure** statistical method among the first ten most accurate ones, as measured by PFs, while **all of the ten least accurate** methods were either pure statistical or pure ML. Similarly, there was only **one pure** statistical method among the first five most accurate ones, as measured by **PIs**, while **all of the five least accurate** methods were pure statistical.

The Accuracy of Combining: M4 Competition

Method	sMAPE	MASE	OWA	Rank	Time
SES	13.087	1.885	0.975	34	8.1
Holt	13.775	1.772	0.971	33	13.3
Damped	12.661	1.683	0.907	22	15.3
Comb	12.555	1.663	0.898	19	33.2
Theta	12.309	1.696	0.897	18	12.7
ETS	12.726	1.680	0.908	23	889
ARIMA	12.669	1.666	0.903	20	3031

The M4 Competition and its 7 Findings

Finding 2: ***The superiority of a hybrid approach utilizing statistical and ML features:*** The biggest surprise of the M4 Competition was a new, innovative method submitted by Slawek Smyl, a data scientist at Uber Technologies, which mixes ES formulas with a Recurrent Neural Network (RNN) forecasting engine. Smyl clarifies that his method does not constitute a simple ensemble of exponential smoothing and neural networks. Instead, the models are truly hybrid algorithms in which all parameters, like the initial ES seasonality and smoothing coefficients, are fitted concurrently with the RNN weights by the same gradient descent method. This method was close to **an impressive 10% more accurate than the Comb.**

The M4 Competition and its 7 Findings

Finding 3: ***The significant differences between the six top performing methods and the rest in terms of PFs.***

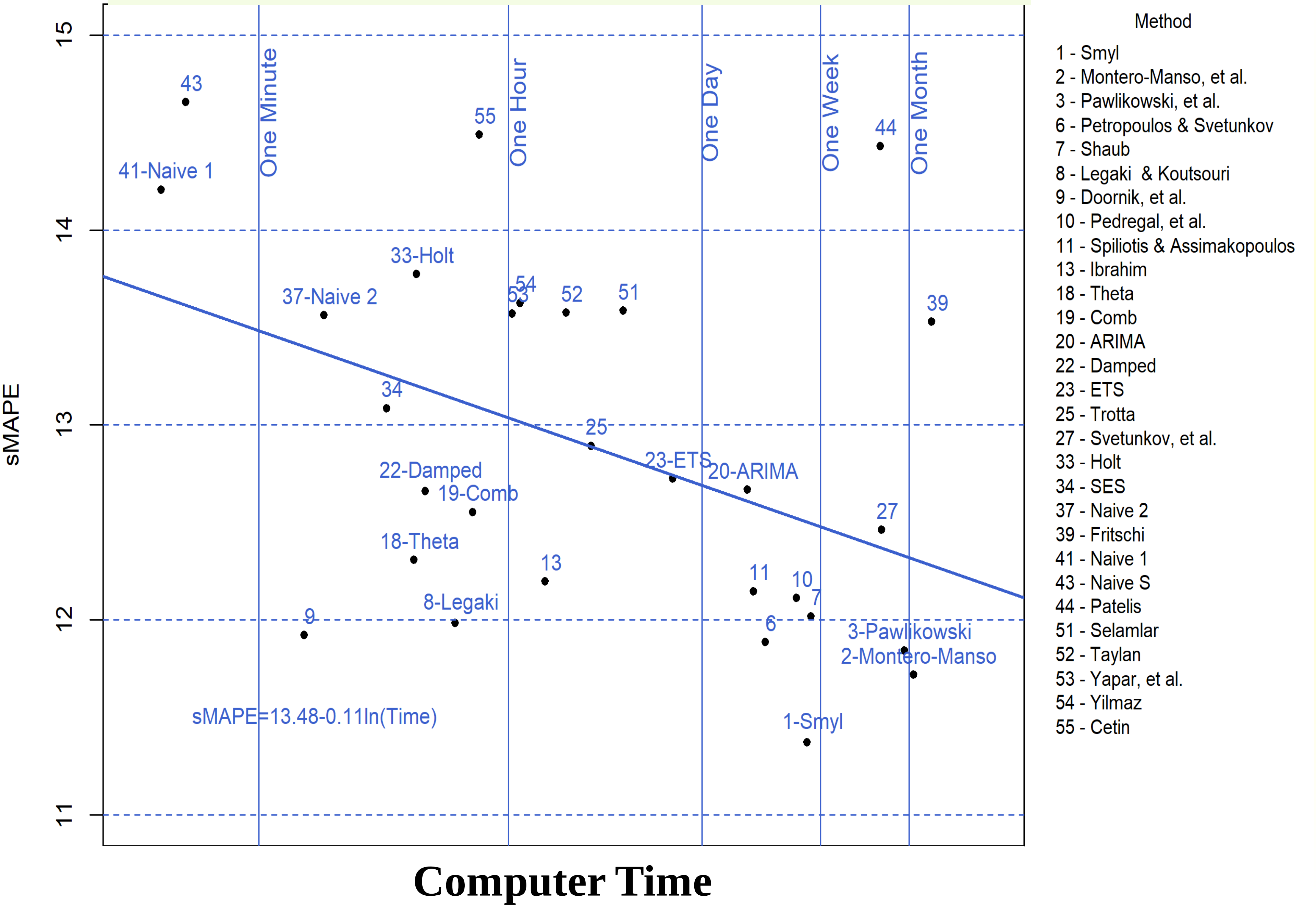
Finding 4: ***The improved precision of the PIs:*** The most accurate and the second most accurate methods of the M4 in terms of PFs achieved **an amazing success in precisely estimating the 95% PIs** as well.

Finding 5: ***More complex methods can possibly lead to greater forecasting accuracy***

Finding 6: ***Using information from multiple series to predict individual ones***

Finding 7: ***The poor performance of the submitted pure ML methods***

Accuracy Vs. Computer Time



The Three Others and the Six Kaggle Competitions

- The Crone et al., 2011 Neural Networks
- Athanasopoulos, 2011 The Tourist Demand
- Hong, Energy Forecasting, 2014, 2016 and 2017

The Kaggle Competition

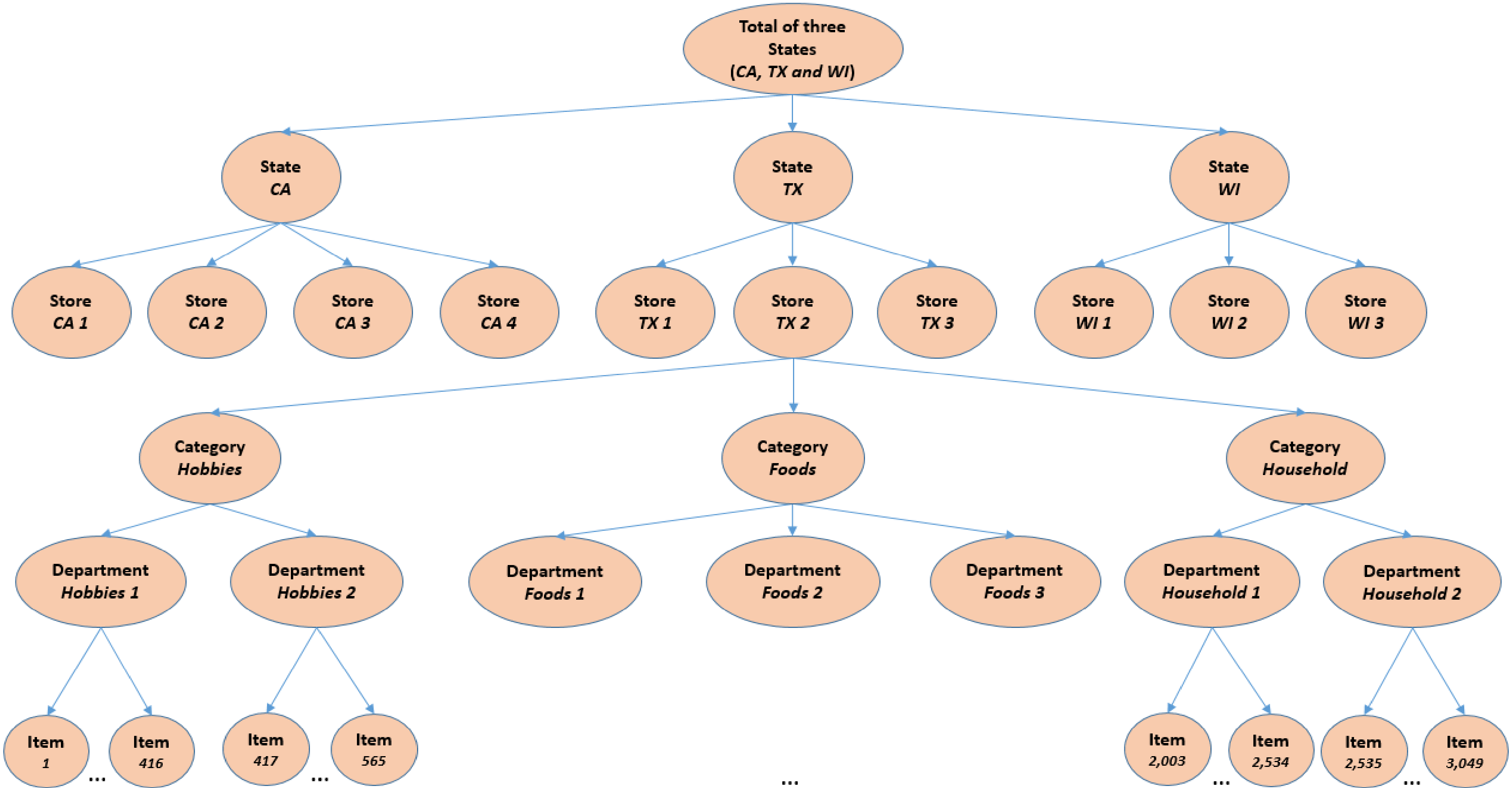
Walmart Store Sales Forecasting	2014
Rossmann Store Sales	2015
Walmart Sales in Stormy Weather	2015
Wikipedia Web Traffic Time Series Forecasting	2017
Corporación Favorita Grocery Sales Forecasting	2018
Recruit Restaurant Visitor Forecasting	2018

Casper Solheim Bojer and Jens Peder Meldgaard, Kaggle forecasting competitions: An overlooked learning opportunity, International Journal of Forecasting, 2020

The M5 Competition (Walmart Data): Forecasting and Uncertainty

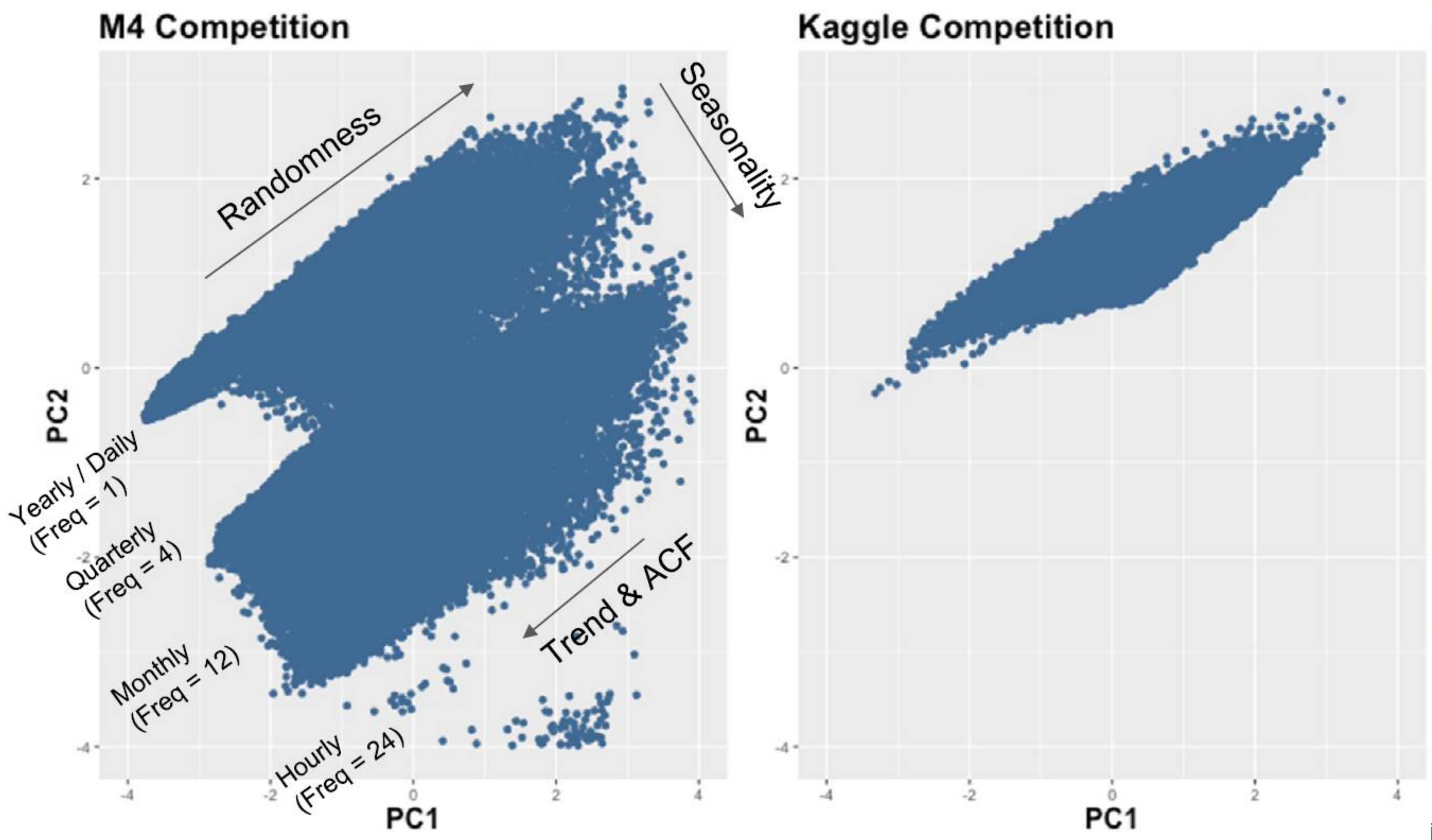
M1, M2, M3 and M4: Covering Practically All Frequencies/Domains

M5: Specific, Hierarchical, Daily Retailing Data, Exhibiting Intermittency

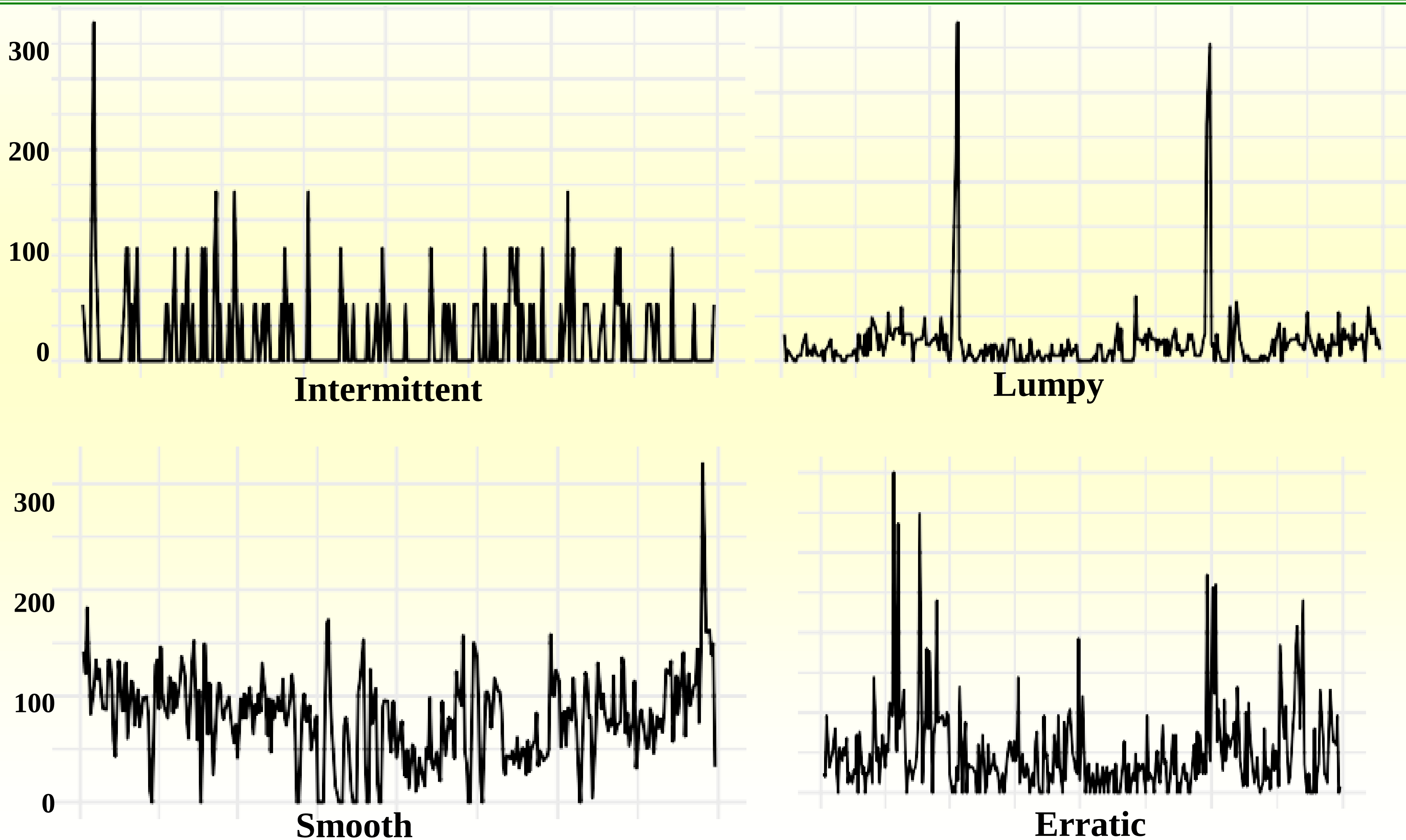


Are Forecasting Competitions Data Representative of the Reality?

C. Fry and M. Brundage / International Journal of Forecasting 36 (2020) 156–160



The M5 Walmart Data: Exhibiting Intermittency (Sporadic demand including zeros)



M5: The Two Challenges (Accuracy & Uncertainty) and its Major Findings

First competition to study uncertainty and analyze its findings in detail

- Finding 1: The superiority of simple ML methods
- Finding 2: The value of combining
- Finding 3: The value of cross learning
- Finding 4: The value or cross validation
- Finding 5: The significant differences between the winning methods and the statistical benchmarks used
- Finding 6: The importance of exogenous/explanatory variables.

A Summary of the M Competitions and their Major Findings

Year	Name	Citations	Competitors	Methods	No. of Series	Major Findings
1979	Pre-M: M-H Study	659	0	22	111	Simple>Sophisticated Combining works
1982	M (or M1)	1,585	7	24	1001	Simple>Sophisticated Combining works
1993	M2	348	5	16	29	Previous, plus human forecasters do NOT improve accuracy
2000	M3	1,693	15	24	3,003	Previous, plus Theta and Forecast Pro improve accuracy. NN/ML less Accurate than statistical methods
Forecasting winter: Between 1979 and 2017 there were minimal improvements in forecasting accuracy while uncertainty was seriously and consistently underestimated						
2018	M4	306	49	61	100,000	Forecasting winter ends Hybrid methods improve accuracy and estimate uncertainty perfectly Combinations still achieve great results Pure ML methods underperform
2020	M5		8,229	?	42,840	Superior performance of ML lightGBM, both challenges The considerable value of CL, both challenges Importance of explanatory/exogenous variab. Value of combining The potential of NN methods in uncertainty
2022	M6					?

Forecasting at Scale: The Architecture of a Modern Retail Forecasting System

TARGET CORPORATION: PHILLIP YELLAND, ZEYNEP ERKIN BAZ, DAVID SERAFINI

PREVIEW *In this first of a three-part article, Phillip Yelland, Zeynep Erkin Baz, and David Serafini, technical leads in the Data Science/AI team at Target, describe their team's efforts to construct a demand forecasting system capable of efficiently generating the nearly one billion weekly forecasts required by the Target Corporation. They highlight the interplay of challenges arising in the contexts of statistical modeling, software engineering, and business practice and explain how the team surmounted obstacles in these three fields of knowledge. Subsequent parts of the article will address the process of implementing the forecasting system and its maintenance in production.*

UNIVERSITY *of* NICOSIA

Thank you