A Vetting of the Acuity of the Bloomberg One-period-ahead Forecasts

Moncef Belhadjali[1], David Leon[2], Edward Lusk[3]

[1]School of Business, Norfolk State University:
Norfolk, Virginia, USA

[2]School of Business and Economics [Supply Chain Management], SUNY: Plattsburgh, Plattsburgh, USA

[3]Emeritus: The Wharton School, [Statistics], The University of Pennsylvania, USA & School of Business and Economics [Accounting], SUNY: Plattsburgh, USA & Chair: International School of Management: Otto-von-Guericke, Magdeburg, Germany

doi.org/10.51505/IJEBMR.2026.10705 URL: https://doi.org/10.51505/IJEBMR.2026.10705

Received: May 16, 2026 Accepted: Jun 23, 2026 Online Published: July 06, 2026

Abstract

Context One of the progeny of the Big Data Analytics [BDA] *circa* 2017 is the recent emergence of a commercial sector offering: SQL: Relational Database [SQL:RDb] Data-Analytics & Artificial Intelligence modular driven platforms that can be used by their clients as "Dedicated Out-source Links". Most of these SQL:RDb firms offer Forecasting Platforms where clients can pass their datasets through 25-30 Forecasting Models [FM]s.

Focus As there are often *Quantity v. Quality* issues with this forecasting-barrage tack, we offer a simple and effective vetting-check. We suggest using a protocol where the Bloomberg One-Period Ahead forecasts are used to create a simple but effective benchmark for deciding on forecasting acuity.

Results For the measure of forecasting acuity, we selected the Absolute Percentage Error [APE] for the One Period Ahead forecasts. For the Vetting-phase, as expected, *usually* the [Random Walk FM] *outperformed* the OLSR-Forecasts. When this was the case, we moved to the Benchmarking-phase, where the Bloomberg forecasts were, in the main, either *inferentially* more effective or equally effective compared to those of the RW FM ***for the* [*APE%*]*-measure***.

Conclusion We suggest that the SQL:RDb firms offer Forecasting Platforms that are also linked to the BBTs-platforms. Using the BBTs as *an initial forecast profiler* may simplify the forecasting process without compromising its acuity.

Keywords: Wilcoxon pairwise method, Non-parametric

1. Introduction

1.1 A Recent Mini-Retrospective As Bob Dylan so apply noted: "*The Times, They are A-Changing*".

Where this is most evident is in the integrated: {**D**ata **A**nalytics & **A**rtificial **I**ntelligence}-domain. *Circa* the 2020s, there was a natural coalescing of: (i) The collection and storage possibilities enabled by "*Super-Computers*", (ii) the "Infinite-capacity" of the "Cloud" for capturing data, (iii) The evolution of the Pivot Tables® Relational Database [RDb] platforms, (iv) into Python™-esque-SQL codes, resulting in: (v) DA- & AI-code-generators capable of creating a plethora of decision-making intel. We label these firms in this novel commercial sector: The SQL:RDb Firms. This, we offer, rather than enabling managers to intelligently process intel in this *Giga-context*, actual creates: *Time-Stress Decision-Panic* & *Information Overload*. To address this ubiquitous conundrum, individual researchers have offered time-tested course corrections—*so as to regain control of the decision-making and planning processes*. For example, in the *Journal of Big Data*, Stockinger, Bundi, Heitz & Breymann (2019) present a "***Must-Read***" discussion on "controlling" the DA- & AI-interface in the SQL:RDb firms where SQL-Modules can probe datasets and spew-out uncountable DA- & AI-*What-If* untested or non-vetted "solutions". These authors present the implementation of a *real-world case* where the Decision Analytics can be performed:

(i) by the *SQL query optimizer of a firm in the SQL:RDb sector,* ***without*** *any* programming required by the user—i.e., this is the standard profile for most all of the SQL:RDb firms, or—*alternately*
(ii) after the managers of the firm edit the *SQL query optimizer* so as to create intel that is likely needed to address the issues under examination.

This is the *crux* of the issue that we are addressing. The above choice: [(i) *or* (ii)] is Binary: Use the SQL query optimizer offered by SQL:RDb firm ***as is***: ***Consequence*** The Managers must sort through the **volumes of output** of the optimizer-module to try to locate the relevant intel for addressing the question of interest—we suggest that this is the: "*Small Needle in Large Haystack Polemic*" *or alternatively*, managers opt to edit the SQL-query optimizer offered by SQL:RDb firm so as to create the intel likely needed to aid in addressing their decision-making issues —we suggest that this is the "*Large Needle in the Small Haystack*". Both can provide useful intel. Occasionally, as management searches the Large Haystack, they become aware of alternative issues that could be addressed. *Or* occasionally, the Small Haystack is too narrowly focused re: the search parameters to the extent that they miss hitting on the best-case option for resolving the issue. Faced with this unavoidable Binary-Choice-set, Stockinger, *et al.* offer: [we added **[*is*]**]:

> *"The design trade*-offs **[is]** *between applying user-defined functions on existing computation kernels vs. partially re-writing the kernel in SQL and thus taking*

> *advantage of the underlying SQL query optimizer. Our performance evaluation demonstrates almost linear scalability for the best design choice."*

In our treatment of this issue, ***when relevant forecasting intel is desired***, our experiential conditioning suggests that the KIS-Protocol is often the best choice. Also see the excellent treatment of the insight-derivation from the operational context as presented by Fahey (2019). For this reason, we offer the following research intel to support vetting-screens to arrive ***expeditiously*** at forecasting intel that will usually inform the Planning Committee.

1.2 Mini-Epilogue As one may expect, the advice of Stockinger, et al. details the next logical and *sorely* needed evolution phase:

> *The SQL:RDb sector should focus* ***on how*** *the* ***quantity*** *of intel possible can* ***be formed*** *into* ***quality*** *decision-making information!*

Each semester, we have been offering three or four *Zoom*™-presentations of these SQL:RDb firms to our faculty where these firms detail their services and encourage the integration of their platforms into ***our*** {AI & DA}-course offerings. For example, here are a sampling of five such SQL:RDb firms*:

Splunk™ <https://www.splunk.com/> Now a *Cisco*™ Company.
CompTIA™ <https://www.comptia.org/>
MongoDB™ https://www.mongodb.com/try/download/community
EC-Council™ https://www.eccouncil.org/
Alteryx™ {<<https://www.alteryx.com/>> *SparkED*

**We have no financial links with any of these firms*.

We have been present at these SQL:RDb workshop presentations. Our interest has been in the planning possibilities afforded by their {DA & AI)-Forecasting Platforms. During these *Zoom* ™-presentations, they usually suggest that their clients: "link-into the ***n***-Forecasting Models on their {DA-AI}-Forecasting-Platform and generate a rich profiling of the ***n***-forecasting profiles! ***Important Intel***: Usually, for these SQL:RDb firms; ***n*** is in Range of [25 to 35]-FMs**!** *This gives us pause, to say the least*; managing such an out-pouring of possible forecasting decision-making-intel, most of which cannot be vetted *re*: the assumptions of these ***n***-FMs, is not consistent with the practice of forecasting for firms and organizations which have "successfully been able to project their ***Past*** in to a relevant and understandable profile of the Tactical or Strategic ***Future*** state of the firm so as to inform the Planning Committees". ***Truly*** *More possible Forecasting options does not necessarily bode-well re: the Quality of the decision-making in planning and guiding the firm into the future.*

During these DA-AI-presentations, we have posed the following question:

> *Have you integrated a link in your forecasting platform to the Bloomberg* TM *Market platform, where routinely, for now more than 20 years, Bloomberg has presented* ***One-Period-Ahead*** *forecasts for Annual and Quarterly Data?*

Paraphrasing, the SQL:RDb-firm presenters indicate:

> *Well, it is the case, that many of our clients are also clients of Bloomberg. However, we have not contracted with Bloomberg to integrate the BBT-Market platform* ***into*** *our Forecasting Platform. Should our clients use our service, there are no restrictions on the use of any other commercial DA or AI platforms.*

The above information, collected over a few years, is the motivation for this research report. Following, we will offer ***vetting-*** and ***benchmarking-***inferential analyses to form an informed opinion on the forecasting acuity of the Bloomberg forecasts. This is needed in the current SQL:RDb-sector so as to enrich the creation of ***quality*** forecasting-intel to inform the planning committee. ***Simply*** If the BBT-forecasts are at least not inferior to those of the more Stat- & Math-FMs that are in the SQL:RDb {DA-AI}-FM platforms—*such as The Class of Neural-Net FMs*—then, asserting the Time to Forecast imperative—"***We need the forecast-intel yesterday!***" the BBT-forecasts can be a very valuable addition to the panoply of forecasters. In addition, such a vetting of the acuity of the BBT forecast is needed as there does not seem to be an analysis of the BBT-forecasts. *This is based on an ABI/INFORM[Global] *Key-Word* [AND [&] OR] Search: [*Bloomberg* & *Firm* & *Forecast*] of 25 Jan [2026].

2. Inferential Tests of the Forecasting Acuity of the BBT-Forecasts

2.1Overview It is, indeed, an anomaly that there has not been an inferential analysis of the acuity of the Bloomberg One-Period-Ahead forecasts reported in the literature*. Perhaps, this anomaly is due to the fact that the BBT-forecasting modelling process is proprietary and so not disclosed in the public domain. Further, ***be advised***—attempts at reverse-engineering of the BBT-Estimates to discover Bloomberg's forecasting protocol is "an inferential fools-errand!"

2.2 The Inference Context: Vetting- & Benchmarking-Screening The purpose of this research report is to create reliable inferential intel so as to inform the Planning Committee as they ponder ***our*** *musing*:

> *How does the forecasting acuity of the BBT* ***One-Period-Ahead*** *forecasts for annual data of firms traded on active exchanges compare to that of other FMs*?

To create such information, we offer the following inferential-testing-protocol. *In each section*, we will elaborate on the issues that are important in creating the inferential intel that is to be used in addressing our *musing*. Following are the essential four elements [A, B, C & D] of our inferential protocol:

[A] ***Power*** *Calculation for the Firm Random Samples and The* ***Skewness*** *Guidelines for Inference Modeling*

[A_1] The **Power** App: Wang & Chow (2007)

In this case, we will be reverse "engineering" the usual desired Power for the Means using the Wang & Chow **(2007)** Power Profiler: App [WC:App] for the Two-Sample pairwise inferential comparison case. The forecasting measures of interest will be APE[percentages]. Specifically, for a False Positive Error of: [α = 5%]—i.e., a 95% level of confidence, for a desired Power [1-The False Negative Error] of 80%—i.e., a z[β] of 84.2%, for the detection Mean[%]s of: Sample A[30%], and Sample B[5%]: -i.e., a *generalized* precision of 15%. Thus, given these parameters, the sample size needed ***per*** Group [A & B] will be ***29***. For this sample-size, we will use an False Positive Error [FPE] [$H_o[Null$[p-value]]]: Toggle-point of ***≈0.10*** for rejecting the H_o.

Caveat Power, in the Mean[%]-domain, is conditioned on the Nature of the Population *as reflected* in the accrued samples. The Power calculation is useful ***iff*** the assumed population has skewness = 0. *If this is not the case, then the inferential tests recommended are Non-Parametric* **[*NP*]**-*Tests*. Very often, in our experience, analysts assume that the sample-size calculated using the WC:App is a tacit-guarantee that the FPE[%]s are reliable decision-making intel for ***any*** of the Mean[Parametric]-inference profiles. If skewness is an issue, ***this is not the case***. Also, interestingly, the Central Limit Theorem [CLT] is no guarantee that skewness will not be an issue to consider in electing the nature of the analysis: Parametric [Means-oriented] or the robust alternative: [Non-Parametric-Tests—usually Rank-Sum Measures].

[A_2] Selecting the Inference Model: **Skewness** Guidelines

As a pre-inferential analysis screening-check, we will test the data-collected for skewness. We have used and *highly recommend* the Pearson Skewness[Median[ν]]: [P_ν^{Sk}]-test and its *exact* standard error to create an inference context for deciding *if* Parametric ***or*** Non-Parametric inferential methods are preferred; however, this is the Sampled-Version[p-value] rather than the Population-version. The computation for the [P_ν^{Sk}]-test is illustrated following:

The Pearson [P_ν^{Sk}] inferential test formula is:

[P_ν^{Sk}] = [3×(Mean[%] – Median[%])] / ***s***] / {[(6×n) × (n-1)] / [(n-2) × (n+1) × (n+3)]}^0.5

where: ***s*** is the standard deviation of the sample.

Computational Illustration Assume we have the following [%]-values

Mean = 77.0%, Median = 8.7%, ***s = 13.5199****, and n = 196. then*

N[*z*] = [3×[77.0% – 8.7%] / ***13.5199***] / [The Standard Error{Sampled-Version]]

Where this standard error is: {[6×196×195] / [194×197×199]}^0.5

Thus, the N[***z***] = [0.151554] / [0.173644] = **0.873**

The [P_ν^{Sk}] *directional* FPE[$H_o[Null$[p-value]]] is: N[**z**] = **0.873** = ***0.191***

This indicates that the chance of taking a random sample of 196 values from a *symmetric population* where: {The Mean **=** The Median} and {**s** is: ***13.5199***}and finding the Mean = 77.0% and the Median = 8.7% *could happen, due to random chance*, but

only **19.1%** of the time. For most analysts, this would suggest that the FPE[$H_o[Null$[p-value]]] of: 19.1%, is a non-trivial indication that the assumption that sampled population was symmetric may logically be ***rejected*** in favor of: H_a that: There is likely ***non-trivial*** Pearson Skewness.
Note We elected and recommend NOT to eliminate any of the outliers causing skewness or use any trimming-screens [including Winsorizing] to mollify skewness as they change the nature of the data.

Skewness A ***Guideline*** *for selecting an Inferential Model* For each inferential analysis, we will test for the level of Pearson ***Sk****ewness:* **[**P_V^{Sk}**]** as this will determine the preferred analytic-platform: *Parametric* [***Mean***[Absolute Percentage Error[APE]%]s] ***or*** *Robust*[*Non-Parametric* ***Mean***[*Rank Sum*[APE%]s]]. Specifically, for a trivial-level of **[**P_V^{Sk}**]**, we will use: The Tukey-Kramer HSD[LS***Mean***[APE[%]s] for creating the directional *pairwise inferential profiles. Otherwise*, we will use the Wilcoxon Method [based upon the Kruskal-Wallis ***Mean***[Rank Sums[APE[%]s]. In the presence of *material* Skewness, the Means, *per se*, are rarely used for inference. Interestingly, experientially, the Wilcoxon Method that uses the Ranks of the Data *usually* tests to have [P_V^{Sk}]s *that fails to reject the* FPE[$H_o[Null$[p-value]]] *even* ***when*** *there is non-trivial skewness for the APE*[*%*]*s*!

[P_V^{Sk}**] Results** For our sample of the S&P$_{500}$, for each of the seven Y-Variates there are 28-Firms for which there is a Mean & Median for which a [P_V^{Sk}]: FPE[$H_o[Null$[p-value]]] *may be computed*. Additionally, as there are three FMs there are in total 21 [P_V^{Sk}]: Y-Variate values. Following, we computed the Mean and the 95%***CI***[Mean] for these 21 **[**P_V^{Sk}]:p-values:

The [P_V^{Sk}]-Profile Y- Variate FPE[$H_o[Null$[p-value]]], n=21 was:
$\hat{x}_{p-values}$= 0.0394 and the 95%CI [$\hat{x}_{p-values}$] was: **[0.021: 0.058]**

Indication We judge this **[**P_V^{Sk}]-profile to indicate that there is clear evidence for the 21 Y-Variates of the S&P$_{500}$ that there is ***non-trivial skewness*** in the [APE%]s over the FMs. To highlight this indication, we have shaded the Non-Parametric inferential section of the testing profile Tables. ***Action*** We will use the Wilcoxon Method [based upon the ***Mean***[Rank Sums[APE[%]s] to create the pairwise inference profiles to address our questions of interest.

[B] ***Key Y-Variates*** *for a Random Sample of Firms Actively Traded on a* ***Major Stock Exchange*** *& Y-Variates* ***Panel Sample-size Issues***

[B_1] *Definition of a* ***Key-Variable***

The Key Y-Variates selected were those that we judged to be ***Exo***genously *as well as* ***Endo***genously Sensitive AIS[*Y*-Annual Variates]. As a guideline in selecting the Y-

Variates, we used the Fraser & Ormiston (2015) Text. We selected the following seven Key Y-Variates for each firm accrued from the BBTs[Income Statement[GAAP]]:

Table 1 ***Exo***genously as well as ***Endo***genously Sensitive AIS [Annual Key *Y*-Variates] Variables used in the inferential analysis

AIS-AICPA Notation	BBT[GAAP] Notation
Earnings for the Common Shares	EARN-_FOR_COMMON
EBITDA	EBITDA
EBITDA at the Margin	EBITDA_MARGIN
Gross Profit	GROSS_PROFIT
Income Statement: Operating	IS_OPER_INC
Pre-Tax Income [GAAP]	PRETAX_INC
Sales: Revenue	SALES_REV_TURN

Discussion We define Key Y-Variates as any BBT-Income Statement [GAAP] Accounting Information System [AIS]-variable that is likely to be used by most investors to decide which firms in various [SIC/NAICs]-groups are profitable investment options. Usually, these Key Y-Variates are: ***Exo***genously as well as ***Endo***genously Sensitive AIS[*Y*-Variates]. We define ***these*** as:

I. An ***Exo***genously AIS[*Y*-Variate] is modulated, *more or less in real-time*, by the state of the economic milieu of the firm, *and*
II. An ***Endo***genously AIS[*Y*-Variate]] is *more or less* controllable by decisions that are made by managers of the firm.

[B_2] *Selection of a* ***Major Trading Exchange***

The S&P$_{500}$ We collected a random sample of 28-firms actively traded on the S&P$_{500}$ through January of 2025; and the same firm-sample for the Year-ending in January of 2026. The firm-sample of 2025 listed the BBT-Annual **Estimate** for 2025. The sample collected in 2026 reported the **Actual Value** of the Y-Variate for 2025. Thus, the dataset for our inferential analysis will be:

2025:Accual[$_{\mathbf{n=10}}$] [Firm[Y-Variates]: Panel [2015:2024 & The BBT-***Estimate*** for 2025]]
2026:Accual [The same Firm[Y-Variates] & The BBT-***Actual Value*** for 2025].

Vetting Note We verified that there were no SEC-restatements for the Panel-values reported for the accruals of 2025 & 2026. The BBT-Tickers are noted in Appendix A.

[B_3] ***Longitudinal Panel Sample-Size*** *for Time Series FMs*

The sample-size forecasting literature recommends selecting [10 : 25] Panel-Y-Variate-values for time series FMs for ***each*** OLSR-[Input X-Range]-variable to obtain meaningful 95%Predication Intervals [95%PIs] for each of the Extrapolations of the

Vol. 10, No.07; 2026

ISSN: 2456-7760

OLSR FM. See: Jenkins & Quintana-Ascencio, (2020). We selected *ten* Panel-Values as we have only *one* Key Y-Variate for Annual Panel Values; a sample of 15 years traversed a likely serious Non-Ergodic-Event section—The *Lehman Bros*™ sub-prime debacle. For this reason, we selected a longitudinal Panel-size of ***ten***.

[C] Indicate the Measures of Forecasting Acuity and the **Time Series Forecasting Models**

[C_1] *Measures of Time Series **Forecasting Acuity***

As the Measure of Forecasting Acuity for profiling the inferential tests, we will use the following non-directional *version* of the Absolute Percentage Error [APE]:

APE[%] = ABS[Forecast – Actual] / Actual Eq[1]

Where the standard directional APE is:

[$APE^{\%}$] = [[Forecast – Actual] / Actual]

In this case, it is possible to flag instances where the Actual is < 0, as our APE will also be < 0. We recommend our version, as it creates more refined de-construction possibilities.

[C_2] *The Three Forecasting Models* [***FMs***[

We have selected ***two time-series FMs and a data-accrual protocol*** that we proffer will enable the creation of forecasts that will offer reasonable ***Vetting- & Benchmarking-*** expectations for evaluating the Bloomberg Forecasting Estimates. We are accruing annual data for firms on the S&P$_{500}$. As noted above, we have selected an event-space that was not atypical for the global-markets. In this timeframe, the trajectory of most all of the Global Indices exhibited a Pearson Product Moment Correlation [PPMC] that is almost, without exception, ***stable*** and > **0**. Consider now the two FMs:

The OLSR Forecasts We are using the Upper Limit of the **OLS** Linear [Two Parameter [Intercept & Slope]] **R**egression [95%**P**rediction **I**nterval] [95%**PI**] as the One-Period-Ahead ***forecast*** of the Y-Variate: ***Noted*** as: **OLSR**[*FM*$_{[\mathbf{UL}95\%\mathrm{PI}]}$]. We expect for ***vetting-purposes*** that this OLRS-Forecast will be ***an extreme value*** in comparison to most of the One-Period-Ahead forecasts made from the Panel of ten-longitudinal values for that Y-Variate.

The Modified RW Forecast The standard RW model of Makridakis *et al* (1982) and Collopy & Armstrong (1992) uses as the RW-forecast-i.e., the last Y-Variate that is used to create the OLSR-intel, noted as: [Y_n^{OLSR}]. However, it is the case, that such a RW FM[: [Y_n^{OLSR}]] projection is very likely to be an under-estimation of the next observed actual value of the Y-Variate in a market trading context, where the usual PPMC-trajectory is increasing over time. In this case, we formed the following FM:

$$\widehat{RW}_{FM} \equiv [Y_n^{OLSR} + [k \times \Delta]] \qquad R[1]$$

where: k *is the number of periods ahead* for the forecast, and Δ is the incremental constant, and Y_n^{OLSR} is the standard RW FM. We recommend the following computation protocol for Δ:

Δ = [[Y-Variate [$t_{m=[n-1]}$]] – [Y-Variate [t_{n-5}]]] / 4] where : *n* is the last Y-Variate for the Panel that is used to fit the OLSR.

In this configuration, the slope is non-restricted. For example, assume that we have the following dataset:

Table 2 Profile for the computation of Δ

Time[t]	t_{---}	t_5	t_6	t_7	t_8	t_9	t_{10}
Panel[A]	t_{---}	12	14	19	17	22	RW_A
Panel[B]	t_{---}	22	17	19	14	12	RW_B

The Δ for Panel[A] is: Δ_A = [[22 – 12] /4] = +2.5; thus,
the ***One***-Period-Ahead forecast will be: [RW_A + 2.5].
The Δ for Panel[B] is: Δ_B = [[12 – 22] /4] = –2.5; thus,
the ***One***-Period-Ahead forecast will be: [RW_B – 2.5].
Note This R[1]-configuration will fit any dataset for any of the directional projections. ***Simply*** The Directional Change effected by the Δ will be in sync with the PPMC of the Y-Variate Panel parameterizes the Δ.

[*D*]Vetting- & Benchmark-Testing the Relative Acuity of the **BBT-Estimates**

[D_1] *Vetting the* ***Detection Power Change*** *for an Aggregation of the Y-Variates*: *Overall Screening for Non-Ergodic Panel Segments*

Previously, we have noted that **29** Y-Variates are needed given the specifications for the Pre-Power Calculation of: [A_1]. This vetting test will use the Aggregation of all the Y-Variate data *for each* of the three FMs. The Vetting Expectation is:

The FPEs of the three Pairwise Wilcoxon indications should all be << ≈**0.10,** where: ≈ ***0.10*** was our conjecture of a reasonable Toggle-point for the FPE[5%] & FNE[80%] conditions that we used for the WC:App.

Rationale Recall, for the Aggregate Analysis, the sample-size for each of the three FMs is; [Firms × Y-Variates] or 196 [28 × 7]. As this is a ratio of: 7 [196/28], this *sevenfold* increase in the testing sample-size should ***dramatically*****[<<]** increase the detection-power relative to that of ***≈0.10***. ***If not***, this would likely suggest that the some of the Y-Variate accruals are basically Non-Ergodic relative to the remaining Y-Variates. The results are presented in Table 3.

The Profiles of the Aggregate APE[*%*]*s* For this Vetting Phase, we need to examine the Aggregate Profiles of the C_2^3[FMs]: {$APEs[^{BBT\ v.}_{Actual}]$ *v.* $APEs[^{OLSR\ v.}_{Actual}]$ *v.* [$APEs[^{\widetilde{RW}\ v.}_{Actual}]$]}

Table 3 The Aggregate Profiles: C_2^3[FMs]:{$APE_{BBT}^{Y[i]}$ v. $APE_{OLSR}^{Y[i]}$ v. $APE_{\widehat{RM}}^{Y[i]}$}

FMs, **[n=196 each]**	**APE[Means]**	**APE[Medians]**
BBT-Estimate	77.0%	**8.7%**
OLSR[*FM*[UL95%PI]]	233.1%	28.6%
$\widehat{RW}_{FM}$	93.3%	11.4%
Overall p-values	**O**[p-value]	**O**[p-value]
Welch[ANOVA]*	0.332	N/A
Median Test*	N/A	<0.0001
Pairwise p-values	T-K HSD*	Wilcoxon*+
BBT *v.* OLSR	0.212	**<0.0001**
BBT *v.* $\widehat{RW}_{FM}$	0.983	**<0.001**
OLSR *v.* $\widehat{RW}_{FM}$	0.287	**<0.0001**

The FM-acuity measure is APE[%] *These Inferential tests are provided in Appendix B and +Indicates that all the Wilcoxon p-values are directional.

Discussion The *overall vetting* results are clear. The sevenfold increase in sample-size has resulted in a dramatic increase from the power-detection toggle-point of ***≈0.10***. See the Shaded p-values in Col[3]. Recall, we are using the Wilcoxon-Method for the directional-FPE-Tests as there was *suggestive* evidence of skewness for the Means and Medians of Table 3. This is an inferentially important result for supporting the stable nature of the 28 Firms that are presented in Appendix B for the seven Y-Variates. **Simply** The veracity of the Benchmarking of the BBT-Estimates likely is not going to be affected by Non-Ergodic segments for the Y-Variates over the 28 randomly selected S&P500 firms.

[D_2] ***Vetting the Nature of the*** $\widehat{RW}_{FM}$ To vet the nature of these [$\widehat{RW}_{FM}$] S&P$_{500}$ Panels, we elected to use the following Vetting-Screen:

For the random sample of S&P$_{500}$ firms, the chance that the $\widehat{RW}_{FM}$ for a particular Y-Variate would have a ***Median*[APE%]*** that *would be > than that of the* OLSR[*FM*$_{[UL95\%PI]}$][***Median[APE%***] for that particular Y-Variate *is very unlikely; for a particular Y-Variate this chance is: ≈5%*[*i.e., 1-≈95%*]. *If there is* ***inferential*** *evidence that the:* [$\widehat{RW}_{FM}$] ***is outperformed by*** *the* OLSR[*FM*$_{[UL95\%PI]}$] over the collection of the 28-S&P$_{500}$ firms for their respective APE%, *then that Y-Variate panel will be* ***flagged*** *as a* ***Non-Ergodic*** panel—i.e., it does not share the profile expected for the S&P$_{500}$ panels and so it will be removed from the study. If this this is not the case, then the $\mathbf{RW_{FM}}$ for that Y-Variate will be used in the following Benchmarking-test. **We used the Medians as they are more aligned with the pairwise Rank-Sums that are used* *by the Wilcoxon Method. However, to be clear the Wilcoxon Test does not use actual Median-values—the Medians are just a convenient anchor for understanding the nature of the pairwise FPE-tests.*

[D_3] ***Benchmarking the Nature of: {The $\widehat{RW}_{FM}$ v. The BBT-Estimate}***

Our **benchmarking** expectation is: *For each of the seven Y-Variates*, the APE% of the vetting qualified $\widehat{RW}_{FM}$ will ***inferentially*** test so that the {FPE[APEs[$\widehat{RW}^{v.}_{Actual}$] *v.* $APEs[^{BBT\ v.}_{Actual}$]]} is such that FPE[$H_o[Null$[p-value]]] ***is not rejected*** suggesting that the APEs[$\widehat{RW}^{v.}_{Actual}$] = $APEs[^{BBT\ v.}_{Actual}$]. This inferential information will be collected from the ***Wilcoxon-Method pairwise JMP-Platform***.

3. Computational Illustration

3.1 Overview We have selected from the Bloomberg Terminals, the S&P500 firm: [*Microsoft*™ Corp (MSFT US)-Y-Variate[Income Statement GAAP[Sales: Revenue]]USDs]. ***Alert*** This presentation only is concerned with illustrating the ***computations*** for one firm: [MSFT] and one Y-Variate:[Sales: Revenue] The computations follow:

Table 4 [*Microsoft*™ Corp (MSFT US)-Y-Variate [Income Statement GAAP [Sales: Revenue]]USDs]

Computational Dataset				
93,580	91,154	96,571	110,360	**125,843**
143,015	168,088	198,270	**211,915**	245,122

Additional Intel The BBT [MSFT[***Estimate/Forecast***]] at time t_{11}] = 277,343.39, and the ***Actual Value*** for MSFT[Y-Variate[Sales: Revenue: [t_{11}]] was: 281, 724.

3.2 The OLSR [*MSFT:*[Sales: Revenue]]*Test Elements* The fitted OLSR: $\boldsymbol{Y^{OLSR}_{t=11}}$ = [51,456.6 + [17,624.58 × 11] = ***245,327.00***

The 95%**P**rediction **I**nterval [95%**PI**] for the [MSFT[$\boldsymbol{Y^{OLSR}_{t=11}}$**]** dataset is:

245,327.00 ± $t_{[5\%,8]}$ × $\sqrt{\textbf{MSE}}$× [1 + 1/10 +[(Mean[Time Index) – 11]^2]/SS_{xx}]^0.5

where: The MSE is the **M**ean **S**quared **E**rror of the Regression-Line *v.* the Actual Panel Values.

Using the values in Table 4, we have for the 95%PI:

245,327.00 ± 2.306 × 14101.15 × [1+ 1/10 +[(5.5 – 11]^2]/82.5]^0.5, or

95%PI ≡[245,327.00 ± 39,380.35]

Thus, the Upper 95&**PI** is: [245,327.00 + 39,380.35] = ***284,707.35, thus*** *the* OLSR[$FM_{[\mathbf{UL}95\%\mathrm{PI}]}$] will be: ***284,707.35***

3.3 The OLSR-forecast

In the context of this research report, we will use this 95%PI of the OLSR as fitted to the random sample drawn from the S&P$_{500}$. The reason for this is that the BBT-Estimate is a One-Period-Ahead forecast. In the context of the trading markets, if the Y-Variates are multiplicative where the PPMC is >0, then the $\widehat{RW}_{FM}$ very likely will be: < than the Upper 95&**PI** of ***284,707.35.*** *This is a vetting test, the computations of which are presented following.*

3.3.1The $\widehat{\mathrm{RW}}_{\mathrm{FM}}$: [MSFT: Sales Revenue] Test Elements

Illustration of the Vetting Test of the *proffered* relationship: OLSR[*FM*$_{[\mathbf{UL}95\%\mathrm{PI}]}$] > $\widehat{RW}_{FM}$.
Using the MSFT [Y-Variate [Sales: Revenue]]: Table 4, we find:
The RW = 245,122 and the Δ is:
Δ = [[Y-Variate [211,915]] – [Y-Variate [125,843]]] / 4]
Δ = 21,518.00
The [$\widehat{RW}_{FM}$[MSFT]] forecast ***for k=1*** is:
245,122 + 21,518 = **266,640**

The above vetting computations for The MSFT [Sales: Revenue] of Table 4 is:
[$\widehat{RW}_{FM}$: **[266,640]] and**
[Upper Limit of the 95%PI[FM]: **[*284,707.35*]]]}**
In this vetting case, the [$\widehat{RW}_{FM}$:**[[266,640]]**is < the **[**Upper Limit of the 95%PI[FM] **[*284,707.35*]** as expected. This is a vetting indication for ONE firm—i.e., MSFT!

3.3.2 Finally, Calculation of the APE[MSFT] for the FMs

The [APE%] for the BBT [*Actual Value* for MSFT[Y-Variate[Sales: Revenue: $[t_{11}]$]] was: 281, 724.
Thus, the **MSFT[APE%]s** are:
$APE[\%]_{BBT}^{Y[MSFT]}$ ≡ [[277,343.39 – 281,724] / 281,724] = 1.6%
$APE[\%]_{OLSR}^{Y[MSFT]}$ ≡ [[284,707.35 – 281,724] / 281,724] = 1.1%
$APE[\%]_{\widehat{RM}}^{Y[MSFT]}$ ≡ [[266,640 – 281,724] / 281,724] = 5.4%

Alert This is just ONE firm to be used in the inferential test. For the inferential test, profiled in *Section 2.2*[A], we will use the entire sample of 28 Firms to evaluate the Y-Variate [Sales: Revenue]. We will ONLY be eliminating the Y-[Sales: Revenue] ***if***, over ALL of the 28-firms in the sample, the $\widehat{RW}_{FM}$[Median[APE%[Sales: Revenue]]] were to be ***inferentially*** > than that of the Median[APE%] of OLSR[FM$_{[\mathbf{UL}95\%\mathrm{PI}]}$][Median[APE%]].

4. The Analysis of the S&P$_{500}$[Y-Variates] Random Accruals

4.1 Overview

The next section of this research report will detail the testing of the Pairwise p-values of the relative forecasting acuity of the BBT-Estimate *vis-à-vis* that of: The $\widehat{RW}_{FM}$ for each of the seven y-Variates of Table 1. This will be the inferential intel used to address the following question of interest for this research report:

How does the forecasting acuity of BBTs measured, as: The $\{\mathrm{APEs}[^{\mathrm{BBT\ v.}}_{\mathrm{Actual}}]$, compare to that of the $\widehat{\mathrm{RW}}_{\mathrm{FM}}$ measured as: The $\mathrm{APEs}[^{\mathrm{RW\ v.}}_{\mathrm{Actual}}]$ BBT for the One-Period-Ahead forecasts?

4.2 Summary of the Vetting Testing

Results As context for probing the various aspects of the question of interest, we have summarized the results of our vetting tests; the intention of these vetting screens is to indicate the care that we have taken to ***form datasets that are germane*** to addressing the investigation of our question of interest. ***Simply*** Vetting is our way of dealing with the *GIGO*-Issue. We have summarized our vetting-screening in the Table 5 following:

Table 5 Restatement of the Vetting Tests

Vetting Screen	**Nature of Screen**	**Indication & *Result***
Pearson $[P_v^{Sk}]$ See: A_2	If there is prevalent non-trivial Skewness, then use Non-Parametric Tests; otherwise, use Parametric Tests for the Pairwise Profiling.	The prevalence of **non-trivial Skewness was clear**; thus, *we used Non-Parametric Rank Tests for evaluating the pairwise profiles for all the FMs.*
Aggregate p-values See: Table 3	Test for Non-Ergodic Y-Variates among The S&P500 for the three FMs by screening the nature of the change in detection power.	**All the p-values** for the Aggregate Sample Size[196] Table 3 for the FMs indicated dramatic increases in detection-precision. *Panels likely Ergodic.*
APE_{OLSR}^{Yi} [>] APE_{RM}^{Yi} See: Tables 6:12	Test to determine if there were to be Y-Variates for which the APE[%] of the OLSR FM inferentially outperformed the APE of the $\widehat{RW}$.	There were **no instances** where a Y-Variate for which the APE[%] of the OLSR FM inferentially outperformed that of the $\widehat{RW}$. *Summary in* Tables 6,- - -,12.

4.3 The Inferential Profiles for The Individual Y-Variates

The next de-construction question is: For the seven S&P$_{500}$ Key-Y-Variates, *individually* how do the BBT-Estimates fare relative those of the $\widehat{RW}_{FM}$[*APE%*] & *OLSR*[*FM*]APE%] for the Non-Parametric Tests [Rank Sum[APE%]]s where the testing p-value will be based upon the Wilcoxon- Method? **Clarification** Why offer intel re: The overall ***Median Test? The actual p-***

values reported by the Wilcoxon Method are for the Rank Sum Mean tests for the three APEs of the three FMs. We have found that decision-makers do not have ***an anchor*** that indicates the nature of the p-values for the pairwise p-values reported under the Wilcoxon Method. Experientially, we find that the Rank Sums align *in relationship* with the Medians. Thus, we recommend testing the three Medians using the standard Median test. The decision-makers can thus use the Medians and the related overall Median tests as decision-making ***surrogates*** for the p-values reported by the Wilcoxon Method.

4.3.1 Earn¬_For common

Table 6 Earnings for the Common Shareholders

EARN-_FOR_COMMON	**APE[Means]**	**APE[Medians]**
BBT-Estimate	46.2%	12.8%
OLSR[*FM*$_{[UL95\%PI]}$]	217.4%	27.6%
$\widehat{RW}_{FM}$	113.9%	28.0%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.235	N/A
Median Test	N/A	0.160
Pairwise p-Values	T-K HSD	Wilcoxon^{*+}
BBT v, OLSR	0.266	0.034
BBT v. RW	0.810	0.111
OLSR v. RW	0.612	0.199

*A discussion of these Inferential tests is provided in Appendix B. $^{+}$Indicates that all the Wilcoxon p-values are directional. This will be the case for all seven Y-Variates.

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 6:

The Vetting Test *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {OLSR[*FM*$_{[UL95\%PI]}$]*APEs*]}, *the FPE*[H_o[$Null$]] *was not rejected, as the p-value was 0.199.* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***28.0%*** and that of the {OLSR[*FM*$_{[UL95\%PI]}$]*APEs*]} was ***27.6%***; it is not unreasonable to interpret this relative intel to indicate that the $\widehat{RW}_{FM}$ was not outperformed by the{OLSR[*FM*$_{[UL95\%PI]}$]. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[*FM*$_{[UL95\%PI]}$]*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}$[*APEs*]; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test *indicates that for the two FMs*: The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {BBT-Estimate}, *the FPE*[H_o[$Null$]] *was not rejected, as the p-value was 0.111. .* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***28.0%*** and that of the {BBT-Estimate} was ***12.8%***; thus, it is not unreasonable to interpret this relative intel to indicate that the {BBT-Estimate} did not exhibit a forecasting profile that was different in the population context. Thus, the

Vol. 10, No.07; 2026

ISSN: 2456-7760

Benchmarking-Test suggests ***that it is likely*** that there is a not a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {$\widehat{RW}_{FM}$ **[≅]** The BBT[FM]} where: **[≅]** indicates that the performance of these FMs is likely not different re: their forecasting acuity.

4.3.2 EBITDA

Table 7 Earnings Before Interest Taxes Depreciation & Accruals

EBITDA	APE[Means]	APE[Medians]
BBT-Estimate	30.0%	8.5%
OLSR[*FM*$_{[\mathbf{UL95\%PI}]}$]	101.9%	35.7%
$\widehat{RW}_{FM}$	34.6%	12.5.%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.232	N/A
Median Test	N/A	0.052
Pairwise p-Values	T-K HSD	Wilcoxon^{*+}
BBT v, OLSR	0.115	0.002
BBT v. RW	0.991	0.024
OLSR v. RW	0.149	0.067

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 7:

The Vetting Test *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {OLSR[*FM*$_{[\mathrm{UL95\%PI}]}$]*APEs*]}, *the FPE*[$H_o[Null]$]] *was rejected, as the p-value was 0.067*. ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***12.5%*** and that of the {OLSR[*FM*$_{[\mathrm{UL95\%PI}]}$]*APEs*]} was ***37.5%***; it is not unreasonable to interpret this relative intel to indicate that the $\widehat{RW}_{FM}$ outperformed the{OLSR[*FM*$_{[\mathrm{UL95\%PI}]}$]. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[*FM*$_{[\mathrm{UL95\%PI}]}$]*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}$[*APEs*]; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test *indicates that for the two FMs*: The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {BBT-Estimate}, *the FPE*[$H_o[Null]$]] *was rejected, as the p-value was 0.024*. . ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***12.5%*** and that of the {BBT-Estimate} was ***8.5%***; thus, it is not unreasonable to interpret this relative intel to indicate that the {BBT-Estimate} outperformed the $\widehat{RW}_{FM}$.Thus, the Benchmarking-Test suggests ***that it is likely*** that there is a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {The BBT[FM] **[>]** $\widehat{RW}_{FM}$} where: **[>]** indicates **outperformed** re: their forecasting acuity.

4.3.3 EBITDA_margin

Table 8 Earnings Before Interest Taxes Depreciation & Accruals at the Margin

EBITDA_MARGIN	**APE[Means]**	**APE[Medians]**
BBT-Estimate	32.7%	8.3%
OLSR[*FM*[UL95%PI]]	91.2%	31.5%
$\widehat{RW}_{FM}$	25.3%	7.7%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.175	N/A
Median Test	N/A	0.022
Pairwise p-Values	T-K HSD	Wilcoxon*+
BBT v, OLSR	0.184	0.0002
BBT v. RW	0.972	0.259
OLSR v. RW	0.118	0.001

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 8:

The Vetting Test *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}[APEs]$} *and The* {OLSR[$FM_{[UL95\%PI]}$]*APEs*]}, *the FPE*[$H_o[Null]$] *was rejected, as the p-value was 0.001.* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}[APEs]$} was ***7.7%*** and that of the {OLSR[$FM_{[UL95\%PI]}$]*APEs*]} was ***31.5%***; it is not unreasonable to interpret this relative intel to indicate that the $\widehat{RW}_{FM}$ outperformed the{OLSR[$FM_{[UL95\%PI]}$]. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[$FM_{[UL95\%PI]}$]*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}[APEs]$; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test *indicates that for the two FMs*: The {$\widehat{RW}_{FM}[APEs]$} *and The* {BBT-Estimate}, *the FPE*[$H_o[Null]$] *was not rejected, as the p-value was 0.259. .* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}[APEs]$} was ***7.7%*** and that of the {BBT-Estimate} was ***8.3%***; thus, it is not unreasonable to interpret this relative intel to indicate that the {BBT-Estimate} did not exhibit a forecasting profile that was different in the population context. Thus, the Benchmarking-Test suggests ***that it is likely*** that there is a not a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {$\widehat{RW}_{FM}$ **[≅]** The BBT[FM]} where: **[≅]** indicates that the performance of these FMs is likely not different re: their forecasting acuity.

4.3.4 Gross profit

Table 9 Gross Profit

GROSS_PROFIT	**APE[Means]**	**APE[Medians]**
BBT-Estimate	11.0%	6.6%
OLSR[*FM*[UL95%PI]]	41.6%	25.7%
$\widehat{RW}_{FM}$	17.8%	6.1%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.019	N/A
Median Test	N/A	0.020
Pairwise p-Values	T-K HSD	Wilcoxon*+
BBT v. OLSR	0.006	0.001
BBT v. RW	0.765	0.099
OLSR v. RW	0.042	0.007

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 9:

The Vetting Test *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {OLSR[*FM*[UL95%PI]]*APEs*]}, *the FPE*[H_o[$Null$]] *was rejected, as the p-value was 0.007.* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***6.1%*** and that of the {OLSR[*FM*[UL95%PI]]*APEs*]} was ***25.7%***; it is not unreasonable to interpret this relative intel to indicate that the $\widehat{RW}_{FM}$ outperformed the{OLSR[*FM*[UL95%PI]]. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[*FM*[UL95%PI]]*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}$[*APEs*]; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test *indicates that for the two FMs*: The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {BBT-Estimate}, *the FPE*[H_o[$Null$]] *was rejected, as the p-value was 0.099.* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***6.1%*** and that of the {BBT-Estimate} was ***6.6%***; thus, it is not unreasonable to interpret this relative intel to indicate that the$\widehat{RW}_{FM}$ outperformed the {BBT-Estimate}.Thus, the Benchmarking-Test indicates ***that it is suggestive*** that there is a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {$\widehat{RW}_{FM}$}**[>]** The BBT[FM]} where: **[>]** indicates **outperformed** re: their forecasting acuity.

International Journal of Economics, Business and Management Research

Vol. 10, No.07; 2026

ISSN: 2456-7760

4.3.5 Is_oper_INC

Table 10 Income Statement Operating Income

IS_OPER_INC	**APE[Means]**	**APE[Medians]**
BBT-Estimate	332.7%	19.3%
OLSR[*FM*$_{[UL95\%PI]}$]	692.3%	39.3%
$\widehat{RW}_{FM}$	120.3%	14.5%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.459	N/A
Median Test	N/A	0.160
Pairwise p-Values	T-K HSD	Wilcoxon*+
BBT v. OLSR	0.754	0.055
BBT v. RW	0.906	0.445
OLSR v. RW	0.491	0.117

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 10:

The Vetting Test **tentatively** *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {OLSR[*FM*$_{[UL95\%PI]}$]*APEs*]}, *the FPE*[$H_o[Null]$] *was* ***not*** *rejected, as the p-value was 0.117*. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[*FM*$_{[UL95\%PI]}$]*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}$[*APEs*]; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test **tentatively** *indicates that for the two FMs*: The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {BBT-Estimate}, *the FPE*[$H_o[Null]$] *was* ***not*** *rejected, as the p-value was 0.445*. Thus, as the Benchmarking-Test suggests ***that it is not likely*** that there is a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {$\widehat{RW}_{FM}$ **[≅]** The BBT[FM]} where: **[≅]** indicates that the performance of these FMs are likely not different re: their forecasting acuity.

4.3.6 Pretax INC

Table 11 Pre-Tax Income

PRETAX_INC	**APE[Means]**	**APE[Medians]**
BBT-Estimate	82.7%	22.1%
OLSR[*FM*$_{[UL95\%PI]}$]	459.5%	41.8%
$\widehat{RW}_{FM}$	326.7%	21.5%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.474	N/A
Median Test	N/A	0.160
Pairwise p-Values	T-K HSD	Wilcoxon^{*+}
BBT v. OLSR	0.606	0.017
BBT v. RW	0.810	0.350
OLSR v. RW	0.939	0.061

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 11:

The Vetting Test *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}[APEs]$} *and The* {OLSR[*FM*$_{[UL95\%PI]}$*APEs*]}, *the FPE*[$H_o[Null]$] *was rejected, as the p-value was 0.061*. ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}[APEs]$} was ***21.5%*** and that of the {OLSR[*FM*$_{[UL95\%PI]}$*APEs*]} was ***41.8%***; it is not unreasonable to interpret this relative intel to indicate that the $\widehat{RW}_{FM}$ outperformed the{OLSR[*FM*$_{[UL95\%PI]}$]. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[*FM*$_{[UL95\%PI]}$*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}[APEs]$; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test **tentatively** *indicates that for the two FMs*: The {$\widehat{RW}_{FM}[APEs]$} *and The* {BBT-Estimate}, *the FPE*[$H_o[Null]$] *was* ***not*** *rejected, as the p-value was 0.350*. Thus, as the Benchmarking-Test suggests ***that it is not likely*** that there is a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {$\widehat{RW}_{FM}$ **[≅]** The BBT[FM]} where: **[≅]** indicates that the performance of these FMs are likely not different re: their forecasting acuity.

4.3.7 Sale_rev_turn

Table 12 Sales Revenue

SALE_REV_TURN	**APE[Means]**	**APE[Medians]**
BBT-Estimate	4.0%	1.74%
OLSR[*FM*[UL95%PI]]	27.9%	19.5%
$\widehat{RW}_{FM}$	14.6%	5.7%
Overall p-values	p-values	p-values
Welch[ANOVA]	0.0001	N/A
Median Test	N/A	0.0001
Pairwise p-Values	T-K HSD	Wilcoxon*+
BBT v, OLSR	0.0002	<0.0001
BBT v. RW	0.165	0.0001
OLSR v. RW	0.059	0.009

Codex & Discussion In this dis-aggregated context, ***for the One-Period-Ahead Forecasts***, we offer the following intel from Table 12:

The Vetting Test *indicates that for the two FMs*: {The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {OLSR[*FM*[UL95%PI]]*APEs*]}, *the FPE*[$H_o[Null]$] *was rejected, as the p-value was 0.009.* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***5.7%*** and that of the {OLSR[*FM*[UL95%PI]]*APEs*]} was ***19.5%***; it is not unreasonable to interpret this relative intel to indicate that the $\widehat{RW}_{FM}$ outperformed the{OLSR[*FM*[UL95%PI]]. Thus, the Vetting-Test ***is confirmed***, i.e.,—The OLSR[*FM*[UL95%PI]]*APEs*]} ***did not outperform*** The $\widehat{RW}_{FM}$[*APEs*]; therefore, we move to the Benchmarking analysis, and

The Benchmarking Test *indicates that for the two FMs*: The {$\widehat{RW}_{FM}$[*APEs*]} *and The* {BBT-Estimate}, *the FPE*[$H_o[Null]$] *was rejected, as the p-value was 0.0001. .* ***Elaboration*** *The Median of the* {$\widehat{RW}_{FM}$[*APEs*]} was ***5.7%*** and that of the {BBT-Estimate} was ***1.74%***; thus, it is not unreasonable to interpret this relative intel to indicate that the {BBT-Estimate} outperformed the {$\widehat{RW}_{FM}$}.Thus, the Benchmarking-Test suggests ***that it is likely*** that there is a difference between the forecasting acuity of the: $\widehat{RW}_{FM}$ & The BBT[FM].

Indication {The BBT[FM]} **[>]** $\widehat{RW}_{FM}$where: **[>]** indicates **outperformed** re: their forecasting acuity.

4.4 Summary of the Y-Variate Vetting Tests for the BBT[APE%]s v. Those of the $\widehat{RW}_{FM}$ *[APE%]s*

Table13 Directional Wilcoxon Method p-values

Y-Variates [WM-Test]	BBT[APE%]	$\widehat{RW}_{FM}$ [APE%]	W[p-value]	Indication
EARN_COMMON Y1	12.8%	28.0%	0.111	BBT≅$\widehat{RW}$
EBITDA: Y2	8.50%	12.50%	0.024	BBT>$\widehat{RW}$
EBITDA_MARGIN: Y3	8.30%	7.70%	0.259	BBT≅$\widehat{RW}$
GROSS_PROFIT: Y4	6.60%	6.10%	0.099	RW>BBT
IS_OPER_INC Y5	19.3%	14.5%	0.445	BBT≅$\widehat{RW}$
PRETAX_INC Y6	22.1%	21.5%	0.350	BBT≅$\widehat{RW}$
SALES_REV_TURN: Y7	1.74%	5.70%	0.0001	BBT>$\widehat{RW}$
Mean Average	***11.33%***	***13.71***	N/A	N/A

Note: > indicates Outperforms and ≅ indicates that inferentially there was not likely a performance difference.

Discussion The de-construction of the Individual Y-Variate results as presented in the seven Tables and summarized in Table 13. ***Simply*** The BBTs created forecasts that were, *in the main* [*excepting Gross Profit*], either inferentially more effective or equally effective with respect to those of: The $\widehat{RW}_{FM}$.

5. Practical Take-Aways

5.1 Overview

Assume that the members of the Forecasting Division intend to share their capsule-profile and recommendations of Table 13 with the Planning Committee of an ***Investment Bank***. Following, we are offering ***our experiential expectations*** of this communication:

From: *The Director of Forecasting*
Subject: *Analysis of the Forecasting Acuity of the Bloomberg Market Platform*
To: Members of the Planning Committee

We have analyzed, in the usual manner, the Bloomberg Forecasting Estimations that are routinely offered at year-end for most all of the firms that we evaluate for developing our performance projections for the coming year. After comprehensive inferential analyses, we suggest the following:

The One-Period-Ahead Forecasts offered by Bloomberg ***are reliable indications*** of what is likely to be the performance profile *for the coming year* for the market trading securities, *the ilk of which*, are our target-firms for our clients.

Vol. 10, No.07; 2026

ISSN: 2456-7760

Of course, should you wish a detailed *PowerPoint*™ presentation of our inferential analyses, we are at your disposal.

Needless to say:
Certainly, we will be tracking the viability of this recommendation each year to assure that this recommendation remains relevant to our needs. As a related *vetting* indication, we offer the following.

5.2 An Additional Vetting Indication of a Rating Service [***Most of this intel is taken directly from their URL and so is not edited by The Forecasting Division.***]

There is a commercial organization called *Ad Fontes Media*™ [<https://adfontesmedian.com/>] that employs panels of analysts that regularly review representative sample content to rate it for **Reliability** and **Bias**. Each panel of analysts comprises one left-leaning, one right-leaning, and one center-leaning analyst. The team considers a variety of factors when rating content.

To determine its **Reliability** score, we consider the content's veracity, expression, its title/headline, and graphics. We add each of these scores to the chart on a weighted scale, with the average of those creating the sample content's overall reliability score.

To determine sample content's **Bias** score, we consider its language, its political position, and how it compares to other reporting or analysis from other sources on the same topic. We add each of these scores to the chart on a weighted scale, with the average of those creating the content's overall bias score.

The **Bias** rating, - - -, ranges from most extreme left to middle to most extreme right. The reliability rating, demonstrated on the chart's vertical axis, rates sources on a scale from original fact reporting to analysis, opinion, propaganda and inaccurate/fabricated information.

Reliability scores for articles and shows are on a scale of 0-64. Scores above 40 are generally good; scores below 24 are generally problematic. Scores between 24-40 indicate a range of possibilities, with some sources falling there because they are heavy in opinion and analysis, and some because they have a high variation in reliability between articles.

Bias scores for articles and shows are on a scale of -42 to +42, with higher negative scores being more left, higher positive scores being more right, and scores closer to zero being minimally biased, equally balanced, or exhibiting a centrist bias.

The following are the overall Bias and Reliability scores for **Bloomberg News** according to our Ad Fontes Media ratings methodology.

Reliability: 42.29 & **Bias: -2.93**

This information offered by *Ad Fontes Media* is in sync with the results of our study as reported above.

Appendices

Appendix A Random Accrual of the S&P$_{500}$ firms[alpha-listed] n=28

S&P$_{500}$ Tickers Randomly Selected						
APH	APPL	AVGO	ECL	EMR	FTNT	GM
GOOGLE	HAS	HD	JCI	JNJ	LLY	MCHP
META	MHK	MPC	MSFT	MSI	NWSA	PG
PSX	ROP	SEE	TSLA	UNH	VFC	WHR

Appendix B Inferential Systemic Effects of Parametric & Non-Parametric Analyses *re*: Skewness

Table B1 Three Groups {A & B & C}and their related APE[Percentages]

Datasets to Illustrate the Skewness Effect										
A	0.44	0.33	0.26	0.43	0.27	0.32	0.45	0.27	0.35	0.39
B	0.76	0.74	0.66	0.78	0.27	0.31	0.23	0.33	0.12	0.23
C	0.55	0.4	0.59	0.33	0.36	0.42	0.25	0.41	0.41	0.45

Discussion

When there is serious skewness in a sample, Parametric [***Mean***-based inference models] are usually not the best inferential models for developing useful intel. Usually, the Non-Parametric-based inference models [NpM[are preferred. A frequently used NpM is the Wilcoxon/Kruskal-Wallis Rank Sum Test, where: the original data is transformed into Ranks and these Ranks are used to create the inference intel. This is important for our study as ***non-trivial skewness abounds*** and we are interested in creating relevant pairwise inference profiles to address our study questions. Following are various profiles that will illustrate the above inferential issues. Table B2 offers the Pearson [P_{Sk}]-skewness profiles of Table B1:

Table B2 **P**earson[Mean – Median] ***Sk***ewness [P_{Sk}] using the Correct Standard Error

Group Table B1	**Mean**	**Median**	**[P_{Sk}] Skewness**	**p-value**
A	0.351	0.34	0.45	0.256
B	**0.443**	**0.32**	**1.42**	**0.019**
C	0.417	0.41	0.21	0.379

Discussion

The clear indication is that there is a dramatic difference in the relative position of the [Means *v.* the Medians] of Panels [A & C]] *vis-a-vis* those of [Panel B]. Due to outliers, Panels A & C have a ***slight*** Right-Hand-Side Mean repositioning relative to their Medians; whereas, this repositioning for Panel B is relatively ***dramatic***. This is profiled using the p-values of the Pearson Skewness test presented in Table B2. This repositing-effect causes a difference in the Welch parametric [ANOVA-profile] relative to that of the NpM: The Wilcoxon Kruskal-Wallis test that uses the relative Ranks of data from Table B1. The Welch test produces an overall *FPE*[$H_o[Null]$] p-value of **0.214**; this usually suggests that there are one or two pairwise p-values that offer suggestive indications of analytical interest. However, for the Wilcoxon Kruskal-Wallis [***WKW***] NpM-test, the overall *FPE*[$H_o[Null]$] p-value is: **0.539**; this usually suggests that there ***not likely*** to be pairwise differences of analytic interest. This difference in the p-values profiles: $FPE^{Welch}_{Table\,B1}$ = 0.214 *v.* $FPE^{WKW}_{Table\,B1}$ = 0.539 is ***an artifact of relative skewness profiles***.

The Simple De-Brief for the Pearson [P_{Sk}] Skewness Inference Protocol is:

I. *Check for Skewness. If the p-value re: the FPE*[*Null of* [P_{Sk}]]*=0 can be rejected at a p-value <0.15, then use* The Wilcoxon Kruskal-Wallis test; otherwise use *the Welch*[*ANOVA*] *test, also*

II. *We recommend, when there is non-trivial skewness, to use the Median-values to judge the nature of the middle of the population from which the random sample was drawn.*

References

Collopy, F., & Armstrong, J.S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, *38*, 1394–1414. https://doi.org/10.1287/mnsc.38.10.1394

Fahey, L. (2019). Getting to insight: The value and use of small data. *J. Strategy & Leadership*, *47*, 27-33.

Fraser, L.M., & Ormiston, A. (2015). *Understanding Financial Statements*. Pearson 11th Ed. ISBN-13 978- 0133874037.

Jenkins, D.G., & Quintana-Ascencio, P.F. (2020). A solution to minimum sample size for regressions. *PLSO One*, *15*. <<https://doi.org/10.1371/journal.pone.0229345>>

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R. Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111–153. https://doi.org/10.1002/for.3980010202

Stockinger, K., Bundi, N., Heitz, J. & Breymann, W. (2019). Scalable architecture for Big Data financial analytics: User-defined functions vs. SQL. *J. Big Data*, *6*, 1-25 https://doi.org/10.1186/s40537-019-0209-0

Wang, H., & Chow, S.-C. (2007). Sample size calculation for comparing proportions. Test for equality. *Wiley Encyclopedia of Clinical Trials*, *3*, 3-4. <https://doi.org/10.1002/9780471462422.eoct005>

ISSN: 2456-7760

Acknowlegemnts

Thanks and appreciation are due to: Mr. John Conners, Senior Vice President, Financial Counseling, West Coast Region, AYCO for his generous philanthropy which funded the establishment of the John and Diana Conners Finance Trading Lab at the State University of New York College at Plattsburgh and the Bloomberg Terminals that are offered therein and were instrumental in this research, Prof. Dr. H. Wright, *Boston University*: Department of Mathematics and Statistics, our colleagues who responded to the SBE Research Workshop *Send-Around* at SUNY: Plattsburgh, NY USA, in particular: Prof. Dr. Petrova, for their careful reading, helpful comments, and suggestions.