Available online at www.elixirpublishers.com (Elixir International Journal)

Computational Engineering

Elixir Comp. Engg. 162 (2022) 55940-55950

ISSN: 2229-712X

Investigation of the Smoothing Effect for ANN-Outlier Replacement Protocols: Vetting and Extensions

Frank Heilig[1] and Edward J. Lusk[2]

[1]Strategic Risk-Management, Volkswagen Leasing GmbH, Braunschweig, Germany.

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

ARTICLE INFO

Article history:

Received: 4 December 2021;

Received in revised form:

28 December 2021;

Accepted: 8 January 2022;

Keywords

Vetting,

Smoothing,

Provoking,

Transformations,

Precision.

ABSTRACT

In previous research reports the Excel™ outlier replacement protocol called: The **A**verage of the **N**earest **N**eighbor Panel-points: [ANN] was investigated. The authors reported that there seems to be a predilection relative to chance for ANN-protocols to produce a reduction in the standard error of the OLS-two-parameter linear regression [OLSR] model compared to that produced by the Basic or unmodified Panel. This is usually termed a Smoothing-Effect and results in more narrow Confidence or Capture Intervals [CI]—i.e., enhanced precision. There were idiosyncratic anecdotal conjectures offered as to why such a penchant may be created by ANN-protocols. We will consider ***Dysfunctional or Gaming Considerations***: If Smoothing is inherent for ANN-protocols this offers an opportunity to make the decision to apply or eschew the application of the ANN-protocol based upon the intention to engineer the forecasting CIs. This being the case, two research questions are begged: ***Is there a Panel-length that: (i) sufficiently mollifies the Smoothing- or Provoking-events,*** or ***(ii) results in a balance between Smoothing- and Provoking-events either of which would render the gaming decision moot.*** We offer inferential tests re: (i) the conjecture that the length of the Panel systematically mollifies the ANN-impact on precision, and (ii) the conjecture that the seriousness of an ANN-impact on the OLSR-CIs is symmetrically balanced. We demonstrate inferentially that: Using the Medians of various ANN-protocols tested over various sample-sizes that mollification is likely the state of nature. However, ***despite mollification*** there seems likely to be asymmetry in favor of Smoothing. This suggests that gaming must be entertained as an opportunistic possibility. Given this, an organizational solution is suggested to mitigate against gaming the application of the ANN-protocols.

Introduction

1.1 *Context* In recent research reports Heilig & Lusk (2020) & Heilig and Lusk (2021): [H&L] investigated various aspects of the effect of the ANN-outlier Panel-point replacement protocol with respect to its effect on the 95%Capture Intervals of the *Excel*™ OLS-two parameter [Intercept:$\boldsymbol{\beta_0}$ & Slope: $\boldsymbol{\beta_1}$] linear forecasting equation [OLSR]:*Excel*:[*Data*[*DataAnalysis*[*Regression*]]]—referred to as: OLSR. The ANN-protocol is an acronym for the **A**verage of the **N**earest **N**eighbor Panel Points. Specifically, assume the Panel of data under examination is:

Panel: $\{\boldsymbol{x_{t=1}}, - - -, (\boldsymbol{x+\epsilon})_{\boldsymbol{t=k}}, - - -, \boldsymbol{x_{t=n}}\}$

Where: $\boldsymbol{\epsilon}$ represents an error, usually assume to be additive, of such a magnitude or by its nature that $(\boldsymbol{x+\epsilon})_{\boldsymbol{t=k}}$ is identified as an outlier using either experiential judgment or a standard outlier screening protocol.

In this case, the ANN-Panel value-replacement is: $(\boldsymbol{x+\epsilon})_{\boldsymbol{t=k}} \leftarrow ([(\boldsymbol{x_{t=k-1}}) + (\boldsymbol{x_{t=k+1}})]/\mathbf{2})$; thus, the Panel-point considered to be an outlier—i.e., not a representative Panel time-series value—is the average of its nearest neighbor values in the Panel. The ANN-protocol is the most basic and most often used outlier replacement-protocol when time series under evaluation are audited GAAP-accounting information reported in the financials of market-traded organizations. The rational for this is that in the usual case such GAAP-panels are characterized by generating processes that often produce associated Panel-point values. Thus, as the near neighbor Panel-Points are usually collectively "tied in association" the ANN-protocol is, under this assumption, the maximum likelihood and logical protocol[i]. In fact, the ANN-protocol is one of the default-options offered in the *Excel*™ functionality: ***FORECAST.ETS.CONFINT***. The default as detailed by Excel are: [*supports up to 30% missing data in the timeline and will automatically adjust for it based on Data completion. The default value of 1 will account for missing points by completing them* ***to be the average of the neighboring points****, 0 will indicate the algorithm to account for missing points as zeros.*]

In this context, H&L used the following relationship:

$$\boldsymbol{Relative_{ANN}}: \left[\frac{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{ANN}}}{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{BASIC}}}\right] \qquad \text{R1}$$

Where: $[\boldsymbol{PECI}]^{95\%}_{[\bullet]}$ represents the Precision of the **Excel 95%OLSR Capture Interval**[ii] for the Panel [•] under analysis. The Basic Panel is the Panel as downloaded. Finally, the standard definition of precision is: 50% of the width of the [1-FPE[$\alpha_{/2}$]]Capture Interval or [[ECI:Upper Limit – ECI:Lower Limit] /2].

as the indication or measure of one of the three states of nature: ***Smoothing***, ***Neutral*** or ***Provoking***. In addition, it is the case that:

Tele:

E-mail address: luskej@plattsburgh.edu

$$\boldsymbol{Relative}_{ANN}:\left[\frac{\sqrt{MSE_{ANN}}}{\sqrt{MSE_{Basic}}}\right]\equiv\boldsymbol{Relative}_{ANN}:\left[\frac{[PECI]^{95\%}_{ANN}}{[PECI]^{95\%}_{BASIC}}\right] \quad \text{R2}$$

and thus, R1 & R2 may be used interchangeably[iii]; to simplify the codex, the $\boldsymbol{Relative}_{ANN}$ will be noted as: $\boldsymbol{R}_{ANN}$

If $\boldsymbol{R}_{ANN}$ is <1.0, the effect of the ANN-modification is labeled as: ***Smoothing*** as the OLSR-$\sqrt{MSE}$-variation created by the ANN-modification decreased relative to the OLSR-$\sqrt{MSE}$-variation of the BASIC unmodified series. Thus, Smoothing produces an Excel Capture Interval [ECI] that is smaller in comparison to the ECI of the Basic Panel. ***Provoking*** is the designation for $\boldsymbol{R}_{ANN}$-ratios > 1.0; in this case, the ECI produced by the ANN-modification is wider than the ECI of the Basic Panel, and ***No Effect*** is the label if $\boldsymbol{R}_{ANN}$ = 1.0; the ANN-modification produces an ECI identical to that of the Basic Panel. Epistemologically, one can offer an elaboration; one may context Smoothing as: Shrinking the ratio of: The [1-FPE[$\alpha_{/2}$]]Excel Capture Interval of ANN-Protocol modification relative to The [1-FPE[$\alpha_{/2}$]]Excel Capture Interval of the Basic Panel: Simply: ***Shrinking the Relative Ratio of the ECIs***[***RRECI***]; thus, we will context Provoking as [***Expanding the RRECI***]

Details of the Research

H&L report a likely Smoothing tendency that is produced by the ANN-replacements tested. This Smoothing tendency was heretofore not reported in the peer-reviewed literature. For this reason, the H&L-Smoothing-studies beg additional testing-arms and inferential elaboration. The focus of this inquiry is:

1. Offer a possible ***Dysfunctional*** aspect to using ANN-outlier replacements,
2. Investigate H&Ls' conjecture that longer Time Series mitigate against the ANN-modification-effects re: the 95%ECIs this is termed ***Mollification***,
3. Introduce a test for Symmetry or ***Balance*** re: Smoothing- & Provoking-effects as a mediating aspect relative to gaming the selection ANN-protocols,
4. Discuss the ***Accrual of the Firms*** and their ***Sensitive Account Panel Variables*** used in inferential testing of the Mollification and Balance conjectures,
5. Offer various ***Vetting Tests***, the intention of which, is to engender understanding and confidence in the validity of the meaning of inferences drawn from the tests to be reported,
6. Present detailed ***Tabular Profiles*** of the results of testing the ***Mollification*** and ***Balance*** conjectures, and
7. Conclude with a ***Summary*** and an ***Extension*** of this study

The Context for Outliers and Focus of the Research Report

Overview

There is a "Dark-Side" to Outlier Replacement Protocols If indeed there are outliers then it is antithetical that they should not be replaced. However, in the context where there is likely to be correlation or autocorrelation, the H&L-studies have documented a weak-inferential[iv] Smoothing-effect tendency. In the data-visualization context, at this point, it would be informative to illustrate a possible disfunction if there is a Smoothing- or Provoking-tendency in using the ANN-protocol. In Appendix A there are three illustrative datasets: PCAOB Panel I, PCAOB Panel II and ANN-Panel. These are profiled in Figure A

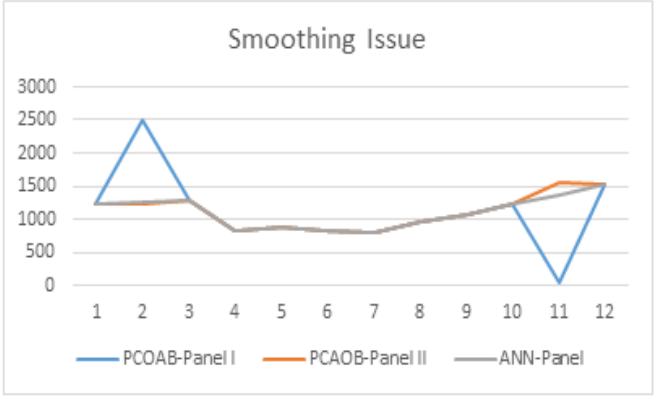

Figure A Panels: Dysfunction Illustration

Dysfunction

Assume the PCAOB-Panel I is the Panel under consideration. Using the Tukey (1977)-Whiskers:Box Plot outlier screen[SAS™[JMP™v.13]SAS(2005), Panel Point[2: $\boldsymbol{t_2}$] & Panel Point[11: $\boldsymbol{t_{11}}$] are flagged as Outliers. This being the case, they should be ANN-replaced. The effect of this ANN[Both[$\boldsymbol{t_2}$&$\boldsymbol{t_2}$]] is:

$$\boldsymbol{R}_{ANN}:\left[\frac{\sqrt{MSE_{ANN}}}{\sqrt{MSE_{Basic}}}\right]\equiv\boldsymbol{R}_{ANN}:\left[\frac{244.5}{553.4}\right]=0.442$$

The ANN[Both[$\boldsymbol{t_2}$&$\boldsymbol{t_{11}}$]-protocol had a dramatic directional and magnitude ratio-effect on the relative Precision of the 95%ExcelCapture intervals as $\boldsymbol{R}_{ANN}$ < 1 or a Smoothing-effect resulted in ***Shrinking the RRECI***. Referencing R2, the implication is that the resulting Precision of the 95%ECI of the ANN[Both[$\boldsymbol{t_2}$&$\boldsymbol{t_{11}}$]-protocol that created the ANN-Panel is: 927.67; whereas, for Precision of the 95%ECI the PCAOB Panel-1 was 2,099.43 or the relative precision ratio is 0.442 [927.67/2,099.43]. Thus, recognizing the likelihood of the outliers at Panel-Points [$\boldsymbol{t_2}$ & $\boldsymbol{t_{11}}$] and applying the ANN-protocol resulted in ***Shrinking the RRECI***. ***It is always the case that smaller CIs are more desirable in comparison to larger CIs.***

Summary: There were outliers identified by a ubiquitous and validated outlier-screening protocol, they were thus replaced by a logical ANN-protocol, and resulted, as expected, in reduction of MSE of the OLSR that produced a smaller 95%ECI. ***This is, of course, the rational of any outlier screening protocol***.

However, here is another, more opportunistic, scenario. Assume that the Panel under consideration is PCAOB-Panel II. In this case, there were no Tukey-outliers. However, assume that the analyst for a reason not motivated by the identification of outliers would like to reduce the size of the 95%ECI. Why would this be an issue? It was just stated that smaller ECIs are more desirable than larger ECIs. ***Yes, true IF there are legitimate outliers; and only if this is the case***. ***Illustrative Actual Case***. A colleague related the following: The organization had a complex-new project evaluation screening protocol. The initial stage to move the project forward for consideration for funding was that The Project was in a forecasting interval for future ROI-projections. What internal-audit of the firm discovered was that different CIs were being used by individuals depending on their desire to move the project forward. Those managers strongly in favor of The Project used Excel-Capture CIs that are relatively larger; those who wanted to block the project selected Random-Effects CIs that are relatively smaller! The same gaming-effect "could" be achieved by selecting an ANN-protocol when it, in fact, is not warranted. For an illustration,

assume that The Project had an actual ROI value of: $299.45 and the evaluation-screen was the 95%ECI of the PCAOB Panel-II: [$296.05 to $2,263.64]. ***Evaluation***: The project would have moved forward as $299.45 is IN: [$296.05 to $2,263.64]. However, if someone wanted to block the project, they could apply the ANN[Both[t_2&t_{11}]-protocol. True, it is not warranted as there are NO Tukey-outliers in PCAOB Panel-II. However, the goal is to block the Project. IF the "outliers" were replaced this would have created a Smoothing-Effect as follows:

$$\boldsymbol{R}_{ANN}\colon \left[\frac{\sqrt{MSE_{ANN}}}{\sqrt{MSE_{Basic}}}\right] \equiv \boldsymbol{R}_{ANN}\colon \left[\frac{244.5}{259.3}\right] = 0.943$$

The re-modulation from PCAOB Panel-II that has a 95%CI of [296.05 to 2,263.64] to ANN-Panel that creates the 95%CI of: [301.78 to 2,157.11] would have resulted in ***Shrinking the RRECI*** thus blocking The Project as 299.45 is not in the ROI-Interval—the nefarious intention.

Gaming Summary In this case, the project fails to be in the ANN-protocol 95%ECI and so that project would not qualify for further consideration.

Elaboration Additional feedback **from H&L(2021, p.100)** continues [**bolding added**],

- - - the following are productive extensions of this research report: 1. One could examine Larger Panels n > 12, to evaluate the impact of ANN-replacements. A sample size of 12 for the OLSR-fit may be close to the practical or reasonable limit of a time series. - - - ***We have reported that for smaller sample sizes there seems to be an increase in the Smoothing proportions. Perhaps the inverse may be the case. For large Panels, perhaps there is more of a balance between the Smoothing and Provoking effects.***

It MAY be the case, that:

(i) ***H&L Conjecture I***: Longer Panels would ***mollify*** the Smoothing effect, or

(ii) ***H&L Conjecture II***: The ***balance*** of the Smoothing- and Provoking-events and/or serious-events may not differ for larger Panels.

In either case, this MAY mitigate against gaming. Specifically, for ***Conjecture I****:Mollification* If the mollification creates such a slight Smoothing or Provoking ANN-effect so that the change in the OLSR ANN:(1-FPE)ECI differs only in an unimportant magnitude akin to OLSR-Noise this would render ANN-gaming moot; or, for ***Conjecture II*** *:Balance* If the Smoothing- and Provoking-effect percentages are not in the mollification-set AND do not test as not being equal in percentage-terms that fact would also render the temptation to game the ANN-protocols moot as it not clear whether the Ann-effect would be Smoothing or Provoking.

Testing Overview

To test the two H&L Conjectures, we have accrued a dataset to investigate the effect of a larger Panel, n=21, on the Smoothing v. Provoking for the four arms: Panels of: {n = 6, 9, 12 & 21}. The rationale for our research interest in the ANN-Effect of outlier modifications is:

IF *the ANN-modification seems to result in a Smoothing- or Provoking-tendency [****Shrinking or Expanding the RRECI****], this raises a possible issue of inappropriate judgment cueing—i.e., a dysfunctional consequence that may introduce a "gaming set of behaviors". If the ANN-protocol often creates a* ***dramatic*** *alteration— smaller* ***or*** *larger in the ECI—this asymmetry re: Smoothing* ***or*** *Provoking could influence the choice that is made to replace outliers. This is relatively insidious as, depending on the decision-makers' Result-Utility-Preference Function, the decision-maker, for various reasons sometimes dysfunctional, may:*

1. eschew the replacement of outliers when needed: An Unwarranted Rejection to Replace an Outlier, or

2. act to make the replacement of outliers when it is not needed: An Unwarranted Election to Replace an Outlier.

Research Agenda: Modification Protocols of The Account Base

Overview

In this section, we will present the research agenda to address the following question of interest:

Would a Long Panel (i) mollify the ANN-protocol's Smoothing- or Provoking-effects to the extent that they are not likely to dramatically affect the 95%ECI of the ANN-modified Panel relative to that of the Basic Panel or (ii) result in a balance between Smoothing- and Provoking-events thus rendering the gaming of the ANN-protocols moot.

Accounting Variable Set for Testing the Panel-Length v. ANN-Impact Effects

For each firm, we selected four (4) Income Statement variables: **{*Gross Profit; Operating Income; Earnings for the Common Shareholders; Shares for Diluted Earnings per Share*}**, and four (4) Balance Sheet Statement variables: **{*Current Assets; Other Assets & Deferred Charges; Current Liabilities; Current Ratio*}**. ***Screening Caveat*** The ECI are the widest confidence intervals compared to the Fixed- and Random-Effects versions. Thus, sometimes the Lower Limits of the ECIs are negative. This can occur when there is anomalously high Panel variation. This would have the tendency to compromise the FPE-Null rejection logic giving an illusion of no difference when in fact in a "standard population" failing to reject the Null may not likely be the case. To control for this, rather than screening for non-Ergodic Panel-profiles, we simply screened-out any cases where the ECI: Lower-Limit was in the negative quadrant.

Panel Lengths The question is: *What is a reasonable Panel-length for testing the Panel Length-variable that has a practical context for usual forecasting problems for Panel-data from market traded firms?* We selected a Panel size of 21 years for the following reasons: (i) annual series are by definition non-cyclical in the Quarterly SEC-reporting context and thus have less variation and so enhanced precision, (ii) the accrual started 2000 through 2020. This is the longest period where the PCAOB was empowered to license Audit LLPs and so presumably avoids the defalcations of the 1990s where the veracity of the reported GAAP-information was sometimes in question, and (iii) a Panel-size of 21 was the Mean of the Panel-sizes of the 181 Series selected by Collopy and Armstrong (1992) from the Makridakis forecasting Panel study: Makridakis et al. (1982, M-Competition)**.** During February 2021, we collected a sample of 17-firms, Appendix B[TableB2[n=21]], from the Bloomberg™ Terminals [BBT] in the John and Diana Conners Finance Trading Lab at the State University of New York: College at Plattsburgh. These firms were randomly selected from the BICS©-listings [other than Utilities] that were in the top 25% of their BICS-group's Market Cap. Thus, the BBT21 is offered as the longest practical forecasting Panel by which to benchmark the H&L datasets and test the H&L Conjectures: re: the Panel-size effect of the ANN-Protocols.

I:Mollification & ***II:Balance***. In addition, the Panels to be benchmarked by the BBT21-Panel are the firms accrued in the H&L-studies that were also accrued from the BBTs. [Appendix B[TableB1[n=12]]. For the H&L-studies there were three Panel sizes: The Download Panel: H&L12[n=12]; The Mid-Length Panel was created by removing three Points: $\{\boldsymbol{x_{t=10}}$ & $\boldsymbol{x_{t=11}}$ & $\boldsymbol{x_{t=12}}\}$ from H&L12. H&L9[n=9] is then: $\{\boldsymbol{x_{t=1}}, - - -, \boldsymbol{x_{t=9}}\}$. Finally, H&L6 removed three Points: $\{\boldsymbol{x_{t=7}}$ & $\boldsymbol{x_{t=8}}$ & $\boldsymbol{x_{t=9}}\}$ from H&L9; thus: H&L6 $\{\boldsymbol{x_{t=1}}, - - -, \boldsymbol{x_{t=6}}\}$. These were accrued February 2020.

ANN-Codex

For each of these Panels: {BBT21, H&L12, H&L9 & H&L6}, we made the following three ANN-replacements.

ANN: Replacement[**E**arly]: The Second Panel Point: $(\boldsymbol{x})_{\boldsymbol{t=2}}$,

ANN: Replacement[**L**ate]: The Next to Last Panel Point: $(\boldsymbol{x})_{\boldsymbol{t=(n-1)}}$, and

ANN: Replacement[**B**oth[Early & Late]]: $(\boldsymbol{x})_{\boldsymbol{t=2}}$ & $(\boldsymbol{x})_{\boldsymbol{t=(n-1)}}$

Testing Variables: The Inference Context for the Focus of Research

To develop informative profiles that address precision impact-effects of the proposed ANN-replacement modifications: {*Basic Panel* [*BP*], *Early, Late & Both*}for the{*H&L6; H&L9; H&L12 & BBT21* }Panels, we measured (i) Smoothing and Provoking ANN-effects, and (ii) also, screened them to focus on serious ANN-impacts.

General Smoothing & Provoking using The Relative Ratio of Precisions:[RRP]

For the RRP blocked by the Panel-size, the following computation is made for each Account-variable of each firm:

RRP[$\boldsymbol{P_j}$] = Precision[$\boldsymbol{P_j}$] / Precision[$\boldsymbol{P_\bullet}$]

EQ1

Where j = {Early : Late : Both} and $\bullet$ is only the Basic Panel [BP] as downloaded.

The Sensitivity Context

RRP[$\boldsymbol{P_j}$] is: *a ratio measure of magnitude of the relative precisions.* The RRP uses the metric $\boldsymbol{R_{ANN}}$: $[\frac{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{ANN}}}{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{BASIC}}}]$ and thus can be used to intuit Smoothing or Provoking tendencies as demonstrated above using [R2].

Where: PECI is the Precision of the Excel 95% Capture Interval[v]. ***Point of Information*** The Panel size must be the same for the numerator and the denominator of the RRP[$\boldsymbol{P_j}$]. This is important as the width of the Precision-interval changes inversely with the size of the Panel and thus could be an insidious confounder if the sample size of the numerator and the denominator were to differ.

Discussion

In this case, we are measuring the ANN-effect as follows:

1. If the PECI of an ANN-modified Panel were to have been less than that of the *BP*, the relative ratio $\boldsymbol{R_{ANN}}$: $[\frac{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{ANN}}}{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{BASIC}}}]$would be <1.0 and thus be located on the Left Hand Side [LHS] of 1.0. This would then be a Smoothing-event resulting in ***Shrinking the RRECI—the lower the ratio the more the shrinkage and so the smaller the ECI***,
2. if the PECI of the ANN-modified Panel were to have been more than that of the *BP*, the relative ratio would be >1.0 and thus be located on the Right Hand Side [RHS] of 1.0. This would then be the Provoking-event resulting in ***Expanding the RRECI the higher the ratio the more the expansion and so the larger the ECI***, and
3. otherwise the relative ratio is =1.0 and so there is NO ANN-impact effect.

The Comfort Zone [CZ]

The logic of using a CZ-filter is to focus on "***Events of Interest***". In addition to testing for the general ANN-Smoothing or Provoking discussed in *4.1*, it is of interest to investigate the ***seriousness*** of the ANN-modifications. The H&L-studies offered profiles for their three panel sizes where the Noise or minor perturbations were reported. They drew the Noise frontier at ±2.5%; this was called the ±2.5%Comfort Zone[CZ] to wit any activity interior to a relative precision displacement of ±2.5% was in the unavoidable Noise-Zone. Thus, in this case, effectively nothing can be done to avoid such a relative Precision displacement and so this is the "***No-Worry-Zone***". As an elaboration of the Simple ANN-effects [≠1.0], we created the CZ-partition of our datasets to examine the Seriousness of the ANN-events. In this regard, we used the measure:

IF $\boldsymbol{ANN^{Serious}_{RRP}}$: $[\frac{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{ANN}}}{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{BASIC}}}]$<[1-2.5%] OR $\boldsymbol{ANN^{Serious}_{RRP}}$: $[\frac{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{ANN}}}{[\boldsymbol{PECI}]^{95\%}_{\boldsymbol{BASIC}}}]$ > [1+ 2.5%], then we will be measuring the Seriousness of the ANN-effects as follows:

1. If the relative ratio: $\boldsymbol{ANN^{Serious}_{RRP}}$ were to be less than the lower-limit of the CZ of 0.975, this would then be the noted as a Serious LHS-Smoothing-event—i.e., a dramatic ***Shrinking of the RRECI***,
2. if the relative ratio: $\boldsymbol{ANN^{Serious}_{RRP}}$ were to be more than the upper-limit of the CZ of 1.025, this would then be the noted as a Serious RHS-Provoking-event—i.e., a dramatic ***Expansion of the RRECI***, and
3. otherwise the relative ratio is IN the CZ: [0.975 : 1.025], and thus there is NO Serious ANN-impact effect—i.e., the impact of the ANN-protocol is IN the ***No-Worry-Zone***.

The Vetting Context

In this case, the vetting test addresses the magnitude of the PECI benchmarked by the related Forecast for each of the four Panels. See Lusk (2017) where another context is offered. Recall, the Forecast is the midpoint of the PECIs; however, it is independent of the width of the ECIs which is a function of the OLS-variation and the Panel-size. Thus, the ratio of the PECI / Forecast will serve as the simplest unitization so that ALL the inferential comparisons will be on approximately the same metric scale. This is referred to as unitization or metric-normalization. For example, the measured units for Current Assets and Current Ratio are very different and so for inferential testing need to be tested on a unitized-scale.

The Vetting measure is the Precision Relative to the Forecast measured as:

PRF[$\boldsymbol{P_j}$] = [PECI [$\boldsymbol{P_j}$] / Forecast [$\boldsymbol{P_j}$]]EQ2

Where j = BP : Early : Late & Both

The Sensitivity Context

PRF[$\boldsymbol{P_j}$] is: *a ratio measure of the magnitude of the precision as unitized/benchmarked so as to neutralize any confounding magnitude effects.*

Discussion

In this case, the PRF is going to be used in a vetting context as follows:

1. *Vetting Test A* We assume that the PRFs blocked by the Panel Lengths: {H&L6 : H&L9 : H&L12 & BBT21} will be inversely ordered relative to the number of Panel-points,

2. *Vetting Test B* We assume that the PRFs of the ANN[Both]-Panels blocked by the Panel Lengths: {H&L6 : H&L9 : H&L12 & BBT21} will result in the a increasing ANN[Both]-effect as the sample size decreases,

3. Both of these vetting expectations are due to the nature of the mathematics involved. The mathematics are that Panel Size is an inverse non-linear driver of Precision and H&L suggest that the ANN-tendency produces on net Smoothing,

4. The vetting-standard is that these two expectations if NOT inferentially in evidence would call into question the ***generalizability*** of the results of the hypothesis tests. Recall the reason for vetting tests is to explore the nature of the generalizability of inferential testing. ***Thus, these vetting tests, if founded, suggest that the variables and the accruals are likely to form a useful sample from a population of non-random effects of firms traded on active exchanges. And to that extent enhance the degree of inferential creditability of the inferential a priori test results.***

5. Testing Profiles: Sample Accruals

The FPE & FNE Credibility of the Inferential Framework

Assume we are interested in the most disaggregated non-directional test of proportions [ToP] for two-sampled populations. This ToP-sample estimate also is a useful sample size "*rule of thumb*" for two-sample inference estimates for variables that are or could be continuous in nature. This will be addressed by providing ex-post Power profiles. As is "standard practice" we used a set of Minimal factors: FPE[90%[CV=1.645]] & FNE[75%[CV=0.675]] as a guide to determining the number of firms to accrue for the tests proposed. Minimal in our testing context indicates that if the number of observations falls below these values that would compromise the inferential precision of the testing partition as triaged. This approach is consistent with controlling the p-value as suggested by Benjamin (2018); additionally, see: Kim, Ahmed & Ji (2018). Simply, we will not be able to disaggregate i.e., block, the testing so that the number in the testing-cells are less than this minimal-value. In this case, we assumed that Population[A] has a proportion of H%A and the other Population[B] has proportion of H%B.

Details

To form the sample accrual information in this two-population context, we used the following standard ToP sample size formula: See Wang & Chow (2007):

Sample size = **[1.645 + 0.675]^2**

$$[\mathbf{1.645 + 0.675}]^{\mathbf{2}} \times [\frac{[\boldsymbol{H\%A} * [\mathbf{1} - \boldsymbol{H\%A}] + [\boldsymbol{H\%B} * [\mathbf{1} - \boldsymbol{H\%B}]}{[\boldsymbol{Abs}[\boldsymbol{H\%A} - \boldsymbol{H\%B}]]^{\mathbf{2}}}]$$

In this case, to initialize the computations, we used as a typical proportion-set for our context: [75% v. 85%]. We selected [75% & 85%] as they give identical results with their binary-partner of [25% & 15%]. This gives more or less, a boundary range of a sample size of 170 per generalized ToP-testing partition. In our study-frame, we have: One accrual set of 22 Firms that have three Panels of size 6, 9 & 12 and one accrual set of 17 firms that has one Panel with 21-Panel Points. Each Panel will generate four (4) ANN-tests: {Basic, Early Late & Both}. Each firm has eight (8) Financial Accounts reported on the BBTs. This gives a total number of projected accrual points of: 2,656: [17×8×4 + [22×8×4]×3]. Owing to our accrual protocols, if there any missing data that Panel was eliminated. The final accrual-set had 2,350 values. The short-fall, not atypical, was 11.5%. This dataset would enable, if deemed necessary, about 13-disaggregation-partitions for testing. This is certainly a sufficient number to engage in productive hypothesis as well as exploratory analyses.

Results: Impacts of the ANN-Protocol v. Length of the Panel

Basic Smoothing & Provoking

There are two research-hypotheses of interest regarding Longer Panels: H&L Conjecture I: Mollification & H&L Conjecture II: Balance between Smoothing & Provoking. We will treat each in order. The central issue addressed by both, as discussed in the introductory section, is: ***According to H&L(2021), there was a preliminary indication that: (i) the magnitude and nature of the impact of ANN-modifications on forecasting precision are inversely related to the number of Time-Series points in the Panel so that the more panel-points, the less ANN modification matters with regard to the precision-effects, and/or (ii) longer Panels may be associated with a balance between Smoothing and Provoking***. This is of interest as either condition will render gaming the ANN-protocols moot.

H&L Conjecture I: Mollification

In this inferential enterprise, we created: (i) four Panels: {n=6, n=9, n=12} with Panel, n=21 as their benchmark}, and (ii) three ANN-modifications ANN:{Early. Late & Both}to probe where in the space defined by the nine: {Panel-Length & ANN-modifications}test-sets move from a high-precision-modification-impact ***not a desirable state*** to a precision-impact that is effectively in the "Noise" or the "No Worry Zone" where the Basic Panel set is effectively not practically affected by the ANN-modification ***a desirable state***.

In this inferential test, we used the RRP[$\boldsymbol{P_j}$] = [Precision[$\boldsymbol{P_j}$] / Precision[$\boldsymbol{P_{\bullet=BP}}$]] EQ1 as the measure. The inferential test, *H1*, will use the Median-values of RRP blocked over the three ANN-modifications & blocked over the three Panels {H&W6 : H&L9 : H&L12}. The operative Null-test for *H1* is:

H&L Conjecture I: $\boldsymbol{H1_{Null}}$ *The Median-magnitude effect of the Smoothing-effect for the BBT21-benchmark dataset is ≥ those from the three H&L:Panels blocked by the ANN-modifications:{Early, Late & Both}. This is a directional test for mollification using the Median as the inferential measure.*

The results of this *H1*-profile are:

Table1. Smoothing Results for the RRP-Median-tests of the Null of H1 *WRT: With Respect To.

ANN-Protocols	**H&L6**	**H&L9**	**H&L12**	**BBT21**
Early	0.962666	0.995305	0.996979	0.99979
p-value *wrt BBT21	<0.0001	0.001	0.079	N/A
Late	0.931080	0.989516	0.994724	0.997552
p-value wrt BBT21	<0.0001	0.003	0.061	N/A
Both	0.816556	0.958340	0.978193	0.991406
p-value wrt BBT21	<0.0001	<0.0001	0.015	N/A

Discussion H1

The focus of *H1*, can be simply summarized as: We are investigating the assertion of H&L (2021, p.99) where: LP: MP & SP are H&L12 , H&L9 and H&L6 respectively that:

It is suggestive from this Chi2-analysis that the Smoothing percentages are inversely related with the Panel sizes: LP:[86.1%[173/201]] MP[90.2%] & SP[98.5%]

The inverse relationship--higher percentage of smoothing-events the smaller the panel referred to by H&L was for the counts or the number of the Smoothing-events. We have extended the test to the magnitude of the RRP-

ANN-effects. Here it is critical to remember that mathematically the width of the CIs is an inverse function of the Panel-size; however, as the RRP-ratio measure [EQ1] uses the same Panel-sizes for each of the events of the Columns of Table 1, the ANN-impacts are not confounded by the changing sample-sizes. Specifically, the measure of interest is the Median of the ratio of the Precision of the ANN-modified series to that of the related Basic Series i.e., RRP. Recall, RRP-ratios < 1.0 indicate that the ANN-modification was Smoothing smaller relative ECIs of the ANN-modification-series compared to the ECI of the Basic-series and *visa-versa* for Provoking-events. Underlying the logic of *H1* is that the OLSR-variation standard error is the root of the Sum of the Squared Errors of the OLSR benchmarked by the degrees of freedom $\sqrt{\boldsymbol{SSE/(n-2)}}$. In this computational frame, the size of the Panel, n, is the benchmark for the OLSR-variation. However, n is also a driver for the SSE. Therefore, *offered without proof*, there is no *a priori* way to anticipate the magnitude or the direction of: $\boldsymbol{\partial SSE/\partial n}$ for any ANN-RRP-effect to wit, ***the only way to understand the relationship between Panel Size and RRP is to form a blocked empirical testing context.*** Additionally, it is the case that as the number of points in the Panel decreases, the width of the 95%CIs increases and thus the impact of the ANN will likely effect more of an leverage impact on the regression-fit given the smaller n. For example, for the test of the ANN[Early], the ratio-impact was very slight for the BBT21 Panels: where it was on the order of 0.00021 [1- 0.99979]. whereas, for the H&L6 Panel the effect was 0.037334 [1- 0.962666]. The impact ratio in the Median measure is: 177.8 [0.037334/0.00021]. However, there is a conditioning aspect that depends on the nature of the generating process. For example, correlated processes a Smoothing tendency is expected as argued by H&L; as an illustration, consider the the sinusoid f(x): sin(x) + x; {x: 1, - - -, 12} where the correlation: [f(x) w/Time] is 98%. Nine of the ten ANN-replacements indexed from $\boldsymbol{t_2}$ through $\boldsymbol{t_{11}}$ produced Smoothing events. Interesting is that even though f(x) is everywhere concave to the OLSR-fitted function, f(x) does have not smooth point to point transition points. This can, and in the case of this ANN-protocol did effected a reorientation of the regression-fit, that in turn created one Provoking-event. ***Inferential Clarification*** In the test of *H1*, we are only using the Median relative ratios as there were a number of Tukey-Box Plot Whisker-Outliers. Thus, rather than filter these outliers or used a trimmed-Mean, we used directional Median-tests of the nonparametric comparisons for each pair produced by the Wilcoxon Method of the Wilcoxon / Kruskal-Wallis Tests (Rank Sums):SAS[JMP]v.13. These Pairwise p-values only use the exact datasets of the Pairs and not the relative-pairwise-data to the overall set of data.

The Inferential Essence of Table 1 re: H1

In this case, using the relative pairwise p-values of Table 1, it is very clear that all of nine Cell-Nulls may be rejected with assurance simply, the directional effect noted in the Null is not founded given the Median-profiles. Note there are in total 18 [3 × $\boldsymbol{C_2^4}$] pairwise-tests. We are only reporting the nine that are the BBT21-benchmarks. Table 1 indicates that all of the relative precision magnitudes trace an inverse-Smoothing relationship as the Panel size increases over the ANN-Protocols tested. The important implications on the test of H&L Conjecture I are:

1. There is statistical FPE-Evidence that the BBT21 dataset overall has a very-low level of Smoothing. The average difference for BBT21 as ANN-Modified relative to the BBT21-Basic is 0.38%. [1- AVERAGE [0.99979 + 0.997552 + 0.991406]]; effectively, this indicates that at a Panel size of 21 the ANN-modifications have a relatively modest effect on the relative precisions of the modified Panels with respect to their Basic Panels. ***Point of Clarification*** The Median Measure of the ANN[Early] is 0.962666 thus the ANN[Early] resulted in a Smaller ECI compared to the Basic Panel; for the ANN[Late] the Median Measure is 0.931080, thus, the ANN[Late] resulted in a smaller ECI compared to their Basic Panels and the Width of the ANN[Late] is ***not as small*** as the width of the ANN[Early] CI as 0.962666[Early] is > 0.931080 [Late]; thus the ANN-impact for the ANN[Late] is more than it was for the ANN[Early],
2. If we use the Comfort-Zone of ± 2.5% for the Medians *this is a simple screen not inferentially tested*, then the Shaded Cells give an approximate frontier or neighborhood where there is a transition from a likely important- or serious-ANN-effect to a Panel-size [not shaded] where the ANN-effect is not likely to be judged "of concern". For example, for H&L12 [lightly shaded], if the ANN-Both is used to replace outliers then the effect is "just about" at the serious juncture as [1 – 0.978193] is 2.18% which is just < than 2.5%,
3. This ***L-shaped*** frontier effectively indicates that Panel Sizes at around 6 will likely produce a worrisome ANN-modification relative to the Basic Panel; if an ANN-Both replacement is used then H&L6 & H&L9 and maybe H&L12 may also be in the "worrisome" set. In this case, worrisome indicates that the ANN-modification materially changes the modified series relative to the Basic Panel a problematic analytic effect, and
4. For generalized smoothing RRP<1.0 , the partition is binary and so only one aspect is in need of testing. In this regard, aggregating over the ANN-events, there are 1,240 instances of the 1,753 tested where the ratio of the ANN to the Basic Dataset was < 1.0 or a percentage of 70.7% In this case, 95%ConfidenceInterval for the LHS-result is: [68.6% : 72.9%]. This is an up-date of the H&L-studies where they reported only the central tendency of 64.3%[314/488] for one arm of the accrual where the Panel was n =12. For completeness, the 95%CI for the RHS:Provoking is: 29.3% [[1-72.9%] : [1-68.6%]] or [27.1% : 31.4%]

Inferential Summary H1[Mollification with respect to BBT21]

This addresses H&L Conjecture I. There is clear evidence that the Null of H1, $\mathbf{H1_{Null}}$, can be rejected in favor that BBT21 the longest Panel, exhibits the lowest relative mollification in that all the p-values moving from BBT12 to H&L6 over trhe ANN-effects exhibit lower p-values suggesting a progressively greater justification for rejecting that $\mathbf{H1_{Null}}$ is the state of nature. As confirmatory, but exploratory, indications the "L"-shaped seriousness figure is certainly consistent with the observed progressive mollification. Thus, $\mathbf{H1_{Null}}$ is duly rejected offering progressive Mollification as the likely State of Nature for the datasets tested. This result leads naturally to the investigation of the following question: As mollification seems likely the relative case using the Median-measures, does this imply that there are not likely to be serious Smoothing or Provoking asymmetries that would offer gaming opportunities?

H&L Conjecture II Balance of Smoothing- vis-à-vis Provoking-events.

In addition to testing for general ANN-Smoothing or Provoking regarding *H1* RRP≠1.0, it is of interest to investigate the ***seriousness*** of the ANN-modification-effects. The logic is: If there is a balance between serious Smoothing and serious Provoking ANN-events, that would likely render gaming the ANN-protocols moot. The H&L-studies offered Profiles for the three panel sizes where the Noise or minor perturbations were reported. They drew the Noise-frontier at ±2.5% that was called the ±2.5%Comfort Zone[CZ] to wit, any activity interior to a relative precision displacement of ±2.5% was in the unavoidable statistical-Noise-Zone. In this case, effectively nothing can be done to avoid such a relative Precision displacement and so this is the "No-Worry-Zone". Thus, as an elaboration of the ANN-effects presented in Table 1, we created the same CZ-partition of our datasets to examine the Serious-ANN-events profile. The scoring used to create Table 2 is: For the full dataset, we used the RRP-measure:

IF $\boldsymbol{ANN_{RRP}^{Serious}}$: $[\frac{[PECI]_{ANN}^{95\%}}{[PECI]_{BASIC}^{95\%}}]$ < [1- 2.5%] OR

$\boldsymbol{ANN_{RRP}^{Serious}}$: $[\frac{[PECI]_{ANN}^{95\%}}{[PECI]_{BASIC}^{95\%}}]$ > [1+ 2.5%], then the numerical binary indication 1.0 was recorded; otherwise 0 was recorded. The <[0.975] test is termed the Left Hand Side-effect [LHS]; the >[1.025] test is termed the Right Hand Side-effect [RHS].

H&L Conjecture II Balance Smoothing v. Provoking

The Null-form for the Serious ANN-effects is: $\boldsymbol{H2_{Null}}$ is:

The percentage of ANN:RRP-effects outside of the ±2.5%Comfort Zonei.e., for the LSH v. RHS will not differ for the pairwise-comparisons for {H&L6 : H&L9 : H&L12 : BBT21} aggregating over the location or nature of the ANN-modifications: ANN[Early, Late & Both] . Note this test is assumed to be non-directional test of the Balance Smoothing v. Provoking as offered in H&L Conjecture II.

Table 2 presents profiles of (i) The LHS-Smoothing, and (ii) The RHS-Provoking ANN-events ***not in*** the ±2.5%Comfort Zone:

Table 2. Percentage of RRP-events NOT in the ±2.5%Comfort Zone.

Panel Size	LHS: Percentage	RHS: Percentage	LHS≡RHS p-value
H&L6: n=463	69.5%	1.1%	<0.0001
H&L9: n=481	44.1%	4.8%	<0.0001
H&L12: n=488	35.5%	5.7%	<0.0001
BBT21: n=321	*28.7%*	14.3%	<0.0001

Discussion

The Binary{1, 0}-scoring for $\boldsymbol{ANN_{RRP}^{Serious}}$ has the desirable property that the Mean is also the related percentage. Further, given this Binary-scoring the Means and also the Medians can be used as there are no Tukey-Box-Plot issues**.** Thus, we reported (i) the parametric Welch ANOVA, as there are Brown-Forsythe variance differences between the LSH & RHS, and (ii) also the Pairwise Wilcoxon Method of the Wilcoxon / Kruskal-Wallis Tests (Rank Sums). If these two p-values are not different then only one p-value is reported.

Results for H2 [Balance: Smoothing v. Provoking]

The inferential information of Table 2 is clear. For the Mean and Median tests $\mathbf{H2_{Null}}$ **is strongly rejected** in favor that there is not likely a ***Balance of Serious Smoothing v. Serious Provoking events for the Row-contrasts***. Actually, the LHS or Smoothing events dominate the Provoking or RHS events and this is a vetting check for the conjecture of the H&Ls studies where for Panels of accounting data from firms traded on active exchanges Smoothing is the tendency observed. ***This result offers a clear indication, that even given the mollification results from the test of and rejection of $H1_{Null}$, there is not a likely balance between the LHS: Smoothing and the RHS: Provoking to allay the temptation to engage in dysfunctional gaming in electing to apply ANN-protocols. As this is the case, it behooves us to further profile certain elaborations of Table 2.***

Elaboration I of Table 2[95%CIs]

As an elaboration of the information of Table 2, the ANN:RRP-effects as profiles and confidence intervals as they relate to: {The LSH OR The events in the ±2.5%Comfort Zone OR The RHS} are offered in Table 3. An indication of this was also given in H&L(2021, Table 3, p.97) where the 95%CI of the RRP **not** in ±2.5%Comfort Zone overall was [50.5% : 55.7%]. Thus, to elaborate on this information, we offer the details on Panel size over the ANN-effects for the three Profiles. Note the sample-sizes for the Panels is given in Table 2.

Discussion

This 95%Confidence Interval profile information is critical in understanding the Panel Size effect vis-à-vis the Panel size relative to the H&L Conjecture II i.e., the Balance between Smoothing and Provoking. This is an elaboration of the result presented by H&L (2021, Table 3, p.97) where it was reported that the balance between the events OUTSIDE the ±2.5%Comfort Zone and the those IN the ±2.5%Comfort Zone was effectively 50% :50%. We see this also in the Overall Row where outside the ±2.5%Comfort Zone there

Table 3. 95%CIs of{Profiles Panel Groups} v. {LHS : IN ±2.5%Comfort Zone : RHS}.

	LHS Profile[RRP<0.975]			In Profile: ±2.5%Comfort Zone			RHS Profile[RRP>1.025]		
Group	**LLimit**	**Average**	**ULimit**	**LLimit**	**Average**	**ULimit**	**LLimit**	**Average**	**ULimit**
H&L6	65.3%	69.5%	73.8%	25.2%	29.4%	33.5%	0.1%	1.1%	2.0%
H&L9	39.6%	44.1%	48.5%	46.7%	51.1%	55.6%	2.9%	4.8%	6.7%
H&L12	31.2%	35.5%	39.7%	54.4%	58.8%	63.2%	3.7%	5.7%	7.8%
BBT21	23.7%	28.7%	33.6%	51.6%	57.0%	62.5%	10.5%	14.3%	18.2%
Average	40.0%	44.4%	48.9%	44.5%	49.1%	53.7%	4.3%	6.5%	8.7%
Numbers		Number	Percent		Number	Percent		Number	Percent
Overall		799	45.6%		852	48.6%		102	5.8%

Table 4. 95%CIs Disaggregated Profiles by ANN:RRP-LHS Effects.

Panel Sizes	ANN-Early LHS: 95%CI	ANN-Late LHS: 95%CI	ANN-Both LHS: 95%CI
H&L6:	50.6% : 58.4%[<0.001]: 66.3%	56.0% : 63.6%[<0.001]: 71.3%	81.0% : 86.5%[<0.001]: 91.9%
H&L9:	26.3% : 33.8%[0.03] : 41.3%	32.3% : 40.0%[0.07] : 47.7%	50.7% : 58.4%[<0.01] : 66.1%
H&L12:	19.7% : 26.5%[0.38] : 33.4%	24.1% : 31.3%[0.39] : 38.5%	40.7% : 48.5%[0.18] : 56.2%
BBT21:	11.5% : 19.2% : 26.9%	18.4% : 26.9% : 35.3%	30.1% : 39.4% : 48.8%
WAverage[MidPoints]	34.5%	40.4%	58.2%

were 51.4% [45.6% +5.8%] of the events and in the ±2.5%Comfort Zone there were [48.6%] of the events or about a 50%/50% balance. Thus, as an elaboration of Table 2, we have expanded the information to include the 95%CIs of the Serious LHS- & Serious RHS-events. For example, Table 3 gives valuable expectation information; we see that the 95%CIs are useful guides in the strength of the rejection of $\mathbf{H2_{Null}}$. For example, if a Panel has n=12 years, then using the 95%CIs the general expectation that an ANN-protocol will be in evidence less than 31.2% of the cases is expected 2.5% of the time. Thus, Table 3 leads to the final disaggregation where we examine the ANN-effects: {Early, Late & Both}. This information is presented in Table 4. ***To avoid information "overload" we will only present the Smoothing effects as they are the most dominant.***

Discussion

In the final set of ANN-profiles Table 4, we offer the 95%ECIs for LHS results: [Lower Limit : **MidPoint** :Upper Limit] blocked by the three ANN-protocols. In addition, we added a p-value noted in **[]s**. The p-value information in Table 4 is the average of the Wilcoxon Pairs Method[Median] and the Tukey HSD-ordered Pairs[Mean] tests for the comparison of BBT21 v. {H&L6 or H&L9 or H&L12}. These 95%CI are now illustrated. For the ANN[Late] for H&L12 there were 51 Serious LHS-events or 31.3%. [51/163]. The 95%CI for this cell-result is [24.1% through 38.5%]. Further, this percentage [31.3%] was tested against the BBT21 result of [26.9%]. The non-directional average p-value was 0.39. The p-value [0.39] suggests that the difference of: 7.2% ABS[24.1% – 31.3%] is not sufficiently large to reject the Null that these observed means could have come from populations with the same means. Simply, the Smoothing difference between Panel sizes of 21 and 12 is not likely to be sufficiently large to make a difference in most practical cases. This is the most disaggregated RRP:ANN-profile and thus provides the details on the impact of the specific ANN-profiles over the four Panel-sizes. Ideally, we would prefer if the ANN-outlier-replacement was "neutral" relative to the Basic Panel and so produced a balance of Smoothing and Provoking of serious LHS & RHS events. Table 4 is likely to be the most useful profiler. For example, if the Panel has n=12 years, then the specific expectation is contingent on the nature of the ANN-protocol to be used. If the outlier replacement is ANN[Early], it is unlikely that Smoothing will occur less than for 19.7% of the cases; if the outlier replacement is ANN[Late], it is unlikely that Smoothing will occur for less than 24.1% of the cases; otherwise, it is unlikely that Smoothing will occur for less than 40.7% of the cases. This is important as it indicates the expectation deconstruction for specific ANN-protocols.

Conversational Inferential Summary: H2 For the inferential indications of the seriousness of the impact of the ANN-protocols profiled from Tables 2 & 3 & 4, there is clear evidence that:

1. the Null of H2, $\mathbf{H2_{Null}}$, of a Balance of Smoothing and Provoking Serious ANN-events can be rejected in favor that Smoothing rests as the dominant tendency overall, and
2. the shortest Panel, H&L6, has the highest LHS/RHS ratio of Table 2[63.2]; and, as one moves systemically to the larger Panels the ratio decreases as follows:

Panel Size v. LHS/RHS: {(6, 63.2) : (9, 9.2) : (12, 6.2) & (21, 2.0)} interestingly, the best fit for this graphic is: y = -44.26ln(x) + 125.47.

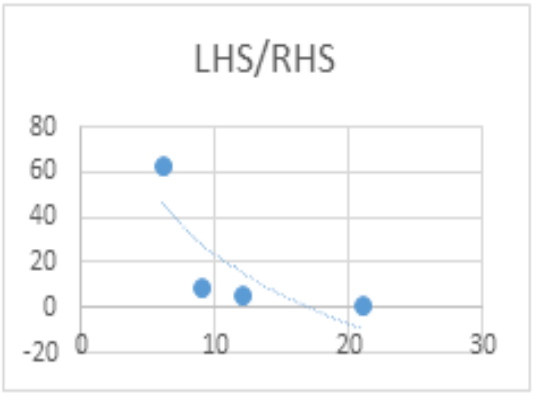

Figure B Log Profile of LHS/RHs

Overall Indications re: H1 & H2

H1: Progressive-Inverse Smoothing mollification is in evidence over the four Panel-sizes: {H&L6 → H&L9 → H&L12 → BBT21} See Table 1. This pertains to the test of H&L Conjecture I: Progressive Mollification which is clearly founded, and

H2: Pertaining to H&L Conjecture II: Balance From Table 2, there is clear inferential evidence that Smoothing [LHS] *is the dominate effect. As an elaboration, not tested, is that the ratio of*]*Smoothing / Provoking*]*-events is a decreasing function of the Panel-size presented in Figure B. This is not inconsistent with the mollification of Table 1 or with the asymmetry of Smoothing v. Provoking in favor of Smoothing. Further, the largest Panel has the lowest ratio of 2.0 of Smoothing to Provoking; unfortunately, this this is not likely to be close enough to 1.0 to estop individual from gaming behaviors in executing ANN-protocols. Tables 2 & 3 & 4 provide details and useful elaborations on the pervasive nature of Smoothing: both in general and of a serious nature—i.e., outside the ±2.5%CZ..*

After we examine the Vetting results, we will address the principal question that is: *Is a Panel-size of n=21, sufficiently long as to allay concerns over Gaming the ANN-outlier protocol?*

Vetting Profile: Confidence Re-calibration Generalizability

Recall, the reason for vetting tests is to explore the nature of the generalizability of the inferential testing reported above for *H1* & *H2*. In the current practice of statistical analyses, it is the usual case when the p-value is the inferential measure that "a background" check of sorts is performed. This speaks to assuring those interested in the inferential results that the context of the modeling methods pertains to the expected population context. Thus, ***vetting is simply the analysis of the analytic protocol that addresses the assurance that the sample accrual and the expected population are in-sync.*** Vetting is thus a derivative of the first work of Fisher (1925), Meta-analysis of Rosenthal (1984), and Meta-Science. See: Aryal & Khanal (2013) and Fahey (2019). An illustration will be instructive. Assume that there is a study to see what set of APPs users prefer regarding; ***Screening streaming-sites dedicated to "First-time-releases" of music-demos***. An accrual population is collected upon which the study will be based. Someone looks at the profile of the accrued population. For this vetting-test, 67.3% of the sample are males the Median-age being 43. As the lead researcher, would YOU commit your research budget to move the APP-analysis forward? *Question rhétorique!!!!*

Vetting the ANN-profile accrual population

There are two vetting tests created using the ratio: Precision / Forecast [PRF[$\boldsymbol{P_j}$]]:

(i) *Vetting Test:A* For the Basic Data-Panel as downloaded, we expect that the widest relative ratio: [Precision / Forecast] will be for the H&L6 and will decrease as the sample-size increases. This is due to the fact that the width of both the Confidence and the Excel Capture Intervals are inverse-functions of sample-size. Thus, Precision benchmarked by the Forecast should neutralize the un-scaled magnitude effect and should result in Precision-magnitudes, the ordering of which follows, in the main, the inverse of the sample-size. In the test of this we will be using the Median as this measure controls for outliers in most cases, and

(ii) *Vetting Test:B* Regarding the impact of the ANN[Both]-protocol re: the ratio: [Precision/ Forecast], it is not unreasonable to expect that the ***difference*** between the PRF of the Basic Panel and that which results after the application of the ANN[Both]-protocol will increase as the sample size decreases. Thus, there would be a move in the direction indicating greater likelihood to reject the logical-pairwise Null as the sample-size decreases. *Rationale*: According to the H&L-studies, it is expected that: (i) the ANN[Both] will likely have a Smoothing-effect in the Panels more than 50% of the time, (ii) for each trial there are two-applications of ANN-protocol: {Early & Late}, (iii) the ANN[Both]-protocol will likely make the greatest impact where the sample size is the smallest, and (iv) mathematically, the smaller the sample-size the larger the precision. Taken together, inferring from the H&L-studies, these four effects are expected to increase the PRF-differences between the Basic Panel and the ANN[Both] as the sample-size decreases. ***Point of Inferential Context*** This vetting test is not inferentially associated with tests of Panel-size re: Precision of the ANN-Protocols as benchmarked by the Precision of the Basic Panel. For example, the Pearson Product Moment and the Spearman association between [RRP[$\boldsymbol{P_j}$] & PRF[$\boldsymbol{P_j}$]] are: [PPM[0.052] & Spearman[0.031]. For both inferential indications the strength of association as measured by the association squared is not indicative of interesting relationships with predictive possibilities.

Vetting Test A

We focused on the un-modified, or the Basic Data-Panel as downloaded. For the Basic Data, we expect that the widest relative ratio: [Precision / Forecast] will be for the H&L6 and will decrease as the sample-size increases. This information is presented in Table 5.

Table 5. Vetting Test A The Size of Unitized-Precision v. Panel Sample-Size.

Datasets for Basic **as Download**	**Ratio: Median[PRF]**	**P-value Pairwise Contrasts**
BBT21	0.359	N/A
H&L12	0.368	[BBT21(v.) H&L12=0.0957]
H&L9	0.428	[BBT21(v.) H&L9=0.01220]
H&L6	0.512	[BBT21(v.) H&L6= <0.0001]

Discussion

Regarding Vetting-test A, the very clear result is that sample-size is the reciprocal driver of the unitized Precision using the Median. The directional decreasing cascading p-values are inferential evidence for the rejection of the Nulls of these blocked comparisons. ***Summary*** **There is clear evidence that as the Sample-Size increases that the Median benchmarked-width of the 95%Excel Capture Intervals decreases and thus, *vis-a-verse*. Thus Vetting Test A is founded.**

Vetting Test B

Here the expectation is that if the H&L-studies are indicative, then there should be a statistically significant increasing difference between the ANN[Both] and its Basic Panel as the sample-size decreases. These results are profiled in Table 6.

Table 6. Vetting Test B The Mollification of the Size of Unitized-Precision v. Panel Sample-Size.

Datasets for ANN[Both]	PRF**:Difference [Basic v. ANN:Both]**	**P-value [Basic v. Both]**
BBT21	0.359 – 0.337 = 0.022	[Basic (v.) ANN:Both=0.3807]
H&L12	0.368 – 0.360 = 0.008	[Basic (v.) ANN:Both=0.2306]
H&L9	0.428 – 0.372 = 0.055	[Basic (v.) ANN:Both=0.1608]
H&L6	0.512 – 0.384 = 0.128	[Basic (v.) ANN:Both=0.0015]

Discussion

The codex for Table 6 is: The Median difference is the PRF[$\boldsymbol{P_{Basic}}$] less PRF[$\boldsymbol{P_{ANN[Both]}}$]. It is the case that for the cascade of the pairwise-comparisons that the ***p-values*** of the differences [**Basic** – *Both*] are decreasing functions of the decreasing sample sizes and the largest effect is for Smallest sample size H&L6 Shaded in Table 6; the order of these three changes were not tested inferentially. ***Summary This is strong vetting evidence that the expected size-interacting mollifications are in evidence. Thus, Vetting Test B is founded.***

Summary of the Vetting Profiles

The vetting of the expectations of an accrual set of market traded firms and the forecasting profiles of the OLSR ECI-precisions provides convincing evidence that these accruals are representative of Ergodic-Panels the sort of which are typically from the population of market traded firms and not produced by random generating processes. This then provides additional FPE-assurance of the inferential results reported for the tests of *H1* & *H2*.

8. Summary Insights and Extensions

Summary

Using both the RRP variable—the Precision of the ANN-modifications benchmarked by Precision of the Basic—unmodified dataset as downloaded—and also applying a screening-filter: ±2.5%CZ, we have created a copious amount of information regarding the ANN-effects created for {Early, Late & Both} outlier replacements tested over four Panel-sizes {6, 9, 12 & 21}. The intent of which is to better understand how the length of the Panel and the ANN-modifications interact re: Precision: Smoothing or Provoking. What we offer as summary insights, reinforced by the founding of the vetting tests, are:

Research Summary Issues of Interest

We offer as a summary the essential take-away points of information.

1. There is statistical FPE-Evidence that (i) the Median Smoothing effects presented in Table 1 are the lowest for the largest Panel BBT21 and, it seems likely that the smaller Panels are more prone to dominant Smoothing-displacements using the Median measures. This was identified as an "**L**" shaped Smoothing-displacement for the ANN protocols suggesting that starting at Cell [Early & BBT21] and

moving in any direction to the Cell [Both H&L6], the Median Smoothing-displacement increases. This **"L"**-trajectory was discussed as an extension of the H&L-studies. Thus, **H&L Conjecture I**: *Mollification* as tested seems to be inferentially founded.

2. Adding the BBT-Panel allows a refinement of the H&L's conjecture of the Smoothing tendency of 64.3% vis-à-vis Provoking of 35.7%. In the discussion for Table 1 [**RRP**] it is reported that the partition of the Smoothing & Provoking was:

For the **LHS:** [Median] the Interval Profile is:
RRP:[Min-Point: 0.108347
→*Median:[Smoothing:70.7%]*→Max-Point:<1.0] and
For the **RHS:** [Median] the Interval Profile is:
RRP: [Min-Point: >1.0 →*Median:{Provoking: 29.3%}* → Max-Point: *1.318535*]
For the **±2.5%CZ RRP**-profile using the overall dataset of Tables 2 & 3:
For the **LHS** [Median], the CZ-Interval Profile is:
Min-Point: 0.108347→*Median: [SS:45.6%]*→ Max-Point: 0.974992]
For the **RHS** [Median], the CZ-Interval Profile is:
Min-Point: 1.02515] *Median:[SP:5.8%]* →Max-Point: 1.318535]
Where: SS is Serious Smoothing & SP is Serious Provoking

This indicates overall as well as for the Seriousness-partition that it is likely the case that the predominant effect is Smoothing: In general, 70.7% or in ratio: 2.4 [70.7% / 29.3%]. For the ±CZ partition, 45.6% or in ratio 7.9 [45.6% / 5.8%]. In this case, referencing the inferential information in the discussion sections of Tables 2, 3 & 4, in general and re: the seriousness-profile there is clear inferential evidence that ***despite*** the founding of Mollification, the **H&L Conjecture II** Balance between Smoothing and Provoking is NOT founded. ***Thus, Smoothing seems the tendency and thus gaming the ANN-protocols may be invited.***

3. As an exploratory result, Hillmer (1984) indicates that when additive-outliers occur ***Early*** rather than ***Later*** in the Panel and current forecasts are recalibrated using such prior-Panel values, the propensity for forecasting errors are increased. This result is certainly the case. However, in the lexicon of forecasting, Hillmers' "tautological" observation has been misconstrued to mean: Replacing observations Early in the Panel has a greater impact vis-à-vis replacing observations Late in the Panel. *Actually, we sort of were of this mindset*. We tested this with the following result. Aggregating over all the Panels and using the RRP-Medians, there are highly significant non-directional pairwise differences in the expected direction for: [Both v.{Early & Late}]: p-values <0.0001: Medians: [0.9543 v.{0.9934 & 9885}]. However, for Early v. Late the non-directional p-value is: 0.067 for which rejecting the Null is usually indicated. However, the direction suggested by Hillmer is not in evidence. If we block the tests over the Panel Sizes, for all of the [Both v.{Early & Late}] and for all of the four Panels, the ordering of the Medians is [Early > Late > Both]; and in the main the non-directional p-values are highly significant for [Both v.{Early & Late}] and much less significant for the Early v. Late contrast. An as indication, observe the Early v. Late v. Both relationships in Table 1. ***In summary, there does not seem to be a strong Early v. Late effect but a very strong and expected [Both v.{Early & Late}]-effect.***

4. ***Final Take-Away*** We examined a very long Panel BTT21 as the benchmark for testing the two H&L Conjectures: Mollification and Balance. Inferentially it is likely the case that even the BBT21, n=21, Panel length will not mollify the Smoothing-events, serious or otherwise, ***to the extent*** that: (i) "most-all" the ANN-effects would be in the No-Worry-CZ, or (ii) there would be a Balance between Smoothing- and Provoking-events. This strongly suggests:

The only way to control the possible gaming dysfunction invited by the ANN-protocols is to require that ONLY outliers that are flagged by one of the standard Outlier Screening Protocols qualify for consideration of ANN-replacement. We prefer the Tukey-Whiskers:Box-Plot Screen.

Further Investigations

The following are productive extensions of this research report:

1. We used the Excel[OLSR] and the Capture Intervals to create the ANN-variable set. Another standard forecasting model is the ARIMA(0,2,2)—i.e., the Holt Model. It would be of interest to research the ANN-impacts on this model,
2. There are three states of ANN-impacts. Smoothing [LHS: $\boldsymbol{R_{ANN}}$ is <1], [Neutral: $\boldsymbol{R_{ANN}}$ is =1.0] or {Provoking: RHS $\boldsymbol{R_{ANN}}$ is >1.0]. It would be a major contribution to develop a forecasting model to anticipate the likely effects of executing a particular ANN-protocol. This may focus on calibrating the Smoothing tendency relative to the nature of the Correlation or the Autocorrelation of the Panel,
3. Adya & Lusk (2016) indicate that a feature affecting the accuracy of forecasting of time-series Panels is complexity. Using their scoring protocol, it would be interesting if or how Complexity and Smoothing interreact in an empirical forecasting context.
4. Perhaps, it would be on a future agenda to test other replacement protocols. Median, Mean and Regression-fill protocols are often used. Additionally, other protocols such as Enders (2010) that address the Missing Data Estimation protocols would be interesting and, to be sure, challenging.
5. Research on the gaming aspect of selecting ANN-replacement protocols to achieve ends that are likely to be judged as dysfunctional re: the intent of replacing outliers so as to ameliorate of the quality of the forecasting process would be most challenging but fruitful,
6. Finally, Hillmer notes that Early outlier replacement protocol are likely to precipitate more profound effects vis-à-vis Late replacements; this makes logical sense as discussed above. In our study, the Tables suggest that the ANN[Late] has more of a Smoothing-impact than does the ANN[Early]. Thus, there are likely to be MORE profound ANN-events occurring for the LHS IF one executes the ANN[Late]-protocol vis-à-vis the ANN[Early]-protocol. An investigative begged by this antithetical result to Hillmer's observation would be a valuable and welcomed addition to the research on outlier replacements.

Acknowledgments

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 were instrumental in this research. Further thanks are due to: Prof. Dr. H. Wright, *Boston University*: Department of Mathematics and Statistics, and the participants of the SUNY: SBE Workshop Series, in particular Prof. Dr. Kameliia Petrova: Dept. of Economics[Statistics] for their helpful comments and suggestions.

References

Adya, M. & Lusk, E. (2016). Time series complexity: The development and validation of a rule-based complexity scoring technique. *Decision Support Systems*. <https://doi.org.10.1016/j.dss.2015.12.009>

Aryal, U. & Khanal, K. (2013). Sharing the ideas of Meta - science to improve quality of research. *KUMJ*, *11*, 75-77.

Benjamin, D*., - - -, & Johnson, V. E. (2018) Redefine statistical significance. *Nat. Hum. Behav.*, *2*, 6–10.

*There are more than 30 individuals listed as authors.

Enders, C. (2010). *Applied missing data analysis*. Guilford Publications; Ltd. ISBN: 978-1-60623-639-0.

Fahey, L. (2019). Getting to insight: the value and use of small data. *Strategy & Leadership*, *47*, 27-33. <https://doi.org/10.1108/SL-03-2019-0034>

Fisher, Sir R.A. (1925). *Statistical methods for research workers*. Oliver & Boyd. ISBN 0-05-002170-2

Gaber, M. & Lusk, E. (2017). Analytical procedures phase of PCAOB audits: A note of caution in selecting the forecasting model. *J. Applied Finance and Accounting*, *4*, 76-84. <https://doi.org/10.11114/afa.v4i1.2811>

Hausman, J. (1978). Specification tests in econometrics, *Econometrica*, *46*, 1251–1272. <http://dx.doi.org/10.2307/1913827>.

Heilig, F. & Lusk, E. (2020). Forecasting confidence intervals: Sensitivity respecting panel-data point-value replacement protocols. *International J. Accounting and Finance Studies*, *3*, <http://dx.doi.org.10.22158/ijafs.v3n2p104>

Heilig, F. & Lusk, E. (2021). Investigation of the effects of simple outlier replacement protocols in forecasting analyses: Are they robust-in-utility? *IAR J Bus Mng*, *2*, 89-101. <https://iarconsortium.org/journal-info/IARJBM>

Heilig, F. & Lusk, E. (2021). Homomorphic relations vis-à-vis excel capture and fixed effects confidence intervals. SUNY: Working Paper: H&L2021.

Hillmer, S. (1984). Monitoring and adjusting forecasts in the presence of additive outliers. *Journal of Forecasting*, *3*, 205-215. https://doi.org/10.1002/for.398003020

Kim, J., Ahmed, K. & Ji, P. (2018). Significance testing in accounting research: A critical evaluation based on evidence. *Abacus*, *54*, 524-537. <https://doi.org.10.1111/abac.12141>

Lusk, E. & Halperin, M. (2016). Client risk calibration in PCAOB audits: An analytic procedures panel risk assignment protocol. *International Journal of Auditing Technology*, *2*, 1-21.

Lusk, E. (2017). Analytic Procedures: A holdback-vetting forecasting model. *Applied Finance and Accounting*, *3*, 65-74. <http://doi.org/10.11114/afa.v3i1.2139>

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*, *2*, 111-1153.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Sage Publications (Applied Social Research 1st Edition: Series[v.6]) ISBN: ISBN 13: 978-0803942462

SAS®. (2005). *Statistics and graphics guide: JMP[6]®*. SAS Institute Inc. ISBN 1-59047-816-9.

Tukey, J. (1977). *Exploratory data analysis*. Addison-Wesley, ISBN-13: 978-0201076165.

Wang, H. & Chow, S.-C. (2007). Sample size calculation for comparing proportions. Test for equality: Wiley encyclopedia of clinical trials. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780471462422.eoct005>

Appendix A

We offer here a dataset constructed to illustrate a Smoothing, Provoking and also thus the effect of both.

Table A1. Illustration of Outlier ANN-Modifications Points {9 & 13}.

Series	1	2	3	4	5	6	25	27	22	33	30	5	1	6	8
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Table A2. Demo-dataset See Figure A.

P-P I	1227.158	2500	1280.824	830.825	884.83	829.813	800.819	967.283	1064.809	1223.922	50	1530.307
P-P II	1227.158	1234.765	1280.824	830.825	884.83	829.813	800.819	967.283	1064.809	1223.922	1544.1	1530.307
ANN P	1227.158	1253.991	1280.824	830.825	884.83	829.813	800.819	967.283	1064.809	1223.922	1377.115	1530.307

The correlation of each vis-à-vis the time-index are: P-P I[–0.39] & P-P II[0.34] & ANN-P[0.28]

Appendix B

We randomly sampled 22 organizations from the Bloomberg Terminals in the John & Diana Conner's Finance Trading Lab at the SUNY:SBE: College at Plattsburgh. For each organization, we selected a Panel of yearly reported information starting 2005 through 2016. This created three forecasting Panels: {**LP**(n=12) & **MP**(n=9) & **SP**(n=6) }. This data is the same that was used in Heilig & Lusk (2020) and Heilig & Lusk (2021).

Table B1. Accrual Firms Tickers found on the BICS-Platform: Bloomberg Terminals.

6758JP	ACN	AIR	AXE	BA	BAE	CVS	EFX	HSY	HUM	HYS
JBLU	LMT	LUV	RAD	ROK	SIE GR	SNA	SPGI	SWK	UTX	WBA

Finally, we randomly sampled from the BBTs for firms that were in the upper 33% of their market capitalization as of 15 February 2021. These are labeled as BBT22. These 17-firms are noted in Table B2. Accrual frame YE: 2000 to 2020.

Table B2. BBT21, n=21.

6501JP	6758JP	BA	CI	DTE	F	FDX	RSDA	VOD
FP	LKOK	KR	HUM	GE	CAT	CVS	VOW	

---

[i] These models mesh well with the generating processes for firms in trading markets as implied by Lusk & Halperin (2016) where most all of the firms in their market accrual had Hausmann (1978)-p-values that suggested rejection of the Random-Effects model rationalizing the likelihood of the Fixed-Effect alternative that is often consistent with near-neighbor Panel-association. This being the case, the maximum likelihood election is to replace outliers or missing values using the simple average of the *nearest* neighbors.

[ii] The OLSR-95%Excel Capture Interval is defined as:
LL→[95%Lower Limit: Intercept – [95%Lower Limit: Slope] × [n +1]
UL→[95%Upper Limit: Intercept + [95%Upper Limit: Slope] × [n +1]
The Precision [PECI] is: [UL –LL] × 50%
This information is produced by the Excel Regression Platform. See Gaber and Lusk (2017) for a discussion of the Capture Intervals vis-à-vis that of Fixed Effects and Random Effect versions.

[iii] It is also the case that the ratios of the Confidence Intervals are identical with the ratios in R2. The only condition is that the size of the Panel is the same for the numerator and denominator for the ratios.

[iv] ***Weak*** in the sense that in the H&L studies there was not an inferential test against a dataset where there was not likely evidence of correlation or autocorrelation.

[v] The RRP is the same whether using the Excel Capture Intervals or the Standard Fixed Effects Confidence Intervals. This homomorphic relationship is discussed in Heilig & Lusk SUNY:Working Paper: H&L2021].