Original Paper

Forecasting Confidence Intervals: Sensitivity Respecting Panel-Data Point-Value Replacement Protocols

Frank Heilig[1] & Edward J. Lusk[2,3,4]

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

[2] Emeritus: [Accounting] School of Economics & Business, SUNY: Plattsburgh, Plattsburgh, USA

[3] [Statistics] The Wharton School, University of Pennsylvania, Pennsylvania, USA

[4] Chair: [Economics] International School of Management: Otto-von-Guericke, Magdeburg, Germany

Abstract

In the practice of Time Series [TS] forecasting there are very often situations where it is prudent to modify certain "outlier" values in the TS-Panel. A simple modification protocol is to replace selected TS-points by the Average of their adjacent Near-Neighbor-points [ANN]. Thus, a research question, not previously addressed and thus of interest, is: Are ANN-TS modifications **balanced**—*50% of the time provoking OLS-forecasting variation, thus reducing the predicative acuity of the 95% Confidence Intervals; and, by symmetry, 50% of the time smoothing resulting in more predicative resolution.* **Research Plan** *To address this question, we: (i) collected accounting information to be forecasted from firms on the Bloomberg™ terminals for Income Statement and Balance Sheet sensitive variables, (ii) formed three ANN-modifications, and (iii) computed 95% Confidence Intervals using the firm-Panel and the three modified Panels.* **Results** *Regarding the research question, surprisingly the ANN-replacement protocols were not balanced. In fact, about 2/3 of the time the ANN was smoothing in nature, and thus about 1/3 of the time the ANN-protocol provoked OLS-variation. We discuss the important implication of this result for forecasting in the economic context.*

Keywords

provoking, smoothing, OLS-variation, outliers

1. Introduction

The quality and utility of a forecast is dependent on the nature of the time series that is used to make the desired projections. This is hardly surprising and is discussed in detail by Hanke and Wichern (2003) that we used in our forecasting courses. However, this core-concept has promoted, and to some extent, is used as the logical and rational justification for the following pre-analysis forecasting data-"organization" protocol:

> *If there are data-points in the TS-panel that do not seem as they would*
> *contribute projections that would inform the decision-making process, then*
> *the analysist is justified in replacing them with more appropriate panel*
> *realizations.*

The usual reasons necessitating such a protocol are: (i) not infrequently there are download/importing errors in accruing data through e-links between the software, such as Excel™, and the Internet as filtered through ever-changing Browser-links, (ii) sometimes outliers in the TS are identified by outlier-screens, such the Tukey (1977) Box-Whiskers-Plot, and others presented in SAS (2014), and, by protocol, they are replaced, or (iii) the analyst decides that certain values in the TS are not likely to be representative and so may bias forecasting projections, and they are judgmentally replaced. Interestingly, this TS-modification *gestalt* seems to be intrinsic to the practice of forecasting; we have seen this in operation over the years being practiced by academic and senior forecasters alike. While one may justifiably quibble with "Judgmental or Expert Opinion" adjustment of TS-points, it is not infrequently the case, in the e-download world, that datasets are corrupted and there are missing data-points [MO] or Additive-Errors [AO] where, in the process of creating the value of a particular data-point, there was a value [ϵ], of a non-trivial magnitude, that was inadvertently "added" to the value reported in the download. This is often noted as:

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

Additionally, when there are Event-Anomalies, such as the 2008 *Lehman Bros. LLP* Sub-prime debacle or the COVID-19 Pandemic, a case can be made for excluding or adjusting for such perturbations. These are sometimes called Level Shifts [LS]. This is a slight misnomer as most LS are temporal spikes or mini-plateaus the effects of which dissipate over time as, for example, a "V-Bound". A true LS is a Step-Level change that establishes a new baseline for the time series. As a point of information, distinguishing between a temporal and permanent LS is fraught with difficulty. For most of the literature, MO, AO, & LS are subsumed under the rubric of Outliers.

As there seems to be: (i) a linguistic behavioral "imperative" for analysts to correct for outliers, (ii) a plethora of outlier screening platforms, the use of which is synonymous with best practices in creating meaningful forecasts and/or data-analytic recommendations, and (iii) analytic-platforms that have output conditioning platforms to adjust for outliers, we should expect that a study that addresses the impact replacement of outliers would be of practical value.

This is the departure point of our research report the focus of which is to:

1) Review the ***Literature*** on Outliers to give focus to our study,

2) Present the ***TS-Forecasting Model*** that we will use to form the forecasts and the related 95% Confidence Intervals,

3) Detail the ***TS-Replacement Protocols*** used to create comparative information,

4) Discuss the ***Accrual of the Firms*** and their sensitive account Panel Accounting variables that are used to develop the TS-forecasts,

5) Detail variable measures sensitive to: ***Forecast & Precision Displacement*** and their ***Impact-Context***,

6) Present detailed ***Tabular Profiles*** of these study impact variables and selected ***Inferential Illustrations*** that can be used to examine aspects of the modifications that would inform their decision needs, and

7) Offer a ***Summary*** and an ***Extension*** of this study.

2. Literature Review

The groundbreaking work of G.E.P. Box and G. Jenkins circa 1976 was the watershed event for the proliferation of research on outliers in the projective-modeling context. One of the research reports at the inception of this study domain was Hillmer (1984). Hillmer examined AO in a simple autoregressive context ARIMA (1,0,0) and also for the Seasonal [B=12] ARIMA-model. His results focused on model "in-adequacy" issues that could be detected by monitoring the one-period-ahead forecasts. In this context, Hillmer used an example presented by Box and Jenkins to illustrate his model-correction protocol for AOs. Hillmer's study suggests that AOs affect the confidence intervals and the forecasts. He observes that the dominant effect is for AOs that occur early in the TS rather than later. Since Hillmer's publication, there have been a number of derivative studies addressing generalized outliers in an ARIMA-context where identification and correction protocols have been researched; usually, these protocols are demonstrated and evaluated using simulations. As these results are along the same lines as reported by Hillmer, we only note the key references. The following are two examples of studies separated by about 10-years. The first is Chen and Liu (1993). They consider four types of outliers/errors in the ARIMA-context: AO, Innovation Outliers (IO), Temporal Change (TC), and Level Shift (LS). For their detection/correction protocol, they offer that judgmental methods aided by statistical methods may be needed to identify which of these four cases is most likely to best characterize the issue under examination. About 10-years later, Hotta, Pereira-Valls and Ota (2004) using simulations addressed ARIMA-error detection and the resulting effect in a design blocked on Disaggregated and the related Aggregated TS-data. They report various effects between the Disaggregated- and the related Aggregated-arms. Interestingly, research reports dealing with ARIMA-models and related Panels used in detection and correction of outliers have waned in the last few years. We did a search on *ProQuest*™: *ABI/INFORM*™ Global[Index] for the last five years for:

106

{*Abstract[ARIMA & Outlier Correction] AND/OR Abstract[Forecasting]*} and found only two indexed research Reports: Rim, Park, Oh, Park and Lee (2016) and Mélard (2016) both of which provide confirmatory information and extensions of the Hillmer study.

3. The Forecasting Model

To generate a forecast, $\hat{Y}_t$, we will use the *Excel*™ Platformv.2016 [*DataAnaysis*[Regression]]. This form will be the OLS-two parameter linear [Intercept & Slope[Trend]] Time Series Regression model [OLSR] used to forecast from a TS-Panel. In the TS-version, the dependent variable is the traditional Y-variate [Ordinate] and the integer Time Index is the equally-spaced independent-variate [Abscissa] where the first point $X_{t=0}$ is indexed as 1 and the points follow as integers to the last point all of which are used in the OLS-fitting.

3.1 OLSR Inference from the Excel Parameter Range Model

The Excel Regression functionality forms a "wide-covering" confidence interval. These 95%CIs are effectively extreme case CI-scenarios as they are produced from the two crisp-end-point parameters of the 95%CI for the intercept and the slope jointly. See Gaber and Lusk (2017). For this reason, we are NOT interested in the capture-rate of these 95%CI; almost certainly, these wide-95%CIs will capture most of the one-period-ahead holdbacks. We are creating these confidence intervals so as to measure the effect of the TS-data-point replacements on the Forecast and the related Precision. Following, we offer the standard notation:

Extreme Left Side [Lower-Limit [LL]] 95% Boundary:

$$\hat{Y}_{LowerLimit:(t=(N+h))} = \hat{\alpha}_N - [\, t_{\frac{\alpha}{2}} \times s_{\epsilon\alpha} \,] + [\hat{\beta}_N - [\, t_{\frac{\alpha}{2}} \times s_{\epsilon\beta} \,]] \times [N + h] \qquad (1)$$

Extreme Right Side [Upper-Limit [UL]] 95% Boundary:

$$\hat{Y}_{UpperLimit:(t=(N+h))} = \hat{\alpha}_N + [\, t_{\frac{\alpha}{2}} \times s_{\epsilon\alpha} \,] + [\hat{\beta}_N + [\, t_{\frac{\alpha}{2}} \times s_{\epsilon\beta} \,]] \times [N + h] \qquad (2)$$

Where: $\hat{\alpha}_N$ & $\hat{\beta}_N$ are the intercept & slope of the OLSR; $s_{\epsilon(.)}$ are the standard errors, $t_{\frac{\alpha}{2}}$ is the t-statistic for inference for the two-tailed 95%CI that has df = [N-2] computed in Excel as: T.INV.2T(5%,(N-2)); h=1 and is the forecasting horizon; N is the last time index in the data-stream and also the number of TS-points used to fit the OLSR.

3.2 Instructive Illustration

It is usual to script an illustration to clarify these computations. This will be done following using the data in Table 1.

SONY *Current Ratio* Panel [2005:2016] for the Balance Sheet. In this case, we downloaded Balance Sheet data for SONY. The SONY-account panel selected was the yearly closing stock price reported on the Bloomberg Market Navigation Platform for SONY as noted in Table 1.

Table 1. SONY: [6758JP-CUR_RATIO]

1.304	1.295	1.128	1.266	1.178	1.280
1.245	0.950	1.018	0.931	0.829	0.845

For a one-period ahead forecast, h=1, we produced the following information:

$$\text{Forecast } \hat{Y}_{(t=(N+h))} = \hat{\alpha}_N + [\hat{\beta}_N \times [N + h]] \tag{3}$$

$$\mathbf{0.82} = \hat{Y}_{(t=(N+1))} = [1.39 - [0.044 \times [12 + 1]]] \tag{3.a}$$

The TS-version of the 95%CIs for the HSY-dataset are:

Extreme Left Side [Lower-Limit [LL]] 95% Boundary for the first projection horizon [h=1] is:

$$\hat{Y}_{LowerLimit:(t=(N+1))} = [1.39 - [2.3 \times 0.0573]] + [-0.044 - [2.3 \times 0.0078]] \times [12 + 1] \tag{4}$$

$$\hat{Y}_{LowerLimit:(t=(N+1))} = \mathbf{0.467} \tag{4.a}$$

Extreme Right Side [Upper-Limit [UL]] 95% Boundary for the first projection horizon [H1] is:

$$\hat{Y}_{UpperLimit(t=(N+1))} = [1.39 + [2.3 \times 0.0573]] + [-0.044 + [2.3 \times 0.0078]] \times [12 + 1] \tag{5}$$

$$\hat{Y}_{UpperLimit:(t=(N+1))} = \mathbf{1.173} \tag{5.a}$$

Using the usual definition of precision—i.e., 50% of the spanning-length of the confidence interval—the precision of this forecast is: 50% × [1.173 − 0.467] = 0.353. Note, as a computation-check, that the mid-point of the confidence interval is the forecast—e.g., [[0.467 + 1.17] × 50%] = **0.82**.

4. Modification Protocols: The Account Base

4.1 Accounting Variable Set for Forecasting

Table 2. Selected Accounts for Each of the Accrual Firms

The firms were randomly selected from the Bloomberg™ Terminals [BBT]; see Appendix A. For each firm, we selected 12 data-points, [2005:t_1] through [2016:t_{12}], from the following Income Statement [IS] and Balance Sheet [BS] accounts.

Income [IS] and Balance Sheet [BS]	Bloomberg Variable Designations
IS: Gross Profit	GROSS_PROFIT
IS: Operating Income	IS_OPER_INC
IS: Earnings for the Common Shareholders	EARN_FOR_COMMON

IS: Shares for Diluted Earnings per Share	IS_SH_FOR_DILUTED_EPS
BS: Current Assets[CA]	BS_CUR_ASSET_REPORT
BS: Other Assets: Deferred Charges	BS_OTHER_ASSETS_DEF_CHRG_OTHER
BS: Current Liabilities[CL]	BS_CUR_LIAB
BS: Current Ratio [CA/CL]	CUR_RATIO

For each of these eight accounts there were four Panels created defined following:

4.2 Panel Modifications

Panel: No modifications: Baseline Panel [*BP*]. The Panel is used as downloaded and noted as: [*BP*]: $\{x_{t=1}, - - -, x_{t=12}\}$.

Panel Modification Replacement Protocols: In the literature review there were no studies that addressed replacement of outliers or missing values using the Average of the Nearest Neighbor points: [ANN] as the replacement protocol. This is interesting as, in our experience, the ANN-protocol is (i) Intuitive in an autocorrelation context in particular for ARIMA(1,d,0) or the ARIMA(0,2,2)/Holt processes given a Fixed Effects generating process as is the case for most traded organizations. See Lusk and Halperin (2016), (ii) it is simple to program using non-VBA Excel functionalities, (iii) there are no "seasonal" blending issues as seasonality is effectively hidden or smoothed by a yearly-index, (iv) it avoids the temptation to make "judgmental" replacements which may offer dysfunctional data-creation or engineering possibilities, and (v) the ANN seems to be non-differentiable—of the same ilk—from many of the other "simple" replacement protocols such as Median or full-Panel averages. Following the various datasets and modifications are detailed.

First, we did not modify the downloaded dataset of the **BP**. For the OLSR-fit, we use all of the Panel data. The **BP** is the benchmark for the ANN-modifications. For each firm & account dataset, we made the following three ANN replacements.

Both: Here we made two ANN replacements: The second Panel point and the next to last Panel point.

Late: Here we made one ANN replacement: The next to last Panel point.

Early: Here we made one ANN replacements: The second Panel point.

For example, we have the **BP**, n =12, and:

$$\textit{Both} \text{ Panel}[\textit{Early}\&\textit{Late}]: \left\{x_{t=1}, \left[\frac{[x_{t=1}+x_{t=3}]}{2}\right], - - -, x_{t=10}, \left[\frac{[x_{t=10}+x_{t=12}]}{2}\right], x_{t=12}\right\} \quad (6)$$

$$\textit{Late} \text{ Panel}[x_{t=11}]: \left\{x_{t=1}, - - -, x_{t=10}, \left[\frac{[x_{t=10}+x_{t=12}]}{2}\right], x_{t=12}\right\} \quad (7)$$

$$\textit{Early} \text{ Panel}[x_{t=2}]: \left\{x_{t=1}, \left[\frac{[x_{t=1}+x_{t=3}]}{2}\right], - - -, x_{t=12}\right\} \quad (8)$$

4.3 Nature of the Replacement Protocols: Smoothing or Provoking Variation

The ANN will, of course, have an effect on the nature of the TS depending on the relationship between the TS-value to be replaced by the ANN-Protocol. The question is: *What is that effect likely to be?*

The ANN-effect is usually measured by the <u>change</u> in the Standard Error [$\sqrt{MSE}$], where: the MSE is the Mean Squared Error of the residuals of the OLSR-Fit of the TS as the denominator in ratio to that of the ANN-modified series as the numerator; note this ratio change as: $\Delta\hat{\sigma}_e$. When $\Delta\hat{\sigma}_e > 1.0$, the ANN-modification created more scaled OLSR-variation than was the case for the original TS: in this case, the ANN-modification is labeled ***Provoking***. When $\Delta\hat{\sigma}_e < 1.0$, the ANN-modification created less scaled variation than was the case for the original TS: in this case, the ANN-modification is labeled: ***Smoothing***. The last case is: $\Delta\hat{\sigma}_e = 1.0$, and the ANN-modification is labeled: ***Neutral***. We are interested in the $\Delta\hat{\sigma}_e$ as this ratio is also the relative ratio of the precision of the TS to any of the ANN-modifications. Thus, Provoking modifications will create a wider Confidence Interval; and Smoothing will create a smaller Confidence Interval. An illustration will be most helpful at this point. Assume that we have the SONY series as presented in Table 3 that presents all the relative measures for the TS and the three ANN-modifications:

Table 3. SONY ANN-Protocol Profiles

Series	$\sqrt{MSE}$	95%CI LowerLimit	95%CI UpperLimit	Precision 95%CIs	$\sqrt{MSE}$ Ratio	Precision CI Ratio
SONY:TS	0.093074	0.467	1.173	0.353	N/A	N/A
*Both***[EQ6]**	0.092573	0.495	1.198	0.351	0.995[S*]	0.995[S]
*Late***[EQ7]**	0.089458	0.497	1.176	0.339	0.961[S]	0.961[S]
*Early***[EQ8]**	0.096354	0.464	1.195	0.366	1.035[P]	1.035[P]

***P** represents Provoking & **S** represents Smoothing

In the SONY case, as a graphical elucidation, all of the four instances presented in Table 3 are also profiled in Figure 1.
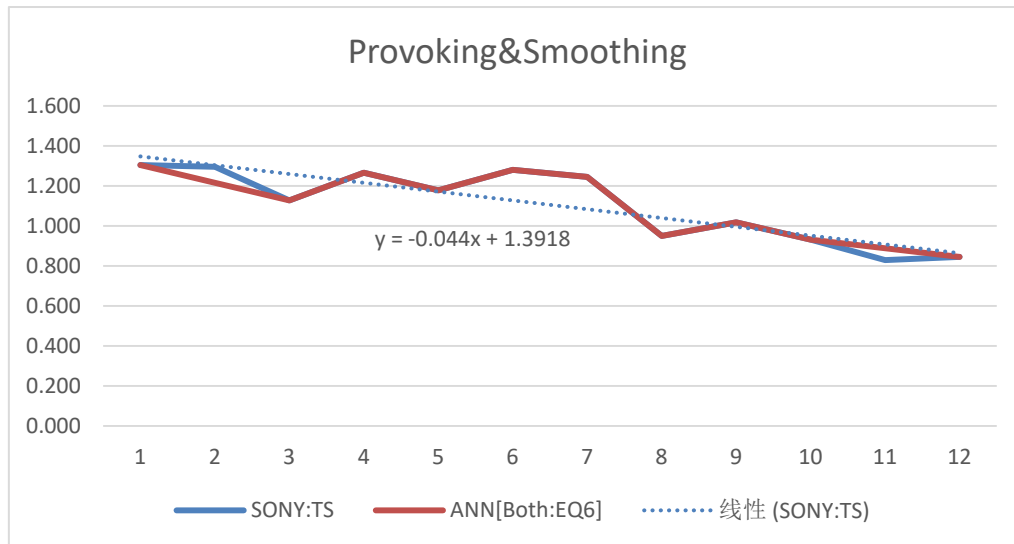
Figure 1. Panel Plot of SONY Contrasting the ANN-Both Modification

The OLSR line[- - -]—i.e., EQ3a and the three ANN-modifications are found on Figure A. The ANN-*Early*[EQ8] modification produces the triangle between Points:[1&3]; The ANN-*Late*[EQ7] modification produces the triangle between Points:[10&12]. Thus, the ANN-*Both*[EQ6] modification is where the *Early* & *Late* are effected. This is an excellent way to inspect the Provoking or Smoothing nature of ANN-Modifications. The ANN-*Early* Changes TS-point: 1.295 to 1.216 [(1.304 + 1.128)/2]; and the ANN-*Late* Changes TS-point: 0.829 to 0.889 [(0.931 + 0.845)/2]. Using the SONY-TS regression line as the benchmark, the ANN-*Early* modification moves TS-point, t = 2, "away from the SONY-TS regression-line". Using the information in Table 3, the OLSR: $\sqrt{MSE}$ for the ANN-*Early*-TS will be higher given the modification created by the ANN-*Early* point modification as it is further away from the SONY:TS line[Point]. Thus the $\Delta\hat{\sigma}_e$ will be: 1.035 [0.096 / 0.093]. As $\Delta\hat{\sigma}_e$ is > 1.0 this means that the precision of the 95%CI of the ANN-*Early* will be > than that of the SONY:TS. This then means that the width 95%CI of the ANN-*Early* will be wider than that of the SONY:TS. This is what we see in Table 3; also this is consistent with Hillmer's research report. Specifically, the width of the 95%CI of the SONY-TS is 0.706 [1.173 − 0.467], whereas, the width of the 95%CI of the ANN-*Early* 0.731 [1.195 − 0.464]; this of course means that the precision which is 50% of the width of the Confidence Interval will follow: i.e., Precision of the SONY-TS is 0.353 and is less than that of the ANN-*Early* that is: 0.366. Thus Figure A and Table 3 indicate that relative to the SONY-TS benchmark there were two Smoothing Modifications: [ANN-*Both* and ANN-*Late*] and one Provoking: the [ANN-*Early*]. *Point of Information* Also of interest: it is the case that even though the ANN-*Both* is composed of the ANN-*Early* & the ANN-*Late* and the ANN-*Early* is Provoking and the ANN-*Late* is Smoothing, taken together—as the ANN-*Both*—they produce a Smoothing Impact relative to the SONY-benchmark. This is not an anomaly; it just means that the OLSR[*Both*] is reoriented by the aggregation of the ANN-*Early* and the ANN-*Late* so that the OLSR[$\sqrt{MSE}$] for the

111

Both-TS that is: 0.092573 compared to the OLSR[$\sqrt{MSE}$] for the SONY-TS that is: 0.093074 produce a ratio that is < 1.0—thus Smoothing: 1.0 > 0.994611 [0.092573 / 0.093074].

5. Hypothesis and Results

The important questions herein cast as: **Hypothetical Conjectures [HC]**, which with we are preoccupied, are the subject of this research report are:

For the ANN-modifications **scripted in the FPE: Null** form:

1) **HC[1]** *The first level of analysis is **Overall***. In this aggregated case, the FPE:Null will be that the ANN-modifications, in the aggregate—<u>overall</u>, produce an equal percentage of: {Provoking & Smoothing} impacts for the accrual Panel-dataset.

2) **HC[2]** A **Logical Disaggregation** is to determine if there are differences individually for the Provoking- and then for the Smoothing-Arms for <u>each</u> of the three ANN-modifications: {**Both**, **Late** & **Early**}. In this case, the FPE:Null will assume that <u>each</u> of three ANN-modifications produce $\Delta\hat{\sigma}_e$ measures that are inferential identical for the Provoking events **and then for the Smoothing Events**.

3) **HC[3]** Finally, a **Judgmental Disaggregation** of interest will be to create a sub-set for each of the three ANN-modifications: {**Both**, **Late** & **Early**}labeled as <u>Serious</u> using the $\Delta\hat{\sigma}_e$ measure. The operative FPE:Null will be for <u>each</u> of these three ANN-modifications that the percentage of Provoking and Smoothing will not be different. *Clarification* In this context, based upon experiential judgement, we proffer that if $\Delta\hat{\sigma}_e$ is < 0.975 in the Smoothing direction **OR** if $\Delta\hat{\sigma}_e$ is > 1.025 in the Provoking direction that the ANN-impact should be labeled as Serious. This screening window of ±2.5% for the $\Delta\hat{\sigma}_e$ measure is essentially trimming the ANN-impact dataset that is created if $\Delta\hat{\sigma}_e$ is ≠ 1.0 and thus is used to count the number of these instances. This is in form the protocol used by Makridakis, Hibon, Lusk and Belhadjali (1987).

To form the usual inferential information, we will proffer an *a priori* context for addressing the testing of these three research questions *Testing Context* Mathematically, the ANN-modifications are neutral point blends—i.e., nearest neighbor or adjacent point-averages. Usually, a Panel of firm accounting data traded on exchanges is characterized by: (i) a well-defined exogenous random-stochastic variation, usually termed white-noise, and (ii) a endogenous value-generating process that is usually an ARIMA process no more complicated than an ARIMA(1,0,0) or an ARIMA(0,2,2)—simply: *autocorrelation with a relatively small random perturbation*. In this configuration, an illustrative expectation that characteristically mimics a traded organization's Panel is a sinusoidal function. See Appendix B. Using an OLSR-fit and applying the ANN-protocol **over the entire Panel in Appendix B** it is clear that <u>most often</u> these 16-ANN-modification move the modified-points toward the adjusted-OLSR-line; thus relative to the benchmark-TS, this results in Smoothing as these are mostly $\Delta\hat{\sigma}_e$ < 1.0. For example, for the data in Appendix B panel the $\Delta\hat{\sigma}_e$ is: 0.59 [0.45/0.76]—a reduction of about 40%. This

112

strongly indicates that overall the ANN-protocol applied over the TS resulted in a major smoothing. *Pedagogic Note* It is instructive for a class presentation to observe that most all of the ANN-modifications move the sinusoidal TS-Panel points towards the basic OLSR-line. Thus, this simple sinusoid-mimicking-graphic, suggests that there is likely a tendency for the ANN-modifications to be more often Smoothing in nature than Provoking. This can thus be formed into an operational hypothesis expectation so as to give an *a priori* context to the three questions above.

6. Inferential and Exploratory Results

To provide a complete rendering of the testing profiles, we have created Table 4. With this results profile, we will offer parametric and, in the service of robustness, non-parametric alternatives to inferentially probe this dataset. Given the nature of this research, as expressed in the *HC*, the testing will be a blend of *a priori* testing—to wit: *HC[1]* and exploratory testing: *HC[2]* & *HC[3]*. For example, we have *no pre-conceptions* as how the relative tests: *Both*, *Late* or *Early* as blocked by Provoking & Smoothing, or how the *Seriousness* screening-partition will play out. Thus, these are not *a priori* tests but rather exploratory in our context and this experiential exploratory profile can be used subsequently to form *a priori* tests.

Table 4. Accrual Profiles over the ANN Modifications

	Provoking[P]	Smoothing[S]	Total[P+S]
Serious[Both]	**12**	**79**	**91**
Serious[Late]	**10**	**51**	**61**
Serious[Early]	**6**	**43**	**49**
ANN[Both]	*49*	*114*	163
ANN[Late]	*61*	*102*	163
ANN[Early]	*64*	*98*	162
ANN[Overall]	**174**	**314**	**488**

6.1 Consideration of HC[1] That Examines the Overall Binary Context: Provoking v. Smoothing

The basic test for Question **HC[1]** is for the FPE: Null—that is of a {50%: 50%}–balance: Provoking & Smoothing. The simplest appropriate test is to form the 95%CI for the Null. In this case, the Expectation-95%CI is: [45.6%: 54.4%]: $[50\% \pm 1.96 \times [50\% \times 50\% /488]^{0.5}]$. The Actual result is that the Smoothing Percentage is: 64.3% [314/488]. Clearly, this is outside the Expectation-95%CI on

the relevant RHS. To reinforce this result: the z-value—using 50% in the standard error computation with no continuity correction—is: 6.3; this produces a directional P-value of < 0.0001. Both of these results indicate that rejection of the Null is the obvious and so logical inferential choice in favor of the likelihood that the ANN-modifications are biased/pre-disposed to Smoothing for Accounting Firm data from traded organizations. The inference, using the counts, is the correct inferential context. However, as additional confirmatory or vetting information, following are the Means & Medians of the $\Delta\hat{\sigma}_e$s and the related exploratory tests.

Table 5. OLS Profiles for the *a posteriori* Classified Provoking & Smoothing Effects

Coded Partition	$\Delta\hat{\sigma}_e$s Mean: Median
Provoking: n=174	1.014 : 1.010
Smoothing: n = 314	0.939 : 0.972

Discussion For Table 5, the overall ANOVA-Welch test has a P-value of <0.0001 and an overall Wilcoxon/Kruskal-Wallis RankSums [W:K-W] P-value of < 0.0001. The *a posteriori* Power was > 95%. This is expected given the fact that the ANN-Impacts are labeled/partitioned according to their magnitudes relative to $\Delta\hat{\sigma}_e$. **Summary HC[1]:** *The operative Null scripted in HC[1] is rejected in favor of: There is a clear difference between the number of **Smoothing** Impacts relative to the number of **Provoking** Impacts where: there were more that were Smoothing in nature*. This follows, in direction, the conjecture that we offered relative to the Sinusoidal mimicking-function in Appendix B.

6.2 Consideration of HC[2] That Examines the ANN-Impact over the Three ANN-modifications: [Both, Late & Early] for Provoking & Smoothing

We tested **HC[2]** using parametric and non-parametric tests individually blocked by Provoking and Smoothing. Specifically, we will be using the cells in Table 4 the numbers of which are in *Italics*. For the Provoking tests the $\Delta\hat{\sigma}_e$s are presented in Table 6 following:

Table 6. Provoking Performance Profile

ANN-Feature	$\Delta\hat{\sigma}_e$ Mean : Median
*Both***[EQ6]***n=49*	1.017 : 1.013
*Late***[EQ7]** *n=61*	1.010 : 1.010
*Early***[EQ8]** *n=64*	1.014 : 1.011

114

The overall ANOVA-Welch test has a P-value of 0.02 and an overall Wilcoxon/Kruskal-Wallis RankSums [W:K-W] P-value of 0.21; as for the pairwise comparisons: only the {**Both** v. **Late**} has a [W:K-W] P-value < 0.1 and a Tukey-Kramer HSD: MCT [T-K] P-value < 0.03. In the case of Provoking, the ANN-Impact is identified for **Both** and is only different in $\Delta\hat{\sigma}_e$ from the **Late** for a P-value < 0.1. The *a posteriori* Power was 64.1%%

For the Smoothing tests, the $\Delta\hat{\sigma}_e$s are presented in Table 7 following:

Table 7. Smoothing Performance Profile

ANN-Feature	$\Delta\hat{\sigma}_e$ Mean : Median
*Both***[EQ6]** *n=114*	0.918 : 0.944
*Late***[EQ7]** *n=102*	0.944 : 0.976
*Early***[EQ8]** *n=98*	0.958 : 0.982

Discussion For Table 7, the overall ANOVA-Welch test has a P-value of 0.002 and an overall Wilcoxon/Kruskal-Wallis RankSums [W:K-W] P-value of < 0.0001; as for the pairwise comparisons: the {**Both** v. [**Late** & **Early**]}: has [W:K-W] P-values < 0.002 and Tukey-Kramer HSD: MCT [T-K] P-values < 0.09. The *a posteriori* Power was 83.7%

Summary[H2]: *For the Provoking- and Smoothing-arms the **Both**-ANN modification is often found to be inferentially different in the measure: $\Delta\hat{\sigma}_e$. Specifically, for the Provoking-arm, **Both** only differs from the **Late**; but for the Smoothing-arm, **Both** differs from the: {**Late** and the **Early**}. This make common sense as the **Both** is composed of the **Late** and the **Early**, so it is not surprising that the **Both** has a more significant directional ANN-impact.*

6.3 Consideration of HC[3] That Deals with Number the Scaled Serious: $\Delta\hat{\sigma}_e$ *Impacts*

The simplest test is to benchmark one Impact-set by the other. The ideal test, in this case, is to create inferential information using the Chi2 analysis. As in the above analysis, this will create an overall inferential measure and if the overall Chi2-infernetial indication suggests rejecting the Null of no relative proportional differences, we will explore this fact. The Provoking & Smoothing Chi2 Table is the sub-matrix that has the counts **Bolded i**n Table 4. This sub-matrix is [3[B;L&E] × 2[P&S]. The overall Chi2 measure has a Pearson P-value of 0.79. This is a strong indication that the Provoking and Smoothing sets are not sufficiently different in proportions to indicate that they are drawn from different populations. This is clear as the column proportions are basically the same. For example, for the **Both**-arm the Provoking column average is: 42.9% [12/[12 + 10 + 6]] and for Smoothing is: 45.7% [79/[79 + 51 + 43]].

115

Summary HC[3] Thus there is no evidence that the magnitude of the ANN-impact differs as between the Provoking- or Smoothing-arm. This effectively is just a test of the symmetry of the ANN-impact. Simply, the ANN-protocol affected the $\Delta\hat{\sigma}_e$ in a similar manner independent of whether $\Delta\hat{\sigma}_e$ was < 1.0 or > 1.0. However, it is also of interest that the percentage of serious ANN-Impacts that are Provoking & Serious in nature is: 16.1% [28/174] compared to those that are Smoothing & Serious that is: 55.1% [173/314]; this may just reflect the Smoothing pre-disposition.

7. **Summary and Outlook**

7.1 Summary

We have clear evidence that replacing points in the Time series of firms traded on exchanges can have an effect of the OLSR-information used in the forecasting context. This effect seems to be systematic. There are more instances that the simple ANN-protocol, that is very often the outlier replacement protocol of choice for forecasters, will create Smoothing effects and so result in Confidence Intervals[CI] that will exhibit "more informed CI—in that they have better precision—i.e., are more narrow CIs". We observe this about 2/3 of the time—64.3% in our test. Further, the ***Both***-ANN modifications, as it is composed of the ***Late*** and ***Early*** seems, as expected, to have more of an impact. Finally, although there was not a proportional difference between the percentage of Serious ANN-effects as between the Provoking- and Smoothing-arms, there was a likely important difference in the percentage of serious ANN-modifications between the Provoking- and Smoothing-arms [overall: 16.1% v.55.1%] that may be a variant of the overall Smoothing tendency—we have not tested this "mollification". The summary ***take-away*** message is:

> *The profile and analysis of the ANN-protocol tested in the accrual-context of accounting data for market-traded organizations Re: forecasting acuity suggest that replacing Panel-points has a non-trivial impact on the $\Delta\hat{\sigma}_e$. This manifests itself overall as well as in the dis-aggregations tested both in the **number** of Smoothing instances relative to the Provoking frequency and also in their **magnitude**. Thus, the ANN-effects need to be considered in electing to replace data in the Firm Panels of Accounting data.*

7.2 Outlook

Further, investigations of a simple way to anticipate the nature of the effects: {Provoking or Smoothing} & {Both, Late or Early} & {Serious or Neutral} are follow-up studies begged by our results. Also, it may be useful to determine if the inherent variability of the Accounting Panels is related to these results. In this case, researchers may select from NAICS groups that have varying levels of variability—such as Utilities v. Software Development firms to test the replacement protocols as conditioned by the inherent nature of the variability of the forecasting context. Finally, it may be interesting to determine if these effects are related to the length of the Panel. The research report of Adya and Lusk (2016) suggests that in the Rule Based Forecasting context Panel Lengths smaller than about 13 seems to be where RAE

instabilities are in the offing. Thus controlling for Panel length may provide interesting test results. Finally, we have selected the ANN-modification as this is the most basic and, in our experience, the most often used when outliers are a foot. At the other end of the spectrum are the Missing Data Analyses detailed by Enders (2010). These are relevant if the rigorous assumptions as to causality are likely to be the generating process(es). In any case, it would be productive to research the effect of these "Enders"-reconstitution protocols on the CIs and the related smoothing and provoking profiles.

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

Box, G. E. P., & Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*. Holden Day ISBN13: 9780816211043.

Chen, C., & Liu, L.-M. (1993). Forecasting time series with outliers. *J. of Forecasting*, *12*, 13-35. https://doi.org/10.1002/for.3980120103

Enders, C. (2010). *Applied Missing Data Analysis*. Guilford Publications; Ltd. ISBN: 978-1-60623-639-0.

Gaber, M. & Lusk, E. (2017). Analytical Procedures Phase of PCAOB Audits: A Note of Caution in Selecting the Forecasting Model. *Journal of Applied Finance and Accounting*, *4*, 76-84. https://doi.org/10.11114/afa.v4i1.2811

Hanke, J., & Wichern, D. (2003). *Business Forecasting*. Upper Saddle River:NJ:USA Pearson: Prentice Hall**.** ISBN-13: 978-0132301206.

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.3980030208

Hotta. L., Pereira-Valls, P., & Ota, R. (2004). Effect of outliers on forecasting temporally aggregated flow variables. *Sociedad de Estadística e Investigatión Operativa Test*, *13*, 371-402. https://doi.org/10.1007/BF02595778

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. https://doi.org/10.1504/IJAUDIT.2016.078167

Makridakis, S., Hibon. M., Lusk, E., & Balhadjali, M. (1987). Confidence intervals: An empirical investigation of the series in the M-competition. *International J. Forecasting*, *3*, 489-508. https://doi.org/10.1016/0169-2070(87)90045-8

Mélard, G. (2016). On some remarks about SEATS signal extraction. *SERIEs: Springer[Heidelberg]*, *7*, 53-64. https://doi.org/10.1007/s13209-015-0137-y

117

Rim, H., Park, S., Oh, C., Park, J., & Lee. G. (2016). Application of locally weighted regression-based approach in correcting erroneous individual vehicle speed data. *J. Advanced Transportation*, *50*, 180-196. https://doi.org/10.1002/atr.1325

SAS®. (2014). *Econometrics and Time Series Analyses 2 for JMP®*. SAS Institute Inc. ISBN 978-1-61290-586-0.

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

Appendix A

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. The Trading Tickers for this study are in the following Table A1. For each organization, we selected a Panel of yearly reported information starting 2005 through 2016 or a time series [TS] of 12 values.

Table A1. Accrual Firm's 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

Appendix B

Here is the dataset that is not atypical for a traded organization. The Excel-generating form for this data is: *Sin*(i) + i + (RANDBETWEEN(-10,10)/100); I = {1, - - -, 18}.

Table B1. Sinusoid Projection

1.92	3.00	3.24	3.15	4.11	5.79	7.60	9.09	9.42
9.46	9.96	11.47	13.34	15.00	15.69	15.69	16.05	17.27