

OLS-Regression Forecasting Confidence Intervals Capture Rates: Precision Profiling in the Forecasting Model Selection Process

Frank Heilig¹ & Edward J. Lusk²

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

²School of Economics & Business, State University of New York at Plattsburgh, Plattsburgh, USA

Correspondence: Edward J. Lusk, Department of Accounting: School of Business and Economics, State University of New York (SUNY) at Plattsburgh, 101 Broad St., Plattsburgh, NY, USA.

Received: January 30, 2020

Accepted: February 25, 2020

Online Published: March 6, 2020

doi:10.5539/ibr.v13n4p14

URL: <https://doi.org/10.5539/ibr.v13n4p14>

Abstract

Forecasting creates projections into an uncertain future. To understand the decision-making implications of the forecast, confidence intervals [CIs] are required. This seems simple enough. However, considering the details of the computations that underlie generating forecasting CIs, one encounters a practical disconnect that can create a decision-making paradox. In the OLS 2-parameter linear regression [OLSR] case, there are two forecasting Models that are often employed: {The Uni-variate Time Series & The Standard two-variable Regression Y:X}. Further, for each of these two Models individually, there are three (1-FPE[α]) %CIs Versions: {Excel, Fixed Effects & Random Effects} each of which is oriented around the same OLSR-forecast value. Given this component configuration, a paradox emerges because each of the forecasting models, {TS or Y:X}, individually produces a forecast with a markedly difference precision profile over the three CI-Versions. In our experience, this is paradoxical as forecasters assume that as the forecasts are the same in each model-class, their Capture Rate—the percentage of time that the actual future values are IN the CIs—should also be the same. To address this seeming paradox, we develop, detail, and illustrate a two-stage OLSR Decomposition and Screening protocol, termed: the [D&S-Triage] protocol that has the following components: (i) Stage A: decomposition of the Model & Version factor-sets to better understand the implications of their Precision differences, and (ii) Stage B: focusing on inferentially significant forecasting model components, create a multilevel quality-algorithm to identify a forecasting model-set that addresses the Quality of the Capture Rate that are the best in their class.

Keywords: balanced scorecard, fixed effects, OLS, random effects

1. Introduction: Precision Profiles as an Inferential Screen in the Forecasting Context

To introduce the irresistible “innate and sometime inane” human disposition to seek a means of fore-telling occurrences, Lusk (2019a) discusses: previsions/ soothsaying/ foretelling/ conjectures/ visions that date from antiquity—*Delphi and the Oracles* being the prototypical example. These ethereal methodologies have given way to statistical and mathematical methodologies that seem to have their genesis and growth-to-coalescence in the 19th century. The luminaires of this wave of change from the *Smoke & Mirrors* of antiquity to reproducible scientific protocols were: Karl Pearson (1857-1936), Sir Francis Galton (1822-1911) and Sir R.A Fisher (1890 -1962). They effectively developed most of the basics that underlie all of parametric statistical analyses that are integral to most Decision Support Systems [DSS] currently in use including: Deep BlueTMⁱ and the now ubiquitous WatsonTMⁱⁱ.

As wondrous though DSS may be as informing the decision-making process there is a contextual drawback. The “glitch”, in the DSS-milieu, is a legal one founded in the protection of the related intellectual property rights of the Macro-platforms that are the integrated generation functions of the DSS. Simply, very often the information reported by the DSS to the decision-makers [DM] is in summary form and does not include the underlying statistical context imbedded in the Macro-platforms that were used in its generation. In this sense, often the DM is “flying blind” in effecting an action plan as there is no probability or likelihood context, empirical or otherwise, for the information produced by the DSS re: the forecasting problem under scrutiny. The guidance needed is one of the four pillars of the Balanced Scorecard [BSC]ⁱⁱⁱ of Kaplan & Norton (1992); in particular, in this case, it is the BSC: Learning Loop [Innovation and Learning Perspective—The time and commitment to generate next generation inferential insights that is at issue. For example, Kaplan and Norton note:

During a year-long research project with 12 companies at the leading edge of performance measurement, we devised a “balanced scorecard”—a set of measures that gives top managers a fast but comprehensive view of the business. The balanced scorecard includes financial measures that tell the results of actions already taken. And it complements the financial measures with operational measures on customer satisfaction, internal processes, and the organization’s innovation and improvement activities—operational measures that are the drivers of future financial performance.

Of the four pillars of the BSC, the one that we have selected as critical in the forecasting context is the Innovation and Learning Perspective. This begs the critical question: How can the organization continue to improve and create value in the forecasting context? This is the BSC “Learning Loop” that we will address in this research report. Simply, the learning loop is the essential way that the organization learns and adapts to focus on measures that have the best chance of providing relevant and reliable Intel to move the organization effectively and efficiently into the future.

The simple idea for this BSC-platform is that experimental inferential designs that are used in informing decision-making situations need to be transparent so decision-makers could learn from the past and fine tune decision-making for the future. This is the point of departure of our study. Following, we will:

1. Discuss the importance of creating empirical likelihood information in the forecasting-milieu for selecting a forecasting model. For illustrative guidance, we have adopted a market-trading context from the S&P500.
2. Detail and illustrate the three standard OLS two-parameter linear Regression [OLSR] 95% Confidence Interval models often incorporated into many DSSs that produce OLSR Confidence Intervals [CIs]—i.e., a statistical likelihood context for the forecasts; specifically, the Excel & Fixed Effects & Random Effects CI-Versions will be detailed,
3. Present the standard OLSR-Model forms: [Y:X]—where the Y-variate to be forecasted is conditioned by the X-variate, and the Time Series form: [TS] where the Y-variate “speaks for itself and is time-indexed, t, over the Panel”. Each of these two models generates the three above mentioned 95%CIs,
4. Introduce a Modeling Protocol to aid decision-makers [DM] in selecting the forecasting modeling forms that are the best in the class of the OLSR-models tested. This protocol is called: The OLSR Decomposition and Screening protocol, termed: the [D&S-Triage] protocol. There are two Stages for this protocol. *Stage A*: The Decomposition of the OLSR- [Model & Version]-component set regarding relative precision, and *Stage B* for the inferentially significant OLSR-components from *Stage A*, a multilevel quality-sensitive algorithm is suggested and parameterized to evaluate the *Quality* of the respective 95%CI *Capture Rates*,
5. The [D&S-Triage] protocol is illustrated for an accrual of firms from the S&P500; the forecasting context selected is a very challenging forecasting context introduced by Lusk & Halperin (2015) and Lusk (2019b) where forecasts are produced and evaluated after Panel trajectory turning points, and
6. Suggest further research possibilities for the forecasting domain.

2. The Inferential Morass: The Forecasting Context and an Empirical Analytic Suggestion

2.1 Overview

We offer, to motivate this research report, that there is a dearth of information on the use of the precision of forecasting CIs as a variable of interest in forming a protocol for selecting particular forecasting models—*this is the lacuna to be addressed in our research report*. We found this lacuna surprising as more than 30 years ago Makridakis, Hibon, Lusk & Belhadjali (1987) [MHLB] offered a very comprehensive set of empirical profiles from the groundbreaking M-study [Makridakis et al. (1982)] on the nature of the capture rates for more than 30 forecasting models. One would have expected that such comprehensive and detailed information would have spawned interest in using CI-precision as a/the/ [at least one] of the selection criterion of identifying forecasting models that outperform others tested—a best in class-set. However, that does not seem to be the case, as a search using: ProQuest™, ABI-Inform™, and Business Source Complete™ for the search terms:

{Forecasting Confidence Intervals & [{Stock Markets OR Economic}] & {Precision Decision Making OR Scoring Protocols}

yielded only five references, three of which are certainly interesting as a contextual setting^{iv} for this research report; however, none of these offered detailed precision-based empirical protocols or suggestions to inform the decision-making process regarding using precision in identifying forecasting models that are best in the class tested.

2.2 Illustration

At this point, it would be instructive, actually essential, to probe the two-stage modeling form of the [D&S-Triage] to better understand the important implication of precision differences in the forecasting context. The inferential information that is required in the forecasting context is the likelihood of a reasonable range for the forecast. These are the confidence or capture intervals, [hereafter: CIs], for the forecast that offer the gambling odds for the values of the population expectation from the specific “Random-Panel” selected. This being the case, if different “standard” forecasting models have different CI-versions, then this could create an inferential decision-making conundrum. An illustration will elucidate the nature of the likelihood precision conundrum that is due to the three CI-Versions: {The Excel-CIs [E]; The Fixed Effects-CIs [F]; The Random Effects-CIs [R]} of the two standard OLSR models.

Here we will use an actual firm traded on the S&P500: Cummins, Inc. [CMI: <https://www.cummins.com/1June2006>]. We will detail all the following computations as well as the various components of the testing design in the next section of this research report. Assume the forecasting model is a Times Series. Following is the line-orientation of the {Excel [E], Fixed Effects [F] & Random Effects [R]} CIs. According to Gaber & Lusk (2017) and Lusk (2017), considering the assumptions underlying the production of the CIs, the orientation that one usually expects is:

Schematic A

[-----E[LL]----F[LL]----R[LL]-----Frc[E,F&R]----- R[UL] ----F[UL]----E[UL]-----]

Where, LL indicates the Lower Limits of the 95% CIs, UL the respective Upper Limits, and Frc the Forecast Value.

For the CMI dataset the actual values are:

Schematic B

[-----117.21----125.92----132.65-----[141.36]-----150.07----156.80----165.50-----]

The forecast [Frc] for all three models is the same, excepting rounding, as the average of the three CIs indicates: AVERAGE [117.21; 165.50] = AVERAGE [125.92; 156.80] = AVERAGE [132.65; 150.07] = 141.36.

Precision Intervals: E[24.1*] F[15.4] & R[8.7]

*[(165.50 – 117.21)/2] = 24.145

Assume that the DM is using a DSS that accepts a TS-Panel from the CMI and then generates a 95% CI that will/[could be used] in forming a stock trading strategy. However, as is usually the case, the DSS produces these 95% CIs but does not report (i) the statistical context or (ii) assumptions upon which the computations are based and (iii) there is no related Capture Rate information created. Therein lies the morass. Using Schematic B the conundrum is evident. It will be the case that the Capture Rate will be directly related to the Precision. If the decision-maker just uses that Capture Rate to select the forecasting model for future use, the Excel Version most always will be the preferred choice—it is the widest so it captures relatively more of the future values. What is needed is the Quality of the Capture Rate vis-à-vis the relative Precision. Also see: Chen & Chen (2016). Therefore, the judgement of the forecaster re: the forecasting problem at hand considering the Quality of the Capture Rate is needed to provide a logical platform for selecting the forecasting configuration that is the best in the class after considering the Precision of the CIs. Although this sounds logical it is more complicated than it appears. The reason for this is that precision is impacted by: (i) the nature of the Panel, and (ii) the appropriateness of the assumptions underlying the statistical configuration of the CIs that is used in the forecasting problem. In order to make a logical inferential choice, one needs to create an inference profile-set for the various components of the forecasting models under consideration, the basic information of which is:

The relative Precision—specifically, the precision benchmarked by the forecast, and

The respective Capture Rates scored by the DM for their Quality.

By analyzing the actual/empirical inference profile sets as conditioned by the nature of the components of the forecasting models, the forecaster can arrive at a model configuration that performs best in the class of models tested—i.e., an empirical triage protocol in the inferential set tested. Thus, the focus of this research report is to detail:

How Precision & Models & CI-Versions & Capture Quality all need to be evaluated in making an intelligent selection of a best in class forecasting model.

2.3 Statistical Caveat

This empirically based protocol-focus is a practical alternative to the following “Theoretically Correct CI-selection Protocol”^{vi}: The Forecaster should/must ascertain: (i) the likely statistical profile of the selected Panel segment vis-à-vis the population set—this could be parametrically Ergodic or Non-Ergodic in the usual four statistical moments, and, for robustness, one must also consider appropriate non-parametric alternatives, (ii) then, specify how this sampled “Ergodic” profile set relates to the statistical inferential assumptions of the generating functions of the CIs underlying each of the six components of the Models & Versions, and (iii) finally, based upon these likelihood assumptions & context fit, select the best, usually likelihood,—in the class—Model&Version&CIs protocol. Based upon our decades of experience in the forecasting milieu, we offer that in the history of humanity to date no forecaster engaged in the practice of forecasting has ever followed such a protocol; simply, forecasters rarely have sufficient time, not to mention the expertise, to effect such analyses, and there are no DSSs, of which we are aware that provide such details. For this reason, we offer that the [D&S-Triage] to be detailed in this research report is a practical alternative to profiling the Assumptions and Panel-Context fit to arrive at the likelihood match for selecting the Model and CI configuration. Following, we will discuss how these various components can be woven into an empirical inferential protocol to select a CI-forecasting set to inform the forecasting-model selection process as required in the BSC re: The evolution of the next generation.

3. The Research Montage: The Forecasting Context

3.1 Overview

In this research report, we will assume that the forecasting model class of choice is the OLS two-parameter [α : Intercept and β : Trend] linear Regression model [OLSR]. In this class there are two models that are usually considered the standard fare: Time Series models, where the X-variate is the time index, and the Y:X Model where the X-variate is a conditioning variate for the Y-independent variate. These are noted respectively as: {TS & Y:X}. For each of these two models there are three (1- α)%CIs versions: {The Excel, The Fixed Effects & The Random Effects}. Without loss of generality, we will use the 95% CIs as this is the default in many DSS as well as in the Audit context. We will present these Models and Versions in detail and illustrate them using the following CumminsTM, Inc. data Panel

Cummins, Inc. [CMI: <https://www.cummins.com/> S&P500 TS[t=1]:1June2006]. We will use the first ten Panel values in the OLSR-fitting and for the 95% CIs. Thus, the three S&P500 values in italics: { $Y_{i:11,12,13}$ } are holdback-prices and may be used to evaluate the Capture Rate of the 95% CIs; the bolded value is the Turning Point [TP] identified using the Lusk & Halperin (2015) Screen. Note: after the TP the trajectory of the CMI-Panel changes dramatically. Thus forecasting after a TP will be a challenging test of forecasting acuity.

Table 1. CMI: S&P500 Panel, n=10: Realizations [Italics, n=3]

110.21	122.25	117	114.82	119.23	126.98	119.92
134.56	134.6	144.72	<i>92.16</i>	<i>94.23</i>	<i>101.21</i>	N/A

For the CMI:X-variate, we have the matched time period values for the Capital Cube Price Latest [CCPL®] of the Capital Cube MarketTM Navigation [CCMN] platform^{vi}. We selected this conditioning variable to form the Y:X-version as it is one of the principal variables in the CCMNP. Specifically, the first ten values are used in the: OLSR [$Y[S&P500] = X[CCPL]$].

Table 2. CMI: CCPL Panel of the CMI, Inc

66.2694	63.21489	53.15605	55.6082	60.63303	54.22442	57.32492
52.59557	50.47699	34.04416	N/A	N/A	N/A	N/A

3.2 Regression Models & 95%CI Versions

Following are the three frequently used CI-variations of the two OLSR models regarding the creation of 95% Confidence Intervals [CIs].

3.2.1 OLSR inference from the Excel Parameter Range Model

The Excel Regression functionality forms a “wide-covering” confidence interval. See Lusk (2017) and Gaber &

Lusk (2017). These 95% CIs are effectively extreme case CI-scenarios as they are produced from the crisp-end-point parameters of the 95% CI for the intercept and the slope jointly. Here we offer the following notation:

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

$$\hat{Y}_{LowerLimit:(t=(N+1,2,3))} = [\hat{\alpha}_N - [t_{\frac{\alpha}{2}} \times s_{\epsilon(\alpha)}]] + [\hat{\beta}_N - [t_{\frac{\alpha}{2}} \times s_{\epsilon(\beta)}]] \times [V]$$

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

$$\hat{Y}_{UpperLimit:(t=(N+1,2,3))} = [\hat{\alpha}_N + [t_{\frac{\alpha}{2}} \times s_{\epsilon(\alpha)}]] + [\hat{\beta}_N + [t_{\frac{\alpha}{2}} \times s_{\epsilon(\beta)}]] \times [V]$$

Where: $\hat{\alpha}_N$ & $\hat{\beta}_N$ are the intercept & slope of the OLSR; $s_{\epsilon(\cdot)}$ is the respective standard error, $t_{\frac{\alpha}{2}}$ is the t-statistic for inference for the two-tailed 95% CI that has $df = [N-2]$; N is the last time index in the data-stream—in our study 10. V is either: V: {N+[1,2,3]}, which is the case for the TS model; or for the Y:X [CCPL] model where the last three CCPL values are derived from the TS: Y[CCPL]:X. These CCPL-projections are noted as: V: {CCPL_{i:11,12,13}} and are used to form the S&P500 Y-projections. [Note: The Bolded values following directly are for the TS-Model and were used in Schematic B.]

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

Extreme Left Side [Lower-Limit [LL]] 95% Boundary for the first projection horizon [H1] is:

$$\hat{Y}_{LowerLimit:(t=(N+1))} = [107.5 - 2.3 \times 3.8] + [3.1 - 2.3 \times 0.6] \times [10 + 1]$$

$$\hat{Y}_{LowerLimit:(t=(N+1))} = 117.21$$

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

$$\hat{Y}_{UpperLimit:(t=(N+1))} = [[107.5 + 2.3 \times 3.8] + [3.1 + 2.3 \times 0.6]] \times [10 + 1]$$

$$\hat{Y}_{UpperLimit:(t=(N+1))} = 165.50$$

Y:X-version of the 95% CIs [Using the TS projections of the CMI of: [H1=42.0; H2=39.7 & H3=37.3]] are:

Extreme Left Side [Lower-Limit [LL]] 95% Boundary for the first projection horizon [H1] is:

$$\hat{Y}_{LowerLimit:(t=(N+1))} = [179.7 - 2.3 \times 13.1] + [-1.0 - 2.3 \times 0.2] \times [42.0]$$

$$\hat{Y}_{LowerLimit:(t=(N+1))} = 84.4$$

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

$$\hat{Y}_{UpperLimit:(t=(N+1))} = [[179.7 + 2.3 \times 13.1] + [-1.0 + 2.3 \times 0.2]] \times [42.0]$$

$$\hat{Y}_{UpperLimit:(t=(N+1))} = 190.27$$

3.2.2 OLSR inference from the Expected Random & Fixed Effects Model

The OLSR assumption, in this case, is that there is a random sampling possibility from a well-defined population of variable data streams of realizations, i.e., a population set of: $\{Y_t^l ; l = Firm\ Index\ where\ l\ is\ large\}$ which constitutes a large blocked or stratified group of “like” firms or repeated bootstrapped Panel sets of sufficient size. In this case, the forecaster believes that the sample estimate and the related CIs are formed under the expectation, $E[Y_t^l]$, i.e., the forecast of the mean of Y_t^l ; this is termed the Random Effects [RE] assumption. Experience suggests that the RE assumption is possible though not likely in many forecasting situations context. In a market-Panel collection such as NAICS or SIC groups, the RE event is not unrealistic. Nevertheless, in the service of completeness, the confidence interval for the client value in the RE context for the TS model is for H1:

$$\hat{Y}_{(t=N+1)} \pm t_{\frac{\alpha}{2}} \times s_{\epsilon} \times \sqrt{\frac{1}{N} + \frac{[(N + 1) - \bar{N}]^2}{S_{tt}}}$$

Where: $\hat{Y}_{(t=N+1)}$ is the value of the fitted regression projected for the next X-index time-point [H1] or $N + 1$ using the parameters produced from the OLSR fit for the N data points; $S_{tt} = [\sum t^2 - [(\sum t)^2]/N]$;

$s_\epsilon = \text{OLS}[N]:MSE^{1/2}$; and \bar{N} is the Mean of the time index for the N data points.

For the Y:X version, H1[CCPL₁₁] is:

$$\hat{Y}_{(t=N+1)} \pm t_{\frac{\alpha}{2}} \times s_\epsilon \times \sqrt{\frac{1}{N} + \frac{[(CCPL_{11}) - \bar{CMI}]^2}{S_{xx}}}$$

Where: $\hat{Y}_{(t=N+1)}$ is the value of the fitted regression projected for the next X-index time-point H1, using the parameters produced from the OLSR fit for the CMI-data points; $S_{xx} = [\sum x^2 - ([\sum x]^2)/N]$; $s_\epsilon = \text{OLS}[N]:MSE^{1/2}$; and \bar{CMI} is the Mean of the CMI-values use to fit the regression; and H1 is the first point after the last data point used in the regression.

3.2.3 OLSR Inference from the Fixed Effect Projection

The assumption is that the object of interest is the jth firm with a single set of data stream realizations: i.e., the set of:

$$\{\{Y\}j^l\}; j = a \text{ particular firm Index over all } l \text{ firms}\}$$

Where this jth Firm has longitudinal dummy-variable integrity from all of the other firms in the population and so the projection is not the average of all the firms but only for that jth firm. In this case, and given the usual assumptions rationalizing the OLSR of the time-series, the confidence interval for the extrapolation for the next point in the firm time stream is:

$$\text{Specifically*}, \hat{Y}_{(t=N+1)} \pm t_{\frac{\alpha}{2}} \times s_\epsilon \times \sqrt{1 + \frac{1}{N} + \frac{[(N+1) - \bar{N}]^2}{S_{tt}}}$$

*Note the Fixed Effects Penalty Multiplier is the only difference in the CIs from the Random Effects form.

For example, for the TS-version for the CMI dataset for H1:

Random Effects:

$$\hat{Y}_{(t=N+1,LL)} = 132.65 = 141.36 - [2.3 \times 5.5 \times 0.7]$$

$$\hat{Y}_{(t=N+1,UL)} = 150.07 = 141.36 + [2.3 \times 5.5 \times 0.7]$$

Fixed Effects:

$$\hat{Y}_{(t=N+1,LL)} = 125.92 = 141.36 - [2.3 \times 5.5 \times 1.2]$$

$$\hat{Y}_{(t=N+1,UL)} = 156.80 = 141.36 + [2.3 \times 5.5 \times 1.2]$$

Where [For illustration]: $S_{tt} = [385 - [55^2]/10] = 82.5$; $[(5.5 - 11)^2] = 30.3$ Therefore, the third component is: $[\sqrt{(1/10 + 30.3/82.5)}] = 0.7$; $\sqrt{MSE} = \sqrt{30.6} = 5.5$; $T.INV.2T(5\%,8) = 2.3$; Finally: the forecast for H1 is: **141.36**: $[107.5 + [3.1 \times (1+10)]]$

For the FE version the only change is in the Penalty Parameter that is: $\sqrt{(1 + 1/10 + 30.5/82.5)} = 1.2$

For the Y:X-version again using the CMI data set for CCPL₁₁:

Random Effects:

$$\hat{Y}_{(t=N+1,LL)} = 129.0 = 137.3 - [2.3 \times 6.2 \times 0.6]$$

$$\hat{Y}_{(t=N+1,UL)} = 145.6 = 137.3 + [2.3 \times 6.2 \times 0.6]$$

Fixed Effects:

$$\hat{Y}_{(t=N+1,LL)} = 120.7 = 137.3 - [2.3 \times 6.2 \times 1.2]$$

$$\hat{Y}_{(t=N+1,UL)} = 153.9 = 137.3 + [2.3 \times 6.2 \times 1.2]$$

Where [For illustration]: $S_{xx} = [30 \ 681.6 - [547.5^2]/10] = 700.8$; $[(54.8 - 42.0)^2] = 163.2$ Therefore, the third component is: $[\sqrt{(1/10 + 163.2/700.8)}] = 0.6$; $\sqrt{MSE} = \sqrt{39.0} = 6.2$; $T.INV.2T(5\%,8) = 2.3$; Finally: the forecast is 137.3: $[179.7 + [-1.0 \times (42.0)]]$

For the RE version the only change is in the Penalty Parameter—that is: $\sqrt{(1 + 1/10 + 163.2/700.8)} = 1.2$

3.3 Point of Information

All of these computations are programmed in a VBA-DSS that creates the required Model&Version information and, thus, information at any horizon. Following, we will detail the multi-stage protocol, [D&S-Triage], that will Decompose and then Screen the accrued forecasting data profiles to generate the next version of the forecasting model that will be inferentially the best in class tested.

4. Foci: Precision and Capture: The OLSR Decomposition and Screening Profiler

4.1 Overview

Recall the context for this research report. For the *Stage A* analysis, we are interested in the nature of the precision differences for the six different test-sets:

TS{Excel[E]; RandomEffects[R]; FixedEffects[F]} & Y:X{Excel[E]; RandomEffects[R]; FixedEffects[F]}

For each of these there will be three forecasts at: {H1; H2; H3}.

In this context, so as to form the initial robust-profile for the forecasting model components in the spirit of improving the decision-making forecast model selection process, we explore the inference for the precision differences over the accrual set of information using the standard ANOVA-model and the related non-parametric models found in: *SASTM[JMPTM.13[2017]*; we mention the JMP-inference platforms to: (i) note that the inferences to be reported herein are founded on basic or standard assumptions that are detailed in most every introductory statistics text, and (ii) thus, provide a simple replication-link^{vii}. The logic of focusing on inferential precision differences is that if it were to be the case that the various standard comparative Nulls could not be rejected, then detailed triage would be moot. As a final note, we elected to use a forecast Panel of three one-period-ahead forecasts: {H1; H2 & H3}. This is the short-term forecasting horizon Median-profile used in the Collopy & Armstrong (1992) and the Adya & Lusk (2016) studies. This horizon election will likely control the OLS-variance and so will enhance power; it is well discussed in these studies that long-forecasting horizons invite high forecasting error. As noted by Makridakis & Hibon (2000, p. 452): “the accuracy of the various methods depends on the length of the forecasting horizon”. These are the same results reported by Brodie, Buccellato & Scheffel (2011, p. 131) who report:

The main objective of this work is to explore the predictive power of individual firms’ turnover and capital expenditure series based on four key business surveys administered by the Office for National Statistics. These are the monthly MPI and MIDSS, the quarterly CAPEX and the annual ABI. The results suggest that the predictive power of the data starts to drop substantially after the end of the second year and the beginning of the third.

4.2 Stage A: Details and Principal Inferential Tests Addressing Precision Decomposition

The above CMI-set of detailed information is informative not only as a computational exercise for instructing these forecast models in an academic class setting, but also to enable replication and encourage further studies. For this research report, with the above details as the context, we offer the nature of the inferential tests that we will use to determine if there is inferential differential evidence from these empirical datasets re: Relative Precision. Specifically, for the CMI data for the Y:X Model; Fixed Effects 95%CI Version for the first horizon the values for the Precision are:

95%CI: Forecast: [LowerLimit: UpperLimit]

95%CI: 137.3: [120.7: 153.9] Where **Precision** is: $[(153.9 - 120.7)/2] = 16.6$.

This is the precision in the Units of measure of the Panel. For comparison purposes across Panel for varying units a normalization is needed. The usual standard is to divide the Precision by the forecast value. Thus the Relative Precision [RPrecision]is;

$[(153.9 - 120.7)/2]/137.3 = 16.6/137.3 = 12.1\%$

This RPrecision may now be compared across precision values from any Panel.

For the Stage A partition, noted as [PT] the analyst will block the forecasting components for the inferential testing environment. Initially, we will use the One-way parametric ANOVA and the related Non-Parametric alternatives for the following One-Way blocking:

PT[1] We will block the accrued datasets on Regression Model *Versions* {Excel & RandomEffects & FixedEffects} re: RPrecision,

PT[2] We will block the accrued datasets on Regression *Models* {TS & Y:X} re: RPrecision

PT[3] We will block the accrued datasets on Forecasting *Horizons* {H1& H2 & H3} re: RPrecision.

Given these One-Way Mean/Median difference test results, we will glean the possible effects of the blocking and thus collect information as to the interaction-effects that are likely to be productive to consider in the full n-Way ANOVA. This test will be used to clarify the inference structure to be used in forming the information set for the *Stage B* where the *Quality* protocol will be developed.

PT[4] Using the above datasets, we will compute the n-Way parametric-ANOVA with the selected Combinatoric Interaction.

With the above discussion, and recalling that all of this RPrecision information was produced by a VBA:DSS, available at no cost as a download, where the run time for these datasets is less than 3-seconds, we now present the results of the accrual of the 32 S&P500 Panels which will be the accrual test-set drawn from the firms noted in the Appendix.

4.2.1 ANOVA [PT [1]] RPrecision Re

Horizons [Times: H(t) =11, H(t) =12 & H(t) =13] We elected to start where it is most likely the case that there are no inferential differences. Experience suggests that for the short-term projection frame there are rarely overall RPrecision inferentially important differences for the Models&Versions. The reason for this is that the error-penalty component is relatively bounded in the short-term. Where there are likely to be major RPrecision differences are in the long-run—effectively after the 6th projection frame. The Horizon-ANOVA results are presented in Table 3.

Table 3. RPrecision Profiles for the Short-run Forecasting Projection Horizons [*IQR is the Inter-Quartile Range—also a standard non-parametric measure of spread/variation in the dataset.]

	Mean[Median]	SDev[IQR*]
H(t) =11	37.5%[18.6%]	46.7%[27.6%]
H(t) =12	40.2%[19.3%]	49.2%[28.3%]
H(t) =13	44.3%[20.4%]	56.0%[32.3%]

*IQR is the Inter-Quartile Range—also a standard non-parametric measure of spread/variation in the dataset.]

Discussion The inferential tests using Tukey-Kramer HSD [$\alpha=5\%$], hereafter [T-K] found no pairwise-differences. The ANOVA overall test had a p-value of 0.42 and the Wilcoxon & Kruskal-Wallis [Rank-Sum] p-value was 0.32. Implication: There is no evidence that there are RPrecision differences in the short-run forecasting horizons. This means that the n-Way ANOVA may be logically consolidated; this will reduce the effect-set of interactions and thus increase the Power and so make the Quality profile easier to configure. Vetting note: The expected increase in RPrecision as well as its variability over the projection horizons is in evidence although not tested inferentially.

4.2.2 ANOVA [PT [2]] RPrecision Re

Forecasting Models: [TS & Y:X] There is copious experiential and theoretical information that indicates that, in the OLSR context, the Time Series and the conditioning model Y:X are, in practice, almost exclusively used as the basic forecasting models. In Table 4 their RPrecision profiles are presented.

Table 4. RPrecision Profiles for the Forecasting Models that are Predominate in the Practice Oriented Forecasting Domain

	Mean[Median]	SDev[IQR]
TS	28.6%[17.9%]	38.0%[17.6%]
Y:X	52.7%[22.3%]	58.6%[81.8%]

The inferential tests indicate that there are differences over the Models. Specifically, conservatively the T-K HSD [$\alpha=5\%$] found the Models [TS vs. Y:X] to be different. The ANOVA [Welch] overall-test [similar to the standard Welch-corrected t-test] had a p-value of <0.0001, and finally, the Wilcoxon & Kruskal-Wallis [Rank-Sum]

p-value was 0.0004. Implication: There is strong evidence that there are RPrecision differences between the OLSR Models: TS and Y:X. Caveat: The Y:X model information used the CCPL^{viii} as the conditioning variable; sometimes the conditioning variable may reduce the variation and so produce more narrow CIs. Sometimes this is not the case. Therefore, this result is conditional.

4.2.3 ANOVA [PT [3]] RPrecision Re

Forecasting Versions: {Excel, Fixed-Effects [FE] & Random-Effects [RE]} In the last partition, we tested the three 95% CIs that are the usual fare. The results of the Lusk (2017) and Gaber & Lusk (2017) studies give preliminary information that the Excel version is likely to produce relatively wide CIs as the Excel-computation uses the extreme case for both parameters in the computation of the CIs. We shall re-consider the utility of this usual assumption in the summary analysis. Further, mathematically, the RPrecision of the Fixed Effects is always greater than that of the Random Effects. Therefore, in this blocked case, one would expect the following relationships presented in Table 5 over the One-Way ANOVA: Versions.

Table 5. RPrecision Profiles for the Model Versions in the One-Way ANOVA

	Mean[Median]	SDev[IQR]
Excel[E]	79.5%[54.0%]	64.3%[79.9%]
Fixed Effects[F]	26.7%[19.5%]	29.0%[13.7%]
Random Effects[R]	15.6%[10.7%]	21.0%[7.8%]

The inferential tests used all three [C_2^3] pairwise differences over the Versions [E&F&R]. Testing results: The T-K HSD [$\alpha=5\%$] found that all three pairwise differences were significantly different. The ANOVA [Welch] overall-test had a p-value of <0.0001 and the Wilcoxon-Kruskal-Wallis [Rank-Sum] p-value was <0.0001. Implication: There is strong evidence that there are RPrecision differences between the OLSR Model Versions: E&F&R. Caveat: this is a “rather” specific rather than a general result based upon the use of the CCPL as the conditioning variate.

4.2.4 Stage A

Summary: One-Way ANOVA Results There is no reason to use the short-term projection horizon for the Stage B screening protocol. Therefore, it is possible to concentrate on the other two main-effects: Model-type {TS & Y:X} and Model-version {Excel & Fixed-Effects [FE] & Random-Effects [RE]}. In this case, these will be examined in a 2-Way ANOVA with interaction. This would give further guidance as to which variables can be expected to be useful in Stage B—the development of the Excel-Screening Protocol (our quality algorithm) for use by the DM.

4.2.5 2-Way ANOVA

With Interaction Component Overall Inference Profile For the partition comparisons, given the above results, we used only the Models {TS & Y:X} and the Versions {Excel & FE & RE}; thus, in these cases there will be a 2-Way layout with interactions in the ANOVA context. In addition, we will use the parametric ANOVA-model, as it is the only standard form that is usually part of most of the statistical platforms. Recall that we used robustness tests in the One-Way context as the Wilcoxon & Kruskal-Wallis [Rank-Sum] test is a part of most of the inferential platforms.

The main effects had overall p-values <0.0001 as expected from the One-Way results presented above for the Least Square Means [LS-Means] as reported by the ANOVA. The six-matched interaction profiles presented in Table 6 and screened using the HSD test-matrixes are most interesting.

Table 6. RPrecision Profiles for Interaction: {Models and Versions} from the 2-Way ANOVA

Model&Version	LS-Means
TS, Excel	41.2%
TS,FE	27.2%
TS,RE	17.6%
Y:X,Excel	100.2%
Y:X,FE	26.6%
Y:X,RE	13.7%

Here we used the T-K HSD [$\alpha=5\%$] profiles to probe the precision of the OLS-space. Following there were three summary indications that will be useful in forming the quality algorithm for Stage B^{ix}:

The First partition is: Y:X, Excel [100.2%] has a RPrecision difference that is different from all of the other five partitions.

The Second partition is: TS, Excel [41.2%] which has a RPrecision difference that is different from all of the other four partitions.

The Third partition is an inferential blend: the Group: {TS,FE[27.2%] & Y:X,FE[26.6%]} shares OLS-inferential links with the Group: {TS,RE[17.6%] & Y:X,RE[13.7%]} the last of which stands alone re: RPrecision.

Implication Although this appears complicated to comprehend in the ANOVA: T-K profile-links, the RPrecisions follows the logic of the LS-mean differences of the six-ANOVA-partitions.

The summary implication of Table 6 is that clearly the Y:X, Excel version exhibits, on average, a CI that is very unlikely to provide useful information; this is consistent with the vetting expectation. However, as indicated by the p-values, the LSM-profile suggests that there may be information in the 2-Way results useful in configuring the Quality Profiling stage. Pedagogical Note: We have found that this set of T-K variables for the 2-Way ANOVA interaction [Table 6] is understandable to the students in our Accounting Information Systems [AIS] undergraduate course. Consider now the Stage B formulation.

4.3 Stage B

The Variable of Interest for the n-Way ANOVA-partition: Capture Rate Screening Protocol. It is standard to post-audit the forecasts that are used in creating decision-making information; this is the learning loop that is recommended by Kaplan & Norton (1992) as noted previously. To underscore the importance of the Balanced Scorecard, we offer a vignette of an interview conducted by Prof. Dr. Kaplan (1993) with Mr. Larry D. Brady, EVP of the Chicago based FMC, Corporation as reported in the Harvard Business Review.

Robert S. Kaplan: What's the status of the balanced scorecard at FMC?

Larry D. Brady: Although we are just completing the pilot phase of implementation, I think that the balanced scorecard is likely to become the cornerstone of the management system at FMC. It enables us to translate business unit strategies into a measurement system that meshes with our entire system of management. For instance, one manager reported that while his division had measured many operating variables in the past, now, because of the scorecard, it had chosen 12 parameters as the key to its strategy implementation. Seven of these strategic variables were entirely new measurements for the division. The manager interpreted this finding as verifying what many other managers were reporting: the scorecard improved the understanding and consistency of strategy implementation. Another manager reported that, unlike monthly financial statements or even his strategic plan, if a rival were to see his scorecard, he would lose his competitive edge.

This is the essential issue that we have addressed in our research report—to wit: how to effectively and efficiently focus the decision making activity of the organization. As expressed by Mr. Brady, the BSC uses screens intelligently selected to ferret out the essential details that are likely to matter in forming actions plans. This is the intent of the [D&S Triage] and the result of its execution. By identifying the essential components of Quality forecasting models from the past this information can be used in the BSC: [Learning and Innovation Phase] to guide future actions. Specifically, after the forecasts and the respective 95% CIs are produced, then, when the actual realizations are known, the capture information would be recorded. Overtime, there will be a number of such forecasts realizations and so many Capture Rate Profiles can be produced and analyzed. At that point, the forecaster can consider the various capture profiles relative to the particular parameters of the

forecasting protocols employed using the [D&S Triage]. This is the Learning Loop of the BSC that enables DMs to make better selections of forecasting models in the future. Consider now an illustration of the Learning Loop for the accrual datasets that we have created.

As we discussed in the introductory section, the driver of the analysis is the Quality of the Capture Rate vis-à-vis RPrecision. In this case, this ANOVA-interaction analysis also needs to be parametrized re: Quality. We have selected the following Quality profiler to screen the CI-capture event for the particular forecasting protocols. An illustration will elucidate these ideas.

4.4 Scored Quality of the Capture Rate

A very simple, intuitive, and relevant method of scoring the capture rate is to use the general IF(AND) Excel-screen configuration as our Quality profiler. Specifically, for each of the THREE Horizons, we used:

```
IF(AND(RPrecision>0, RPrecision <=P1, HB=1),"Best",
IF(AND(RPrecision >P1, RPrecision <=P2, HB=1),"Norm",
IF(AND(RPrecision >P2, RPrecision <=P3, HB=1),"Worst","WorstWorst")));
```

Where: the Pi, [i=1,2,3] are ordered as: [0<P1, - - -, <P3].

4.4.1 Discussion of the Excel Codex

This Excel codex represents our quality scoring algorithm; it is termed: Ex[A]. For the Ex[A], HB represents the Holdback or Actual value for a particular projection horizon. The indication HB=1: means that the Holdback at that particular horizon was in that 95%CI. While the HB is evaluated in the units of the Panel, the Actual Precision of the Panel is calibrated relative to the forecast value by dividing Precision by the forecasted value so as to enable the comparison of the various forecasting models. As discussed above this is noted in Ex[A] as RPrecision. This computation will be illustrated in the next section. The coding test for the Holdback is programmed in the DSS-Excel worksheet as follows: IF(AND(HB>=LL, HB<=UL,1,0). This coding asks IF the Holdback is in the 95%CI, where: LL is the Lower-Limit and UL is the Upper-Limit of that particular 95% CI—i.e., IF these two conditions are TRUE, THEN Excel records = 1; if FALSE, then Excel records 0 meaning the HB is NOT in the 95%CI. This is the evaluation of the Holdback value using the 95%CI. It has ONLY two states: TRUE≡The HB is IN the CI; otherwise: The HB is NOT IN the CI.

Then using, Ex[A], the following information is created:

IF the “RPrecision” is in the Interval [0, P1] AND HB=1, then “Best” is recorded. IF either condition is not TRUE, then Excel moves to the next conditional. IF the “RPrecision” is in the Interval [P1, P2] AND HB=1, then “Normal” is recorded, IF this second conditional test is FALSE, then, Excel takes up the third conditional. IF the “RPrecision” is in the Interval [P2, P3] AND HB=1, then “Worst” is recorded; IF this third/last conditional test is FALSE, then Excel records “WorstWorst”.

4.4.2 Fixing the Discriminators

The next question is then: What are reasonable values of: {P1, P2, P3}? This is the critical question as this is the precision screening conditional that parameterizes Ex[A]. We formed these {Pi} screening values using the most complete confidence interval study over various horizons reported in the peer review literature—MHLB and also to a lesser extent the research report: Lusk (2019). Specifically, we used the results of Exhibit 1 of the MHLB report where the averages of the percentages that fell outside of the CIs were reported. To parametrize the {P1, P2, P3} of Ex[A], we took the average of the two largest values of Exhibit 1. This gives an interior CI spanning interval of: [100% – (Average [34.5%; 36.1%])] = 64.7% or a relative RPrecision of 32.4% [64.7%/2] for the mid-point P2. To give a benchmarking variation to derive values of P1 & P3, we used a variable calibration of P2×[50% &150%] applied to the average of P2. This gives as the lowest simply calibrated value: P1 as: 15% [16.2%=32.4%×50%] and for P3: 50% [48.6%=32.4%×150%]. Therefore, for P2 we used 35%; in this case, for simplicity, the three values are: 35±15%. In the Lusk (2019, p.14) study of the setting Capture Intervals for use in the Bloomberg™ Terminals, the interval capture RPrecision was reported as: 32.5% which provides reasonable confirmatory information.

The Ex[A] profiler may be illustrated using the Ryder System [R: <https://ryder.com/en/>] Panel of monthly data, n=10 Starting 1st April 2007 for the TS: Excel-Model and the Excel 95%CI-Version, we have the following information in Table 7.

Table 7. Ryder System Precision Profile using Ex[A]

	LL	F	UL	HB	RPrec	Capture	Label
H1	32.94	48.26	63.57	57.61	31.7%	1=Yes	Norm
H2	31.54	47.74	63.94	60.91	33.9%	1=Yes	Norm
H3	30.13	47.22	64.31	68.47	36.2%	0=No	WorstWorst

The actual values of the Ryder dataset are as noted in the first four columns of Table 7: All of the CIs are the 95% CIs produced using the Excel-TS: Forecasting platform. As noted above, as it is the usual case that the magnitude of all of the Panel variables is different, it is necessary to create a normalization so that all of the forecasts can be compared. The simplest calibration is to use the Forecast value. Therefore, the RPrecision in Table 7 [noted as RPrec] or relative precision is computed as: 31.7% $[\frac{((63.57 - 32.94)/2)}{48.26}]$. The Labels in Table 7 are created using [D&S Triage & Ex[A]] illustrated as follows:

HB1= 57.61 and is in the 95% CI for the first horizon [H1]; thus the coding for the two conditions is: [1=Yes] AND [15% < 31.7% <= 35%], thus as BOTH AND conditions are satisfied the Ex[A] Label given is “Norm”.

Note that the WORSTWORST label results for the third horizon: as the HB3 is > than the UL [68.47 > 64.31] and so the AND condition is never TRUE. Also note for example, the same result would have obtained if the RPrecision were to have been > than 50% regardless if the HB were to have been in the 95% CI or not. This is the case as the Excel conditional is AND and not OR.

4.5 Screening the Inferential Measure

Over time the particular [D&S Triage & Ex[A]] filter will be applied to a number of Panel tests thus developing a rich statistically powerful Panel dataset for developing inferential information for making a decision as to the nature of the forecasting configuration that, considering the past Panel information, is the best in the class tested. To illustrate this screening phase, we randomly collected 16 firms for which there were 32 Panels identified using the Lusk-Halperin:Turning Point [TP] Screen. We also used the CCPL as the Y:X conditioning variable to the S&P500 that was the Y-variate. The 16 firms were tracked over the time period 2005 to 2013. The panels selected were identified as the ten monthly closing prices preceding a stock price TP—dramatic change in the trajectory of the stock price. The details of this TP analysis are best found in Lusk and Halperin (2015) and Lusk (2019b). For each of these 32 Panels there were Two Models: {TS & Y:X} and Three CI-Versions: {Excel & Fixed Effects & Random Effects}, and Three Horizons: {H1 & H2 & H3} or 576 data points: $[32 \times 2 \times 3 \times 3]$. This profiling would be typical for inspecting the contribution components of the various forecasting protocols. These data-profiles should be aggregated to create the valuable post-forecast audit inferential information that will be used to select the forecasting model in the future. This is, of course, the reason for the follow-up in the post-auditing phase or the “Learning-Loop” offered as one of the platforms of the Balanced Score Card.

Given the ANOVA profiles presented and discussed above, in particular the 2-Way ANOVA-effects with Interaction, Table 8 was created using the screening function, Ex[A].

Table 8. Summary Screening of the Selected Inferential Pairs

Ex[A]/Labels Given	TS[Excel]a	TS[RE]b	TS[FE]c	Y:X[Excel]d	Y:X[RE]e	Y:XTS[FE]f
Best&Normal	24	10	21	0	7	20
Worst&WorstWorst	72	86	75	96	89	76
Best&Normal%	25%	10.4%	21.9%	0.0%	7.3%	20.8%
Worst&WorstWorst%	75%	89.6%	78.1%	100%	92.7%	79.2%
P-value[x&y]	<0.01[a&b]	<0.03[b&c]	>0.60[a&c]	<0.03[d&e]	<0.01[e&f]	<0.01[d&f]

4.5.1 Elaboration of the variables of Table 8

Using the D&S Triage & Ex[A] for the six-capture 95% CIs, each of which had n =96, we recorded the number of instances where Best OR Normal were recorded and also where Worst OR WorstWorst were recorded. This is then re-calibrated as the relative percentages. In the P-value row, we presented conservative two-tailed test of proportions, not using the continuity correction, for the indexed [x&y] within Model&Version comparisons as an

initial screening-set—i.e., there are $[C_2^3]$ comparisons within each model-block. Also the vetting expectation for the Y:X[Excel] testing arm is founded as there were no instances where Best OR Normal were recorded as the precision was in most of the cases > than P3. For example, selecting components that have the highest within block performance profile—to wit: the TS[Excel]a and the TS[FE]c, we made the following computations:

$$z\text{-cal} = \left[\frac{\text{Abs}(25\% - 21.9\%)}{\left[\frac{25\% \times (1 - 25\%)}{96} + \frac{21.9\% \times (1 - 21.9\%)}{96} \right]^{.5}} \right] = 0.51$$

$$P\text{-value} = [T.DIST.2T(0.51, 10000)] = 0.609$$

This suggests, strongly, that the Null is not likely to be rejected—and so there is likely no difference in the Capture percentage between the TS-Excel and the TS-FixedEffects trials relative to precision. This is very important as there IS a tendency to reject out-of-hand Excel-versions assuming that their relative precision is excessively large. However, in this case, considering the Quality of the Capture Rate, this is not born out. But rather, either of these models would be reasonable choices if a capture rate of on the order of 25% to 20% were to be acceptable.

4.5.2 Inferential Impact Re: The Screening Protocol

There is a DM-judgement aspect to ferreting out the best in class forecasting-model-set. This is the case, as even for this reduced-set where the horizons were not in the possible selection-mix, there were, nonetheless, 15 pair-wise comparisons $[C_2^6]$. Thus, our common-sense recommendation is to select a few of the Model&Versions that have the highest *Quality Capture Rates*. This will reduce the inferential set and make the choice even more logical and defensible for the forecasters. In the illustrative case, there were, in fact, the following three *Quality* comparisons of obvious interest: {TS[Excel] & TS[FE] & Y:X[FE]} as presented in Table 9.

Table 9. Selected Final Screens

TS[Excel]a: 25%	TS[FE]c: 21.9%	Y:X[FE]f: 20.8%
p-value: [a&c] = 0.61	p-value: [c&f] = 0.85	N/A
p-value: [a&f] = 0.48	N/A	N/A

Using the pairwise test of proportions for the three information sets in Table 9, we found that:

TS[Excel] & TS[FE] & Y:X[FE] are not pairwise inferentially different for their capture profiles labeled Best or Norm.

4.5.3 Testing Summary and Implications

Recall the relatively challenging forecasting task tested in this study, specifically: (i) the last value in the forecasting model was a Panel TP, and (ii) the Relative Absolute Error analysis reported by Lusk (2019b) found that there was no forecasting model that outperformed the Random Walk [RW]—the Occam’s version as the forecast/model. Even for this context, using the [D&S Triage & Ex[A]], we learned the following related inferential impacts that will certainly inform the forecasting model selection process:

1. For these three modeling possibilities {TS[Excel] & TS[FE] & Y:X[FE]} a reasonable expectation is that the average of the capture rates will be in the 95% Confidence Interval: [14.21% to 30.93%]. This is not a trivial result for the forecasting context addressed in this research report. However, it is possible that the capture rate of these models is not sufficiently high to warrant their use for the particular forecasting context under examination. Simply, we have used the [D&STriage & Ex[A]] protocol to identify the best among the various possibilities tested. Sometimes the best is not, in fact, useful given the a priori forecasting requirements. This is, of course, valuable information.
2. There is a strong inferential indication that the 95% CIs of the RE-Model are too narrow to garner any serious interest in the RE as forecasting model in a challenging context, even if the stringent assumptions for its use were to be founded,
3. There is also a strong inferential indication, albeit idiosyncratic, that the Y:X Excel version has a precision profile that renders its use moot for this accrual set; this was also suggested by Gaber & Lusk (2017) in another context,

4. The Lusk (2019b) report was not accurate in characterizing the CCPL as not a useful conditioning Market Navigation Platform-variable. Rather, as an elaboration, it seems that in the TP-context using the CCPL as a conditioning variable did in fact emerge as a possibility not inferentially different from the TS-versions as noted above. However, in a portfolio profile, the return would have to be commensurate with the relatively low capture rate—specifically: the Risk & Return where the Quality Capture Rate is at best 30% [i.e., much less than 50%] and so may not be sufficiently high to justify using this model in the TP-context, and finally

5. It is also recommended, based upon our experience, and consistent with the dynamics of the BSC: Learning Loop, that every six months or so in a normal economy where there are no “catastrophic” events such as that created by the Money-Mad-Morons of Lehman Bros™, LLP aided and abetted by an shockingly inept SEC, that the [D&S Triage & Ex[A]] protocol be recalibrated using an up-dated longitudinal dataset starting at the end-point of the previous longitudinal dataset.

5. Summary and Outlook

5.1 Summary

The purpose of this research report was to offer an experientially based protocol on the use of confidence interval precision variations to inform the forecasting decision-making process. Thus, we offer that the [D&S Triage & Ex[A]] protocol is a comparatively reasonable, systematic, and time efficient triage-protocol to arrive at a forecasting model that is the best in class tested. Recall that in the introductory section, we noted that a comprehensive set of searches in the peer review literature found no protocols on calibrating the utility of forecasting CIs relative to the nature of their precision for guidance in selecting a forecasting model. To fill this surprising lacuna, we presented the Stage A: Decomposition and the Stage B: Quality algorithm called the: [D&S Triage & Ex[A]] protocol. The overview of these two stages are:

5.1.1 Overview of Stage A

In Stage A—i.e., the decomposition stage, 42 statistical combinations were initially identified blocking on the combinations for the horizons. Such a profile-set would be a very challenging Human Information Processing task for even the most skilled decision-makers. However, with a carefully designed decomposition using the One-way and n-Way ANOVA partitions, it was possible to make sense of the various precision profiles. In the example presented, we first tested the various One-way ANOVA effects; we found that there was no evidence that there were relative precision differences over the horizon-set. Therefore, we eliminated them in the 2-Way ANOVA; this reduced the effect-set from 42 effects tests to the six-effects produced in Table 6 [42/ [Sum[C_i^3 , $i=1,2,3$]].

5.1.2 Overview of Stage B

Once the important inferential components are identified, then using a holdback Panel, a simple Quality algorithm was parameterized to create a second set of statistical profiles so as to select the CIs and the forecasting model forms that are experientially the best for that forecasting context respecting the Quality of the Capture Rate. In this case tested, the following three viable alternatives: TS[Excel] & TS[FE] & Y:X[FE] were found to be the best in the classes tested.

5.2 Outlook

The Balanced Scorecard teaches that in a dynamic context, analysis and innovation are required to promote and evolve operating effectiveness and efficiencies. In this vein, we would hope that more research is produced on the OLS-model variants, such as: Moving Average Models, models in the Exponential Smoothing class, in particular the: ARIMA (0,1,1)/Simple Exponential, ARIMA(0,2,2)/Holt Model [and the Winters variant] or ARIMA(1,1,2)/Damped-Trend, as well as Judgement Models such as Rule Based Forecasting (Collopy & Armstrong (1992), and finally taxonomy models such as offered by: Adya & Lusk (2016). In this regard, the [D&S Triage & Ex[A]] protocol would be used in the same way as illustrated in this research report. In fact this is one of the strengths of the [D&S Triage & Ex[A]] protocol; it is a passpartout protocol that can be applied in many situations where reasoned profiles delineations are of interest. A study begged by this research report would be to examine a sample of the M-Study that was used in the MHLB report to determine if the [D&S Triage & Ex[A]] protocol would enhance the capture profiles of the many models tested. Finally, it would also be very useful to re-calibrate the Best: Normal: Worst: & WorstWorst demarcations using the MHLB results as well as other relevant empirical studies as benchmarks.

Acknowledgments

Thanks and appreciation are due to: Dr. H. Wright, Boston University: Department of Mathematics and Statistics, the participants at the SBE Research Workshop at SUNY: Plattsburgh, and Mr. Manuel Bern, Chief of Internal Audit: TUI International, GmbH, Hannover, Germany and the comments offered by reviewers of the International Business Research journal for their careful reading, 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*, on-line. <https://doi.org/10.1016/j.dss.2015.12.009>
- Brillinger, D. (1974). *Time Series: Data Analysis and Theory*. SIAM [International Series in Decision Processes] ISBN: 13: 978-0030769757.
- Brodie, P., Buccellato, T., & Eric Scheffel, E. (2011). Assessing the accuracy of business-level forecasts. *Economic & Labour Market Review: [Office for National Statistics]*, Apr, 119-134. <https://doi.org/10.1057/elmr.2011.46>
- Chen, T. L., & Chen, F. U. (2016). An intelligent pattern recognition model for supporting investment decisions in stock market. *Information Sciences*, 346(7), 261-274. <https://doi.org/10.1016/j.ins.2016.01.079>
- Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time-series extrapolations. *Management Science*, 38, 1394-1414. <https://doi.org/10.1287/mnsc.38.10.1394>
- Duttle, K. (2016). Cognitive skills and confidence: Interrelations with overestimation, overplacement and overprecision. *Bulletin of Economic Research, suppl. Supplement 1*, 68, 42-67. <https://doi.org/10.1111/boer.12069>
- 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>
- Kaplan, R. (1993). Implementing the Balanced Scorecard at FMC Corporation: An interview with Mr. Larry D. Brady. *Harvard Business Review, Sept-Oct*, 143-147.
- Kaplan, R., & Norton, D. (1992). The Balanced Scorecard: Measures that drive performance. *Harvard Business Review, Jan-Feb*. Retrieved from <https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2>
- Lusk, E. (2017). Analytic Procedures: A Holdback-vetting forecasting model. *Applied Finance and Accounting*, 3, 65-74. <http://dx.doi.org/10.11114/afa.v3i1.2139>
- Lusk, E. (2019). Short-Term forecasting in the trading markets: Suggested capture intervals for Bloomberg FA-Downloads, *International Journal of Accounting and Financial Reporting*, 9, 1-19. <https://doi.org/10.5296/ijافر.v9i1>
- Lusk, E. (2019a). Time series forecasting in stock trading markets: The turning point curiosity. *International Journal of Research in Business and Social Science*, 8, 1-16. <https://doi.org/10.20525/ijrbs.v8i4.283>
- Lusk, E. (2019b). A benchmarked evaluation of a selected CapitalCube™ interval-scaled market performance variable. *Accounting and Finance Research*, 8, 1-12. <https://doi.org/10.5430/afr.v8n2p1>
- Lusk, E., & Halperin, M. (2015). An empirical contextual validation of the Capital Cube™ Market Trading variables as reflected in a 10-year Panel of the S&P500: Vetting for inference testing. *Accounting and Finance Research*, 5, 15-26. <https://doi.org/10.5430/afr.v5n1p15>
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting*, 4, 451-476. [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., & Winkler, R. (1982). The accuracy of extrapolation (time-series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1, 111-153. <https://doi.org/10.1002/for.3980010202>
- Makridakis, S., Hibon, L. E., & Belhadjali, M. (1987). Confidence intervals: An empirical investigation of the series in the M-Competition. *International Journal of Forecasting*, 3, 489-508. [https://doi.org/10.1016/0169-2070\(87\)90045-8](https://doi.org/10.1016/0169-2070(87)90045-8)
- Mykland, P., & Zhang, L. (2017). Assessment of uncertainty in high frequency data: The observed asymptotic

variance. *Econometrica*, 85, 197-231. <https://doi.org/10.3982/ECTA12501>

Wan, X., Yang, Q., Jiang, P., & Zhong, P. (2019). A hybrid model for real-time probabilistic flood forecasting using Elman neural network with heterogeneity of error distributions. *Water Resources Management*, 33, 4027-4050. <https://doi.org/10.1007/s11269-019-02351-3>

Appendix

Table A1 indicates the firms randomly selected from the S&P500. There were 16 firms and 32 TP-trials selected. For example, for CMI there were three [3] TPs flagged using the screen in Lusk & Halperin (2015).

Table A1. The S&P500 Tickers Selected Firms from the CCMNP

CMI[3]	CSX[3]	DOV[1]	ETN[2]	FLS[3]	GD[2]	GE[2]	HON[1]	MCO[2]	NOC[1]
PBI[1]	PH[3]	PWR[2]	R[2]	SNA[2]	TXT[2]	N = 32			

ⁱ <<https://www.chess.com/article/view/deep-blue-kasparov-chess>>

ⁱⁱ <<https://www.ibm.com/watson?S>>

ⁱⁱⁱ <<https://www.investopedia.com/terms/b/balancedscorecard.asp>>

^{iv} These three research reports provide a reiteration of the focus of our work and so another viewpoint. Duttler (2016), Mykland, & Zhang (2017). Wan, Yang, Jiang & Zhong, P. (2019).

^v For elucidation see the first three chapters in Brillinger (1974) where he discusses the statistical bases of point process estimation.

^{vi} In 2014 Prof Lusk received access to the complete Panel of the S&P500 of the CapitalCube™ Market Navigation Platform [CCMNP] a commercial product of AnalytixInsight™. Various tests of this dataset were performed over the years focusing on forecasting Stock prices just after a Turning Point—defined as a dramatic change in the trajectory of the stock price just after the turning point [TP]. In this research report, we used the same TP as the forecasting focus. For more information see: Lusk & Halperin (2015).

^{vii} Finally, while all the basic Model&Version&Horizon profiles discussed above are produced by a VBA:DSS, the ANOVA-decomposition was not programmed into the VBA:DSS as this would have required a contractual-monetary agreement with SAS™. However, there are many standard DSS-ANOVA-links available, some at no-cost, that are simple to integrate into the above mentioned VBA: DSS.

^{viii} The CCPL: Variable is: The CapitalCube™ Price Latest. This is an important variable in the CapitalCube Market Navigation Platform [CCMNP]. It was used extensively in a series of tests of the acuity of various variables in the CCMNP; see Lusk (2019) for more detailed information on the CCPL and its TP-Profile.

^{ix} The Exact T-K link-profiles[A-B-C-D] summarized in the paper are: [Y:X, Excel{A}]; then → [TS,Excel{B}]; then → TS,FE{B&C}; Y:X,FE{B&C}]; then → [TS,RE{C}; Y:X,RE{C}].

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).