

Analytical Procedures Phase of PCAOB Audits: A Note of Caution in Selecting the Forecasting Model

Mohamed Gaber¹, Edward J. Lusk¹

¹The State University of New York (SUNY) at Plattsburgh, Plattsburgh, NY, USA

Correspondence: Edward J. Lusk, the State University of New York (SUNY) at Plattsburgh, Plattsburgh, NY, USA.

Received: October 17, 2017

Accepted: November 15, 2017

Available online: November 23, 2017

doi:10.11114/afa.v4i1.2811

URL: <https://doi.org/10.11114/afa.v4i1.2811>

Abstract

The best-practices execution of PCAOB audits requires the use of Analytical Procedures at the Planning and the Substantive Phases. This often finds the auditor using the standard OLS two-parameter linear regression forecasting model [OLSR] to project account-values from the Planning Phase to balances expected at Year-End so as to effect a variance analysis at the Substantive Phase. This is the point of departure of our study. We examine the practical effect of using the OLSR model in a time-series context of the audit. Specifically, this research report provides information on the use of the OLSR model as the model of choice in the audit context compared to the ARIMA(0,2,2)/Holt model which is usually the standard choice for an exponential smoothing model in the presence of autocorrelation of data in the time-stream; autocorrelation is the usual case for longitudinal series taken in the audit. Results: We find that there are reasons to condition the selection of the forecasting model in the Analytical Procedures context based upon autocorrelation in the data-stream. When the time-stream of data exhibits autocorrelation the OLSR model fails in a statistically significant manner to capture the next or one-period ahead client value at the same rate as does the ARIMA/Holt model. This then has implications for the False Negative Investigation Error.

Keywords: big-data, Holt, forecasting confidence intervals

1. Forecasting as the Principal Platform in Forming an Analytical Procedures Investigation Projection

1.1 Context

The Public Company Accounting Oversight Board [PCAOB] Requires the Use of Analytical Procedures [AP] in the Planning and the Substantive or Completion phases of certification audits. As noted by Arens, Elder, Beasley & Hogan (2015, p.193):

“Analytical procedures are required at the planning phase as part of Risk assessment procedures to understand the client’s business and industry and to assist in determining the nature, extent, and timing of audit procedures.

Analytical procedures are also required during the completion phase of the audit. Such tests serve as a final review for material misstatements or financial problems and help the auditor take a final “objective look” at the audited financial statements.

Analytical procedures performed during the audit planning generally use aggregated data to help understand the client’s business and identify areas where misstatements are more likely. In contrast substantive analytical procedures used to provide audit evidence require more reliable evidence, often using disaggregated data for the auditor to develop an expectation of the account balance being tested.”

This linkage between the Planning and the Substantive phase all but proscribes that there is a projection at the Planning stage, typically formed analytically using previously reported and certified past data and forecasted as an expectation for the client’s Year-End value. The modelling form used to generate this projection is very often the standard linear regression model. In fact, the AICPA (2012) initiative: *The Clarity Project* dealing with Analytical Procedures presents a carefully detailed study called the *On the Go Stores* case analysis where the forecasting technique used is the standard regression model.

After the forecasting projection is made the auditor may examine the difference between the AP forecasting expectation,

E , and the actual client's account value, A . This is usually referred to as the disposition phase of the AP protocol where if the magnitude of the directional difference between E & A is relatively large, then the auditor is required to consider the meaning of this difference, note in the working papers the likely reason(s), and finally to justify the decision whether or not to further investigate the account.

Typically in our experience and that of our colleagues in the audit world, for PCAOB, Private Company, as well as Internal Audits, the Expectation, E , is formed by using the standard OLS two-parameter linear regression forecasting model [OLSR] to project expected account values from the Planning Phase to balances expected as of Year-End of the audit so as to effect a disposition analysis at the Substantive Phase.

1.2 Model Appropriateness

The Type 3 Error Despite the wide-spread use of the OLSR as the forecasting model of choice for the AP projections, theoretically the OLSR is not well suited for analyses where the audit datasets are time-series. The reason for this is there is a mismatch in the assumptions underlying the OLSR and the nature of the data in a time-series. Specifically, OLSR assumes that the longitudinal set of error terms—i.e., the result of the filter of the OLSR applied to the time-series—is composed of independent and identically distributed Normal random variates (Tamhane & Dunlop (2000, p.347)); however, for most economic time-series the data points in the economic time stream are produced by a relative consistent economic generating functions/processes so that the measured data points are associated in a Pearson sense over time—called autocorrelation. This mismatch leads to a Type 3 error or model misspecification often characterized as selecting the wrong model for the decision-making context. Avoiding model misspecification can be rather complicated. For example, Collopy & Armstrong (1992) formed a forecasting model that blends forecasts from the OLSR, exponential smoothing models and a Radom Walk projection to arrive at a more accurate forecasting model. Findley (2007) addresses the misspecification problem by comparing various models to identify the most acceptable modeling configuration in a time-series domain. Finally, Allam, Gagan, Bahaa & Maher (2012, p.42) note: *The autocorrelation problem in the model emerges when the two neighboring scenes are correlated, and that influences the credibility of the model.* They detail the modifications that were needed to re-cast the model to produce more relevant information for their study. Fortunately there are other classes of models that were developed to deal with time-series datasets where autocorrelation is produced by the data generating processes called: **AutoRegressive Integrated Moving Average** models: [ARIMA (p,d,q)]. These models are sometimes referred to as Box-Jenkins Models (1970) [BJ] and are refinements on groundbreaking work of Holt (1957) and Brown (1956) in the 1950s. At this point an illustration of the Type 3 error and a simple alternative to the OLSR will serve as a motivation for our study.

1.3 Illustrative

Example of the Type 3 Error Assume that we have selected a time-series of Account Receivable net balances at the end of each quarter, $n = 30$, over about seven years as detailed in Figure 1.

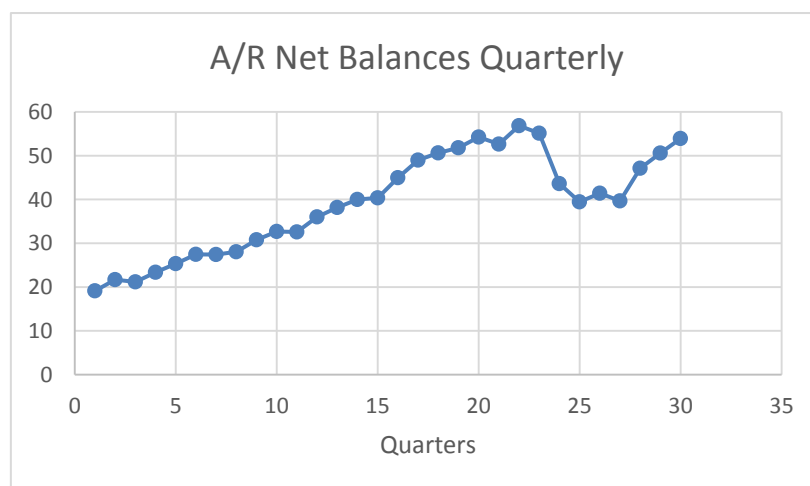


Figure 1. Illustration of the Time Stream for AP Testing

The Pearson autocorrelation [AC] over the first six progressive-differences, i.e., lags, has the range of [87.4% at Lag[1] to 36.8% at Lag[6]] in a classic decreasing profile indicative of an autoregressive process. The first to two AC values are positive and outside the BJ 95% confidence bounds this is a strong indication of positive AC. Further, all six of the Ljung-Box Q (1978) p-values [LBQ] are <0.0001 providing confirmatory evidence of an autocorrelated time stream. After we filtered or fitted the ARIMA(0,2,2)/Holt [Hereafter the Holt] to this data, the residuals had the following profile: The AC over the six first-differences, i.e., lags, has the range of [-33.0% at Lag[1] to 30% at Lag[6]] with no

discernable pattern. The first to two AC values are not outside the BJ 95% confidence bounds this is an indication of a failure to reject the usual Null. Further, all six of the Ljung-Box Q p-values are >0.05 providing confirmatory evidence of a non-autocorrelated time stream. The simple de-brief is that the Holt model captured all the statistically relevant information in the time-stream as the residuals were devoid of structure—i.e., noise.

However, if we had ignored the AC information of the time-stream and used the OLSR to form a model of the data then the residuals of the OLSR filter produced the following profile: The AC over the six lags has the range of [84.5% at Lag[1] to -21.5% at Lag[6]] in a decreasing profile. The first to two AC values are positive and outside the BJ 95% confidence bounds this is a strong indication of positive AC. Further, all six of the Ljung-Box Q p-values are <0.0001 providing confirmatory evidence of an autocorrelated time stream. The simple indications are: (1) that the OLSR model left structural information in the residual time-stream and so was not the correct or proper filter, and (2) the Holt model performed the task as expected and left the residuals as a time-series with no evident structure. In this case, if the auditor were to have selected the OLSR model, this would have been a Type 3 error.

1.4 Summary

This is a simple example of the disparity created by making the Type 3 error that we shall note as: **Ignoring the pre-filtering profile and selecting the “wrong” model.** The pertinent question is:

If the auditor does make the above Type 3 miss-cue, to what extent is there jeopardy as far as the audit is concerned?

This is the point of departure of this study. Given the above information, and our experience where we note that there is a predilection for the use of the OLSR in making AP projections, we offer information on the critical question of interest:

In the presence of autocorrelation in the data time-stream is the Analytical Procedures’ inference using the OLSR at variance with that of the theoretically correct Holt Class of forecasting models?

Following we will:

1. Specify an analytical procedures protocol that one may glean from AS 5 that is likely to pass the test of the PCAOB regarding the creation of audit evidence,
2. Discuss the choice of the time-series datasets in the audit context and examine the variations of the OLSR inferences possible in the audit context,
3. Detail a protocol rationalizing the choice of the Holt model in the audit context that will serve as the inferential benchmark for the OLSR,
4. Present the inference protocols that will be used to examine the question at issue,
5. Examine the comparative results, and
6. Summarize the results and make recommendations.

2. Forecasting Projections: From the Planning Stage to the Substantive Stage

Analytical Procedures Forecasting Protocol As a summary of the context, at the Planning Stage of the Audit the In-Charge collects data from the past performance of the client. This longitudinal dataset—i.e., a single firm time-series Panel, is inputted into a forecasting model and a projection is made for the expected value at Year-end—e.g., 31Dec. Then, if the actual client value IS in a reasonable confidence interval as produced by the forecasting model, the likelihood is that Extended Procedures [EP] are not warranted. If, on the other hand, the client value is NOT in the reasonable confidence interval as produced by the forecasting model, then the likelihood is that EP are warranted. In this context, what seems to be often used is the 95% confidence interval [CI] although the 80% CI and 90% CI are also found not infrequently. Also we have never seen a one-sided test used in the AP protocol.

3. Model Triage: The OLSR or the Holt? It Is More Than a Coin-Flip

Forming the Test-set The driver of the selection of the forecasting model is the nature of the Time-Stream. *A-priori*, the likelihood is that economic data selected at equally spaced intervals over time will exhibit association. If the modeling format is time-series, meaning the X-Variate is the time index, then the Holt model should be the likely choice for the auditor to make. However valuable such intuitions may be, it is recommended that an empirical test be used in creating audit evidence from a forecasting model to be used in the service of AP. Following on the A/R(net) balances illustration above, we suggest testing the autocorrelation of the residuals of the two models so as to make a reasoned decision as to selecting an appropriate forecasting model to execute the projective AP trail. Specifically, we have had success with using the Ljung-Box Q p-values [LB] in the following: *Model Knock-Out Protocol*: [MKoP]

For the first six-lags, if more than three LB p-values of the residuals for the time stream under audit are < 0.01 , then there is likely a sound basis to reject such a model for use in the Analytical Procedures Phase.

We examined the datasets that we accrued [See endnote iv] and based upon this and our related experience, we offer the

following conditionals for the expected incidence:

MKoP[1]: If both the OLSR and Holt fit the MKoP, then look for another Model [Our experience: *This happens rarely*];

MKoP[2]: If both the OLSR and Holt do not fit the MKoP, then use either Model [Our experience: *This happens rarely*]

MKoP[3]: If the OLSR does not fit the MKoP and the Holt does fit the MKoP, then use the OLSR [Our experience: *This happens very rarely*]

MKoP[4]: If the OLSR does fit the MKoP and the Holt does not fit the MKoP, then use the Holt Model [Our experience: *This happens the preponderance of the time.*]

4. The Models Used in the Testing: Four Versions of the Confidence Interval Testing

There are three frequently used variations on the OLSR model regarding the creation of Confidence Intervals [CI] whereas the Holt has effectively one CI form. To have the richest test, we will use the three OLSR and the one Holt inference intervals pitched at the 95% level of confidence. Consider now these four testing modalities.

4.1 OLSR Inference from the Excel Parameter Range Model

The *Excel*TM Regression functionality of the *Data Analysis Tab* allows the creation of a “wide-covering” confidence interval. These are effectively extreme case CI-scenarios as they are produced from the crisp-end points of the 95% CI for the intercept and the slope jointly, but are unconditioned. Here we offer the following notation:

Extreme Left Side [LowerLimit[LL]] 95%Boundary for actual client YE-value: $Y_{[t=(\Omega+1)]}$

$$\hat{Y}_{LowerLimit:(t=(\Omega+1))} = [\hat{\alpha}_{\Omega} - [t_{\frac{\alpha}{2}} \times s_{\epsilon(\alpha)}]] + [\hat{\beta}_{\Omega} - [t_{\frac{\alpha}{2}} \times s_{\epsilon(\beta)}]] \times [\Omega + 1] \tag{1}$$

Extreme Right Side [UpperLimit[UL]] 95%Boundary:

$$\hat{Y}_{UpperLimit:(t=(\Omega+1))} = [\hat{\alpha}_{\Omega} + [t_{\frac{\alpha}{2}} \times s_{\epsilon(\alpha)}]] + [\hat{\beta}_{\Omega} + [t_{\frac{\alpha}{2}} \times s_{\epsilon(\beta)}]] \times [\Omega + 1] \tag{2}$$

Where: $t_{\frac{\alpha}{2}}$ is the t-statistic for inference that has $df = [(\Omega)-2]$; Ω is the last time index in the data-stream—i.e., The last data point or N.

4.2 OLSR Inference from the Expected Random Effects Model

The OLSR assumption, in this case, is that there is random sampling possibility from a well-defined population of variable data streams of realizations, i.e., the set of:

$$\{Y_t^l ; l = Firm Index where l is large\}$$

which constitutes a large blocked or stratified group of “like” firms. In this case, the auditor 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. In our experience, the RE assumption is possible though not likely in the audit context. Nevertheless, in the service of completeness, the confidence interval for the client value in the RE context is:

$$\hat{Y}_{(t=\Omega+1)} \pm t_{\frac{\alpha}{2}} \times s_{\epsilon} \times \sqrt{\frac{1}{\Omega} + \frac{[(\Omega+1)-\bar{\Omega}]^2}{S_{tt}}} \tag{3}$$

Where: $\hat{Y}_{(t=\Omega+1)}$ is the value of the fitted regression projected for the next X-index time-point, $\Omega + 1$, using the parameters produced from the OLSR fit for the Ω data points; $S_{tt} = [\sum t^2 - [(\sum t)^2]/\Omega]$; $s_{\epsilon} = OLS[\Omega]:MSE^{1/2}$; and $\bar{\Omega}$ is the Mean of the time index for the Ω data points.

4.3 OLSR Inference from the Fixed Effect Projection

The assumption is that the object of interest is the j^{th} firm with a single set of data stream realizations: i.e., the set of:

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

Where this j^{th} 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 j^{th} 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:

$$\hat{Y}_{(t=\Omega+1)} \pm t_{\frac{\alpha}{2}} \times s_{\epsilon} \times \sqrt{1 + \frac{1}{\Omega} + \frac{[(\Omega+1) - \bar{\Omega}]^2}{S_{tt}}} \tag{4}$$

Adding the unit constant may seem a trivial modification but this increases the width of the confidence interval compared to that of EQ[3].

The next critical issue is to profile the precision, defined as the 50% of the width of the CI, of these three CIs so as to make a reasoned ordering of them for judging the reasonability of the time-series accrued for testing our question of interest. In this regard, it is obvious that the precision of the Random Effects [RE] CI:

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

is always less than the precision of the Fixed Effects [FE] CI:

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

as the latter has a unit value, i.e., 1.0 in the root function, and all the other values in the root function are positive.

As for the Excel precision: [EQ[2] – EQ[1]]/2 it is not clear for data series often found in the PCAOB account milieu, what the precision relationship is for the Excel precision compared to that produced by EQ[3] & EQ[4]. One supposes as EQ[3] is a mean projection that the precision for the Random Effects: CI will be less than that of the Excel CI precision as EQ[1] and EQ[2] are driven by the end-points of the standard errors of the parameters. Finally, for the Excel precision compared to that of the Fixed Effects EQ[4] based upon our experience, we proffer that more than 50% of the time the Excel’s Precision will be greater than the FE’s Precision.

Summary expectations for the relative precision of the three OLSR CIs:

- 1.) FE > RE; 2.) Excel very likely > than the RE & 3.) Excel more than 50% of the time > than the FE.

We shall use these relationships as a reasonability check on the accrued time-series that we will use in the test of the integrity of the Holt model.

4.4 Inference from the Holt Projection

The Holt model is rather complicated viewed in the relief the OLSR inferential versions. For this reason we offer the excellent URL link where the Holt model is presented and detailed with an illustrative example: [<https://www.otexts.org/fpp/7/2>]. The Holt model and the related 95%CI that we are using are in the *Time-series* platform of the JMP™v.13/SAS™. See JMP (2005).

4.5 Illustration of the Computational Details

Assume that we have selected a time-series of Account Receivable net balances at the end of each quarter, n =30, [The Client YE-Value was **52.96**] over about seven years as detailed in Figure 1 above¹.

4.5.1 Populating the Excel Equation Set

N= Ω =30 and the t-value is 2.048407141795 noted as 2.05

Intercept: The Lower Limit = 21.87 – [2.05 × 2.32] = 17.12

Intercept: The Upper Limit = 21.87 + [2.05 × 2.32] = 26.62

Slope: The Lower Limit = 1.12 – [2.05 × 0.13] = 0.85

Slope: The Upper Limit = 1.12 + [2.05 × 0.13] = 1.38

Thus the 95%CI Limits for the Forecast at the next period—i.e., the client value are:

¹Here are the longitudinally ordered values in this time-stream in the case that the reader would care to verify the computations. Assumed A/R series found in Figure A [Series 10 in the Collopy & Armstrong Download], n=30 [The Client Holdback is: 52.96] is:

19.13	21.67	21.17	23.37	25.36	27.45	27.41	28.06	30.77	32.69
32.59	36.02	38.16	39.99	40.38	44.98	48.99	50.62	51.79	54.25
52.66	56.86	55.11	43.62	39.45	41.42	39.67	47.12	50.61	53.93

EQ[1] Lower95% Confidence Interval = $17.12 + [0.85 \times [30 + 1]] = \mathbf{43.45}$

EQ[2] Upper95% Confidence Interval = $26.62 + [1.38 \times [30 + 1]] = \mathbf{69.52}$

Summary: 52.96 \subset [43.45 to 69.52]

4.5.2 Populating the OLSR [REffects] Equation Set

$$\mathbf{51.74} = 56.48 - [2.05 \times 6.19 \times \sqrt{\frac{1}{30} + \frac{[(31)-15.5]^2}{2247.5}}]$$

$$\mathbf{61.23} = 56.48 + [2.05 \times 6.19 \times \sqrt{\frac{1}{30} + \frac{[(31)-15.5]^2}{2247.5}}]$$

Summary: 52.96 \subset [51.75 to 61.23]

4.5.3 Populating the OLSR [FEffects] Equation Set

$$\mathbf{42.95} = 56.48 - [2.05 \times 6.19 \times \sqrt{1 + \frac{1}{30} + \frac{[(31)-15.5]^2}{2247.5}}]$$

$$\mathbf{70.02} = 56.48 + [2.05 \times 6.19 \times \sqrt{1 + \frac{1}{30} + \frac{[(31)-15.5]^2}{2247.5}}]$$

Summary: 52.96 \subset [42.95 to 70.02]

4.5.4 Finally, the Holt 95% CI Produced by JMPTMv.13/SASTM

Summary: 52.96 \subset [49.02 to 65.48]

5. The Research Testing Protocol for Benchmarking the OLSR Inference Variations

We are interested in testing the inference differences of the OLSR Models v. those of the Holt model for the AP phase of the audit where a forecast is the driver of the decision to effect an analytical procedures protocol. To reiterate the testing context:

Assume that the In-Charge has a forecast and the 95% CI derived only using model generated² information. Then if the Client Substantive YE-value is in the CI, extended procedures are not likely to be entertained. If the Client Substantive YE-value is not in the CI, extended procedures are more likely to be effected. In this context, one needs the Client Substantive values to conduct the benchmarking test. Unfortunately we do not have this empirical information. As a reasonable surrogate, we have selected a set of time-series data where there are holdback values. In this case we will use these datasets and select the first holdback value to be the client YE-value. Further, given the nature of the study we have selected series that seemed to be appropriate for filtering using the Holt³ model and also the OLSR. The inferential test will use the CI inclusion rates to test the question of interest.

5.1 The Datasets

We have selected the time-series set used in the Makridakis et al. (1982), Collopy & Armstrong (1992) and Adya & Lusk (2016) studies. These time-series have been used in a large number of studies over the years and in this sense have been vetted for inclusion in our testing protocol. Specifically, there are 181 time-series of economic and demographic time-series data. For our study, we have selected as a random probability sample, i.e., with replacement, 106 series. In this regard, we used the RANDOMBETWEEN(1,181) functionality in *Excel*TM to form the random numbers for selecting the series. Of the 106 series⁴ used there were 82 unique series and 24 duplicates.

5.2 The Selected ARIMA Model

We selected the ARIMA (0, 2, 2) or the Holt Model as it is the simplest model with a level/intercept and trend/slope

²This is to say that there are no rule-based judgmental adjustments to the forecasts the sort of which are detailed by Collopy & Armstrong (1992).

³See Collopy & Armstrong (1992) C&A who also have selected the Holt and the OLSR models to initialize their judgmental Rule-Based Forecasting protocol.

⁴The enumeration of the series as originally scripted in the above cited studies is given following. The series noted in Bold are those that have duplicates. Series 178 had two duplicates: 1,4,6,9,10,13,14,16,17,18,**20,22**,23,24,28,29,31,**43,45,49,50**,51,52,54,**56,57**,62,63,64,65,66,67,71,73,74,76,78,80, 81,86,87,**94**,97,99,100,101,102,103,**105**,108,**110**,114,**118,120**,121,**122**,123,**124**,127,132,138,**141**,142,143,**146,150**,151, 152,153,154,**157**,161,162,**165,167**,170,171,176,**177,178**,179,**181**.

component in the ARIMA model class. Also this model is, in most cases, one of the platforms in the commercially available statistical software. We have the *JMP*, v.13 of the SAS Institute and the Holt model has been included in this software over the last three decades. Finally, the Holt model was selected by Collopy & Armstrong (1992) as one of three models to form the basic platform in their Rule Based Forecasting system; one of the other models in their platform was the OLSR model. For the OLSR information set we used the REGRESSION platform in the *DataAnalysis* platform in *Excel*(2016).

5.3 A-Priori Expectations

The operational hypothesis to address the research question of interest is:

H [OLSR v. Holt] The Test *Null* will be the usual version for the test of proportions:

There will be NO expected differences between the percentage of client YE-account values IN the 95% forecasting confidence intervals as produced by the three OLSR versions mentioned above compared to those produced by the Holt.

We will produce the various 95%CI and determine if the Client Value is in the Interval. If so this will be recorded as a success. We will use this Inclusion Percentage as the principal test of *H*. As a related info-set we will report on the relative precision of these four CIs as it may be necessary to condition the interpretation of the test of *H* on the relative precisions.

5.4 Results

The results are presented in Table 1.

Table 1. Analysis of the Models as the Filters of the Sample Time-series, n = 106

Models	Inclusion 95%CI	%Relative Precision	Holt Ratio Benchmark
REffects	40.6%	8.6/5.7	0.83
Excel	84.9%	23.7/15.8	2.30
FEEffects	80.2%	20.3/14.0	1.84
Holt	95.3%	15.6/9.0	N/A

5.5 Discussion

Initially, we offer the profile on the relative precisions of the time-series where we proffered the following relationships:

1. FE > RE [Test Result: 100% of the Cases];
2. Excel > RE [Test Result: 100% of the Cases] &
3. Excel > FE [Test Result 82.0% of the Cases].

For the third comparison, the lower limit of the 95%CI is 75.9% and excludes 50% suggesting that the third vetting conditional was reasonably the case. *Conclusion* The accrual of the time-series seems appropriate for the test of *H*.

For purposes of analysis and testing of *H*, we are assuming that the appropriate model for the AP analysis is the Holt. This seems justified in for all but three of the 106 randomly accrued time-series MKoP[4] obtained. Thus using the Holt as the gold standard and assuming that there is no differential bias over these four models as far as the False Negative Investigation Error [FNIE] or the False Positive Investigation Error [FPIE] are concerned, we can directly draw the inference to test the question of interest, the *Null* of *H* which, here re-expressed, is:

There is NO evidence that percentage of time that the Client YE-Account value is IN the 95%CIs produced individually by the three OLSR model forms used to make forecasts from the Planning Phase to the Substantive Phase will be different from those using the Holt model?

The inferential indication is the Inclusion Percentage over the four models—i.e., Col2 in Table 1. The Holt model captures the Client Value 95.3% of the time; note that this is what one would expect for the 95% CI. This is certainly not unexpected as most all the data in the randomly selected series exhibited clear autocorrelation. This result is certainly an additional validation of the appropriateness of the random sample of accrued time-series and also of the logic of using the Holt results to judge or benchmark the OLSR results. As for the relative statistical directional comparisons from Table 1:

Holt v. RE: [95.3% v. 40.6%]: test p-value = < 0.0001

Holt v. Excel: [95.3% v. 84.9%]: test p-value = < 0.006

Holt v. FE: [95.3% v. 80.2%]: test p-value = < 0.0004

In simple summary, *there is no evidence that the OLSR outperformed the Holt Model relative to the inclusion of the client value in the Substantive Phase 95%CI.*

5.6 Testing a Possible Coincidental or Anomalous Occurrence

A question begged by these test results for H is: *Perhaps the reason for these results is that the precision of these models is biased in favor of the inclusion for the Holt model—i.e., the Holt model produces very wide 95% CIs?* This test information is produced in Table 1 Col 3 where the Mean & the Median [Mean/Median] are presented. This is the relative benchmarked precision as computed as:

$$\text{Client Value} / [(\text{Upper Limit of the 95\% CI} - \text{Lower Limit of the 95\% CI}) * 50\%]$$

The results are also clear. The RE interval is relatively smaller, as expected, than the others and this explains the low inclusion percentage. In this case, even if the In-Charge could rationalize a Random Effects' context, which in most audits would be problematic, the RE model is not a reasonable inferential choice. As for the Excel and the Fixed Effects CIs, both of which find currency in use in the audit context, their relative precisions are wider not smaller than that of the Holt: 23.7% & 20.3% respectively compared to the Holt of 15.6%. In fact these relative precisions compared to the Holt's have p-values of: < 0.009 for the Excel result and < 0.07 for the Fixed Effects. Therefore, there is a strong indication⁵ that the Holt inclusion results are not due to a relatively larger 95% CI.

6. Overall Summary and Recommendations

6.1 Summary

There is clear evidence that when there is a single firm Panel—i.e., a longitudinal segment of data that qualifies as a time-series which exhibits, according to the MKoP, autocorrelation that using the OLSR models in any of the three forms is not appropriate in the audit context. In fact, of the least inappropriate model forms, those of the Excel & Fixed Effects OLSR models will likely create more False Negative Investigation Errors [FNIE] in the range of [10.4% [95.30% – 84.90%] to [15.1% [95.30% – 80.20%]]. This failure to consider EP investigations may likely unnecessarily increase the risk of failing to investigate when appropriate and so possibly compromise the veracity of the audit opinions.

6.2 General Recommendations

It is critical to test for autocorrelation when the In-Charge selects a time-series to form the forecast for the AP testing at the Substantive Phase. In these cases the MKoP seems a simple test for guiding the model selection process. There is a high risk relative to the FNIE in making the Type 3 error—i.e., using the wrong model. In the future, we would hope that actual audit datasets could be collected and used to re-test the MKoP as we used time-series data from a forecasting context. Additionally, we offer that the recent publication Lusk (2017) where a DSS is presented that offers a holdback-vetting protocol focused on the OLSR is best conditioned by a pre-screening phase to consider the results of the MKoP triage as a precondition to the launching of the holdback-vetting phase. This should further enhance the applicability of that modeling system. Finally, one hopes that such research results will be considered by the PCAOB, the SEC & the AICPA in providing guidance as to the forecasting model selection protocols for the executing the Analytical Procedures phase of the audit.

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 [In particular: Professors: Neal Duffy, Razvan Pascalau & Dhimitri Qirjo of the Economics' Group] and Mr. Manuel Bern, Chief of Internal Audit: *TUI International, GmbH*, Hannover, Germany for their careful reading, helpful comments, and suggestions. Finally, we appreciate the suggestions of the reviewers of *Applied Finance and Accounting* that improved the nature of this paper.

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>
- AICPA. (2012). Audit Guide: Analytical Procedures, *American Institute of Certified Public Accountants, Inc.* New York, NY USA: https://competency.aicpa.org/media_resources/208125-analytical-procedures-aicpa-audit-guide & <<http://www.aicpa.org/InterestAreas/FRC/AuditAttest/Pages/ImprovingClarityASBStandards.aspx>>
- Allam, M. M. H., Gagan, K., Bahaa, S. A. L. A., & Maher, M. D. (2012). The auditing quality and accounting conservatism. *International Management Review*, 8, 33-50.
- Arens, A., Elder, R., Beasley, M., & Hogan, C. (2015). *Auditing and assurance services: An integrated approach*. Pearson Publications. ISBN: 13:978-0-13-406582-3.
- Box, G. E. P., & Jenkins, G. (1970). *Time-series analysis: Forecasting and control*. San Francisco: Holden-Day. ISBN:

⁵The Ratio results in Col 4 of Table 1 are confirmatory to this analysis.

0-8162-1104-3.

- Brown, R. (1956). *Exponential smoothing for predicting demand*. Arthur D. Little Inc. Cambridge, Massachusetts, USA.
- 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>
- Findley, D. (2007). Optimality of GLS for one-step-ahead forecasting with REGARIMA and related models when the regression is Misspecified. *Econometric Theory*, 23, 1083-1107. <https://doi.org/10.1017/S0266466607070430>
- Holt, C. (1957). Forecasting trends and seasonal by exponentially weighted averages. *Publication Memorandum: Office of Naval Research, Folio: 52*.
- JMP®6. (2005). *Statistics and graphics guide*. SAS Institute Inc. ISBN 1-59047-816-9.
- Lusk, E. (2017). Analytic Procedures: A holdback-vetting forecasting model. *Applied Finance and Accounting*, 3, 65-74. <https://doi.org/10.11114/afa.v3i1.2139>
- Ljung, G., & Box, G. E. P. (1978). On a measure of a lack of fit in time-series models. *Biometrika*, 65, 297-303. <https://doi.org/10.1093/biomet/65.2.297>
- 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>
- Tamhane, A., & Dunlop, D. (2000). *Statistics and data analysis*. (1st ed.). Prentice Hall, Upper Saddle River, NJ USA: ISBN: 0-13-744426-5

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](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.