

An Associational Examination of the CapitalCube™ Effect Context for the MPV over the Linguistic Partitions: Testing Sensitivity & Specificity

Edward J. Lusk¹ & Michael Halperin²

¹ The State University of New York (SUNY) at Plattsburgh, 101 Broad St., Plattsburgh, NY, USA & Emeritus, Department of Statistics, The Wharton School 3730 Walnut St., University of Pennsylvania, Philadelphia, PA, USA

² Director Lippincott Library of the Wharton School [Retired], 3420 Walnut St. University of Pennsylvania, Philadelphia, PA, 19014, USA

Correspondence: E. Lusk, SBE SUNY Plattsburgh, 101 Broad St. Plattsburgh, NY, USA 12901.

Received: January 11, 2017

Accepted: February 20, 2017

Online Published: February 28, 2017

doi:10.5430/afr.v6n2p1

URL: <https://doi.org/10.5430/afr.v6n2p1>

Abstract

In this third examination of the CapitalCube™ Market Navigation Platform [CCMNP] we have selected the previously vetted set of embedded variables: Market Performance Variables [MPV] for their Linguistic Qualifiers [LQ] considering their directional market effects or MPV[LQ[*Neutral: Unfavorable: Favorable*]]]. In the testing, we are interested in the *Sensitivity* and the *Specificity* of these vetted variables over the annual S&P500 Panel from 2005 to 2013. The inference framework employed a Median Split: **High** or **Low** for each of the 13 MPV tested and a random selection to avoid the FPE-jeopardy that is part of the Chi2 testing model. We used the Tamhane & Dunlop cut-off to identify Chi2 cells effects of interest and used these to develop the *Sensitivity* and the *Specificity* tests. Results: We were able to reject the *a priori* Nulls proffered for the testing protocols indicating that one may reject the supposition that the labeling of the LQ is formed by random processes in the CCMNP.

Keywords: Sample Effectiveness, FPE Screening Jeopardy

1. Introduction: Setting the Analytic Context

1.1 Previous Research Summaries

To date, in two research reports, the CapitalCube™ market navigation platform [CCMNP] *CapitalCube™* [<http://www.capitalcube.com/>] has been investigated by Lusk & Halperin [L&H] who have examined the nature of many of the constituent variables that are offered by the *AnalytixInsight™* [<http://www.analytixinsight.com/>] group. These examinations have not addressed the market inferential information where the longitudinal historical information as recorded in the CCMNP is used to forecast the trajectory of the stock from the past information to the end of reducing the false positive signals as to direction expectations. Rather L&H in Lusk & Halperin (2015) used the following two summary market variables that are usually readily available as they are reported information from the trading exchange:

1. [Fifty-Two Week *Low*] & [Fifty-Two Week *High*];

Also this market context for the rolling range of the bell prices is contexted by the CapitalCube Price range as described in the following variables:

2. [CapitalCube Price Range *Min*] & [CapitalCube Price Range *Max*].

In this case, the CapitalCube Price is formed by various manipulations of the constituent variables of the CCMNP. We were not provided with the details that create the measured and reported values for the CapitalCube Price; however, we were given an overview of the essentials that create the *CapitalCube Price Latest* [CCPL]. Specifically, CCPL is a projective rolling variable—i.e., longitudinal—adjusted for Split/Spins, and benchmarked by a large number of market performance measures.

Also L&H selected three other CCMNP Decision Making Variables [DMV]:

1. *Current Price Level Annual* [CPLA] this is a ratio of the closing market price to the range of previous trading-day values going back one year in time. In this case, the scale of the CPLA is effectively in the Unit interval.

2. *Scaled Earnings Score Average Latest* [SESAL]: this is an earnings variable from the SEC reporting requirements for listed firms. The CCMNP adjusts the reported earnings with as many as 50 re-calibrations firm measured markers such as *Return* and *Working Capital* to create SESAL as an aggregate longitudinal variable in the Range [1 to 100].
3. *Previous Day Closing Price Latest* [PDCPL]: is the stock bell-price as adjusted for Stock splits and any sort of Stock spin-offs going back a number of years.

In a 2015 research report L&H examine the associational aspects of these eight variables using both Spearman and Pearson associational measures. Additionally, they examined various factor or correlational-groupings using as the triage the usual eigenvalues over the factor space. In this regard, they used the standard cut-offs of $\sqrt{0.5}$. In conclusion they found in the associational domain where the Null was no Spearman or Pearson association that for a large number of tests there was consistent and expected directional association in the S&P500 Panel. In summary, L&H report that there is no reason to believe the eight constituent variables mentioned above are in total, or in major part, formed from random generating processes.

In the second paper [in this analytic series] Lusk & Halperin (2016) investigated the relationship of 13 selected Market Performance Variables [MPV] relative to their Linguistic Qualifiers [LQ]. Each of the MPV have a unique set of LQs; details on each are found on: <http://www.capitalcube.com/blog/index.php/glossary/>. In this endeavor, we benefited from the recent work of Pachamano & Fabozzi (2014) who offer that there are many aspects that are involved in the successful market navigation platform. Specifically, they suggest:

“Financially screening investment candidates typically relies on identifying important factors that influence investment performance. Factors can most generally be separated into two groups—fundamental factors and macroeconomic factors—but some asset managers consider additional alternative factor categories, such as technical factors, analyst factors, and social responsibility factors. -- - Macroeconomic factors can include gross domestic product (GDP), consumer sentiment index, business confidence index, investor sentiment index, and broad market indices such as the S&P 500 Index.” p.239

In addition, we recommend the following excellent treatments that deal with programming approaches to selection and partitioning of trading portfolios: Sevastjanov & Dymova (2009) and Petter, Tü tinc ü & Fabozzi (2014). These treatments were used in developing the selection of a full range of variables: CapitalCube Decision Making Variables [DMV] and the context variables, the Market Performance set of indicators, the Linguistic partitions, and finally the scoring that we did to give tri-orientation to the context of the LQ. This rich array of indications will, we believe, enhance the inferential information to be reported in this study.

For example, one of the MPV was *Accounting Quality* which had the following LQ: *Conservative Accounting, Aggressive Accounting, Sandbagging, & Non-Cash Earnings*. L&H used a set of experts and trained and qualified students to triage the LQ into one of three categories: {Neutral, **Unfavorable** or *Favorable*}. The intention of this research was to determine: (1) if the distribution of LQ on the scale: {Neutral, **Unfavorable** or *Favorable*} was differential from a random assignment, (2) if there was general agreement among Experts and Trained pre-professionals regarding these LQ indications, and (3) how often their agreement coincided with a second expert vetting. The results as reported in Lusk & Halperin (2016) is that one can, with confidence, reject the relative three Nulls strongly suggesting that there are inferential differences over the various LQ. This vetting rationalizes the use of the labels: {Neutral, **Unfavorable** or *Favorable*} in ferreting out differential indications over the MPV: LQ[DMV] imbedded set. This is the point of departure of this research report. In this paper, we will use the MPV set reported above as nuanced by their unique Linguistic qualifiers [LQ] relative to their directional indications {*Neutral: Unfavorable: Favorable*} now as profiled by the four decision-making variables [DMV]: CPLA, SESAL, PDCPL & CCPL.

2. Continued Analysis of the CapitalCube Navigation Platform: Vetting the Internal Consistency

2.1 Pr écis of this Research Report

We are interested in the *Sensitivity* and *Specificity* of the CapitalCube market navigation platform [CCMNP] as expressed over the S&P500 Panel as a validation indication for the relationship of the [LQ] for each of the Market Performance Variables over the four Decision Making Variables [DMV] CPLA, SESAL, PDCPL & CCPL relative to the directional indications: {*Neutral: Unfavorable: Favorable*}. In this research report:

Sensitivity of the CCMNP over the Panel will be indicated if the DMV are differentially distributed over the Median partition of their respective DMV.

Specificity will be indicated if the proportion of the Non-Neutral directional indications are relatively prevalent compared to the Neutral indications.

These tests will be effected by examining the relationships for a Median-Split triage: {*High: Low*} for the DMV formed at the end of the Panel—i.e., after all of the Panel data values have been recorded by CapitalCube.

This testing is not offered as an indication of the predictive validity of the CapitalCube navigation platform; we will address this aspect in a subsequent paper. In this paper, we are focused on the distribution of the performance profiles of these variable sets. The logic is: ***If there is no evidence of Sensitivity or Specificity for the DMV, for example, one cannot reject the random generating process as populating the CapitalCube Panel as expressed through the S&P500 over the accrual period for the DMV, then there is little need to examine the predictive nature of the CapitalCube navigation platform.***

The Inferential Basis There is NO clear statistical measure with False Positive Error [FPE], or for that matter, a False Negative Error [FNE] *a priori* distributional validity that surrogates for the “profile” as there is likely to be dynamic and non-idiosyncratic non-independence over various sets of MPV[LQ[blocked over the Median-split of: DMV[CPLA;SESAL;PDCLA&CCPL]] over [N;UF:F]; however, if the Null were to be the State of Nature, then there will not likely be numerous insidious structural associations that suggest *Sensitivity* or *Specificity*. For this reason, we will use the following *Sensitivity* and *Specificity* relationships which are founded on the inferential test of the Null:

- 1.) General profile displacement overall as the test of *Panel Sensitivity*,
- 2.) For *Specificity*, we will examine the percentage of {**Unfavorable & Favorable**} Chi2 effects relative to the percentage of Neutral Chi2 effects over the coded LQ designations where we proffer that *Specificity* will be in evidence if more of the {**Unfavorable & Favorable**} LQ have TD indications than do the Neutral LQ.

Panel Sensitivity Test H1 In this case, we will use the percentage of the differential frequency counts as measured by the Chi2 significance tests as a measure of interaction of the firms over of the Panel relative to the LQ for each MPV aggregated over the Panel for the Median blocking. Assuming the Null is the State of Nature, there should be very few Chi2-significant LQ-Median Partition effects. For this inference protocol, we will use the heuristic implied in the Tamhane-Dunlop (2000, p.324) [TD] Cell-Chi2 frontier value of 1.0 [rounded] as a suggestive proportional differential as between the Median partitioned profiles. This seems a reasonable surrogate for non-directional interaction and so sensitivity. As for the inferential test, we will set the expected proportion of TDs for the Null State of Nature at five times the random expectation of 5% or **25%**. This seems a strong test of *Sensitivity* and should allay any concerns as to pseudo-sensitivity given that we expect that there may be statistical non-independence over some of the LQ within some of the MPV for the Panel (Note 1). We will use the number of TD differentials overall as the computation of the association percentage and then test the 95% directional CI for inclusion of 25%. If the test expectation of 25% is in the observed 95%CI then this will not provide the rationale for rejecting the Null of no association.

Panel Specificity: H2 iff: H1 is founded—i.e., the Null of *H1* is rejected—then for *H2*: if the Null is the true State of Nature then we expect over the annual Panel which covers 2005 to 2013 which traverses the Lehman Bros.2008 sub-prime debacle that the market would be destabilized for a year or so and so bias the market in the direction of the Null. See Jorion & Zhang (2009), Allen & Carletti (2010) & Mishkin (2011)]. Given this, then there would be no reason to expect that *Specificity* should be in evidence. In measuring *Specificity*, we will take advantage of the scoring of the various LQ as reported by Lusk & Halperin (2016) scored as: {Neutral, **Unfavorable** or *Favorable*}. If the Null is to be the case, then we proffer that there should be no difference in the percentage of Chi2 effects for the LQ of the Variables of “Interest” [VoI]—those identified as {**Unfavorable** or *Favorable*}—compared to the percentage of Neutral Chi2 effects. *Rationale*: As the non-Neutral LQ are expected to draw more decision maker inferential attention compared to the Neutral LQ, we expect, in the *Specificity* case, that there would be a higher percentage of these with TD indications than for the Neutral LQ. This then will be the Null test. The Null will be rejected if we observe that the percentage of TD Chi2 indications for the LQ VoI is inferentially greater in percentage terms than the percentage TDs in evidence for the Neutral LQ over the MPV. This then will provide support for the *Specificity* of the MNP of CCMNP.

In this particular test frame so as to deal with a reasonably trimmed or Winsorized datasets, even though we are using a Median triage for the TD indications, after consultation with Mr. Gautam Pasupuleti, COO [gautam.pasupuleti@analytixinsight.com], we screened/window-blocked the MPV as follows:

2.2 Conditioning the Panel for the Internal Consistency Testing

To rely on the Median Triage we have screened values outside the CapitalCube range parameters. Specifically, for:

CPLA [We screened values > 1.0 or < than 0; there were only a few such values in number <0.01%]

SESAL [We screened a few values > 100; there were only a few such values in number <0.01%]

PDCPL & CCPL are non-isomorphic translations but there are none the less associational at least over sections of the Panel as we have seen above. In this case, based upon the advice of the CapitalCube, Mr. Pasupleti, COO, we screened for high side outliers as the PDCPL. Specifically, the recommended screen for the CCPL was 2 times the max[PDCPL] which was 2,697.68 [2 x 1,348.84].

This then will be the Panel expression that we will use in the testing for the internal consistency of the LQ for each of the MPV for each of the DMV.

3. The Selected Set of Market Performance Measures & Their Linguistic Qualifiers

We received from AnalytixInsight on 9 April 2015 a Panel-download of the CapitalCube variable set encoded in the S&P500 from 2005 to and including 2015. This Panel offered 20 MPV each of which has a unique set of LQ. We eliminated some of the MPV due to a death of measured values. See Lusk & Halperin (2015) for more details. Also, the annual section of the Panel is from 2005 to 2013. Therefore, we will use only this annuals section of the Panel as we need to have a consistent accrual for the Panel.

Chi2 Sample Protocol In this section we will detail the Chi2 Cell protocol. CapitalCube offers a different linguistic codex for each of the performance variables that are part of CCMNP. For the specific LQ: Testing protocol, we will use as the inference measure the Chi2 triage classification on a cell-by-cell basis using the Chi2 Cell Contribution [C2CC] as the measure of association triage strength. We will use the TD heuristic measure of 1.0 [Rounded] as the indication of an interesting partition effect. This raises an inferential issue; there is a well-documented sample size bias of the C2CC test relative to “erroneously” rejecting the Null due to FPE p-value sensitivity to the sample size. See Cho & Gaines (2007) and Lusk & Halperin (2014). As the sample size for the Panel is large, we are, after all, examining a Panel over ten years of monthly data for about 500 firms, we have scaled the sample size to not invite the FPE anomaly—i.e., *we believe that we can reject the Null of No-Association when in the practical context there is no convincing evidence of association.* Therefore, we created a random sampling of the S&P500 Panel where each cell grouping has an expectation of 220 label indications (Note 2). Additionally, as the S&P500 panel is not always populated by the CCMNP with LQ indications we factored up the sample size by 50% to fill the void and so enhance reasonable inference relative to *Sensitivity* and *Specificity*. For example, assume we had an MPV with four (4) LQ indications over the DMV for the High and Low Median Split (2). In this case the random sample would be n = 7,040 [Median Split[2] × LQs[4] × Missing Data Factor[2] × 220]. We will then also note the number of values on a cell-by-cell for which the Cell-Chi2 is greater than the TD frontier value of 1.0. We will be using these cell to examine the inference relationships on an intra-variable triage partition. Following we provide an illustration of the sampling logic needed to form a reasonable inference metric for vetting the DMV of the CCMNP.

Illustration of the Sampling Logic and the TD-measure

The following is a table where the four DMV are profiled for the MPV: *Upside: Downside Risk* which has three LQ: {Neutral, **Over-Valued** & *Under-Valued*} for the Median-splits of the DMV for the full-year panel: 2005 to 2013.

Table 1. Full Annual S&P500 Panel for *Upside: Downside Risk* A: Neutral B: **Over-Valued** C: *Under-Valued*

CPLA	A	B	C	Total	SESAL	A	B	C	Total	PDCPL	A	B	C	Total	CCPL	A	B	C	Total
	6575	9215	8551	24341		5924	10668	7589	24181		6334	8936	8980	24250		5971	12928	5421	24320
Low	58%	41%	56%	50%	Low	53%	48%	50%	50%	Low	56%	40%	60%	50%	Low	53%	58%	36%	50%
	158	333	128			19	14	0.6			87	437	303		20	295	611		
	4688	13063	6585	24336		5278	11488	7468	24234		4929	13342	5968	24239		5292	9305	9724	24321
High	42%	59%	44%	50%	High	47%	52%	50%	50%	High	44%	60%	40%	50%	High	47%	42%	64%	50%
	158	333	128			19	14	0.5			87	437	303		20	295	611		
Total	11263	22278	15136	48677	Total	11202	22156	15057	48415	Total	11263	22278	14948	48489	Total	11263	22233	15145	48641

Table Codex: For each cell, the first entry is the Count for the Median-split, the second number is the Column Percentage [rounded] and the last value is the Chi2 Cell Contribution as reported by SAS, *JMPv.12*. In this case, we have presented the full Panel for the MPV of *Upside: Downside Risk* for the Median splits for the four DMV for each of the LQ: {Neutral, **Over-Valued** & *Under-Valued*}. For 15 of 16 cells, there are very large C2CC values. However, even though the sample size is very large there is nonetheless a cell configuration SESAL:C where there is

no TD indication of a variation from the marginals. For the sampling results, we will note the Pearson Chi2 probability which is an overall indication of proportional variance compared to the marginals. This will be noted in the header in the Total column for each of the MPVs. For p-values less than 0.002, we will use an “*” as the designation.

Table 2. Sample Result for *Upside: Downside Risk A: Neutral B: Over-Valued C: Under-Valued*

	CPLA				SESAL				PDCPL				CCPL						
	A	B	C	*	A	B	C	*	A	B	C	*	A	B	C	*			
Low	970	1336	1227	3533	Low	883	1571	1112	3566	Low	936	1263	1274	3473	Low	890	1862	770	3522
	59	41	57	50%	54	49	52	51%	57	39	60	50%	54	57	36	50%			
	27	54	20		4	3	0.1		19	76	49		6	34	88				
High	664	1916	924	3504	High	745	1662	1029	3436	High	698	1989	836	3523	High	744	1384	1383	3511
	41	59	43	50%	46	51	48	49%	43	61	40	50%	46	43	64	50%			
	28	54	20		4	4	0.1		19	75	48		6	35	88				
Total	1634	3252	2151	7037	Total	1628	3233	2141	7002	Total	1634	3252	2110	6996	Total	1634	3246	2153	7033

Notice for the random sample of 7,040 that the percentages are essentially identical validating the expected sampling result. For example, the mean of the absolute differences in the percentages between Table 1 & 2 over the 12 instances was 0.0083. Also, the FPE relationships are essentially proportional to the ratio of the realized sampled values as reflected in the similarity of the proportional in Tables 1 & 2. This argues also for the heuristic that we have used where we set the sample at 7,040 which replicated the proportional profiles and also offered representative C2CC values.

4. The Effect Variable as Profilers for the CapitalCube Variables of Interest

Recalling the definitions of the various Linguistic Triage variables as presented Lusk & Halperin (2016), for each of the selected Market Performance Variables [MPV] with respect to the four Decision Making Variables [DMV] presented above we, will now examine the relative associations of the DMV for each of the MPV over their various Linguistic Qualifiers[LQ]. It is critical to bear in mind in reflecting on the following analytic profiles that this is an associational profiling where the *independent* action is the Median partition of the MPV at the end of the Panel i.e., after all the data variables are recorded. This Median partition creates the relative proportions, basically 50% for each, which are tested for TD significance for the reported partitions for the DMV.

For example, consider Table 2—the sample result for *Upside and Downside Risk*. For the CPLA there are 7,037 firms that have been given a CPLA value by the CCMNP in the random sample of 7,040. As expected, the Median split of CPLA produces 3,533 CPLA values that are *less than or equal to* the Median and 3,504 that were *greater than* the Median. Given the LQ partition {A:Neutral, B:Over-Valued & C;Under-Valued}, IF the Null is the State of Nature, then the expectation is that over this LQ partition, the percentages in the Low & High partitions would be not different than the marginal or in this case 50%. This means, of course, that there is NO likely LQ effect; meaning that we do NOT learn anything if a firm is labeled as **Over-Valued** or *Under-Valued* relative to the variable: CPLA. **Short story:** this LQ: {Neutral, **Over-Valued** & *Under-Valued*} has NO information value and is likely to ONLY serve as a “distraction” or “false information queuing” for an intelligent assessment of the market. **However, the contrary is the observation.** We see that the TD indications are all >> 1.0 indicating that one may reject the NO-Effect profile of the LQ on the DMV CPLA. Specifically, for the **Over-Valued** LQ indication for the CPLA measure, there are many fewer in the Low partition than expected relative to the marginals and so many more than expected in the High partition. This is an indication in the *Sensitivity* domain.

What follows is our analysis of this critical question: *Are these various LQ for the MPV likely to follow the Null?* If this were to be the case then respecting the LQ codex the CapitalCube Market Navigation Platform would likely, at best, to be of *no value*. These MPV will be presented in alphabetical order. In this profiling, we will use the sample values for each of the four DMV; we will give the percentage for the Low and the High median splits. This will be most useful in nuancing the TD indications as to whether for a particular split there is more or less activity than expected *over the aggregation of the panel*. Finally for each of the MPV as a summary of the inference information, appended to the Table, we will note the overall number and percentage of TD-indications, the number and percentage of TD-indications for the Variables of Interest [VoI]—{**Unfavorable** & *Favorable*}, and then number and percentage of TD-indications for the Neutral variables [N]. Also, to be clear, this is ONLY a test of the Null; if we reject the Null this does NOT indicate that the CCMNP has market informational utility, rather it only means that it addresses the

Sensitivity and *Specificity* issues. The market information utility testing will be the subject of the next study. Following, we will present 12 MPV and their unique LQ without market impact comment as the focus of this paper is the Null testing NOT the market information utility testing.

1. **MPV: Accounting Quality** [A=**Aggressive Accounting**, B=*Conservative Accounting*, C=Non-Cash Earnings, D=Possible Sandbagging]

	CPLA	A	B	C	D	*	SESAL	A	B	C	D	*	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	*
Low	550	378	173	482	1583	Low	604	344	173	428	1549	Low	509	338	230	541	1618	Low	558	376	211	488	1633	
	49	50	51	57	52%		54	46	51	51	51%		45	45	67	65	53%		50	50	62	58	53%	
	2	0.1	0.1	5			2	3	0.1	0.1			13	10	12	24			3	3	4	4		
High	571	378	169	358	1476	High	513	403	168	411	1495	High	612	418	112	286	1428	High	563	379	131	352	1425	
	51	50	49	43	48%		46	54	49	49	49%		55	55	33	35	47%		50	50	38	42	47%	
	2	0.1	0.1	6			2	4	0.1	0.1			14	11	15	27			3	2	5	4		
Total	1121	756	342	840	3059	Total	1117	747	341	839	3044	Total	1121	756	342	827	3046	Total	1121	755	342	840	3058	

Summary: TD%: [24/32]; TD%VoI: [14/16]; TD%N: [10/16]

Illustration of a Relational Analysis For each of the MPV to be examined, we will note the presumed or tacit order for the LQ referencing the results of Lusk & Halperin (2016) where they report the exhaustive directional indications: {Neutral, **Unfavorable** or *Favorable*} for the various LQ for each of the MPV. This will aid in understanding the *Sensitivity* and *Specificity* of the profiles. For *Accounting Quality*, Lusk & Halperin report the tacit order as follows: Positive Indication of Market Potential to a Negative Indication of Market Potential. Therefore for Accounting Quality: [B=*Conservative Accounting*[*Favorable*] > C=Non-Cash Earnings & D=Possible Sandbagging as Neutral > A=**Aggressive Accounting**[**Unfavorable**] or B>C≈D>A. For each of the MPV over the DMV, we report the overall Chi2 p-value which, in fact, forms the overall inference for the various DMV for that MPV; we will shade these p-values <0.1. For example, the CPLA for Accounting Quality the overall Chi2 p-value was 0.0004 and is reported as the header in the Total Column for CPLA as “*” which suggests that for the CPLA over the four LQ one may reject the proportional marginal as the characterization of the marginal split for the four LQ; this suggests that there are possible interesting profiling effects. Using the TD indications, we see that for PDCPL & CCPL the LQ: C[Non-Cash Earnings] has proportional distributions that vary in an important way from the marginals. For D[Possible Sandbagging] there were more values in the LOW median partition than expected for CPLA & PDCPL. All of the TD Null testing information is concisely summarized as: **Summary: TD%: [24/32]; TD%VoI: [14/16]; TD%N: [10/16]**. This indicates that Twenty-Four of Thirty-Two [24/32] are TD: Significant—equal to or greater than 1.0. Further, for the non-Neutral variables or what we are calling the Variables of Interest: VoL {A & B} 14 of a possible 16 are TD significant. Finally, for the Neutral variables: N: {C & D} there are 10 TD for the possible 16.

2. **MPV: Borrowing Capacity** [A = **Constrained**, B = Limited Flexibility, C = *Quick and Able*, D = Some Capacity]

	CPLA	A	B	C	D	*	SESAL	A	B	C	D	0.01	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	*
Low	381	28	1586	390	2385	Low	289	29	1769	358	2445	Low	407	36	1443	386	2272	Low	391	52	1665	340	2448	
	60	41	46	60	49%		46	43	51	55	51%		64	53	42	60	47%		62	76	48	52	51%	
	15.1	0.1	10.1	14.8			3	1	0.1	2			38.5	0.1	22.3	20.8			15.8	8.9	5.6	0.1		
High	251	40	1891	260	2442	High	342	39	1693	291	2365	High	225	32	2004	260	2521	High	240	16	1815	310	2381	
	40	59	54	40	51%		54	57	49	45	49%		36	47	58	40	53%		38	24	52	48	49%	
	14.8	0.1	9.9	14.4			3	1	0.1	2			34.7	0.1	20.1	18.7			16.3	9.2	5.7	0.1		
Total	632	68	3477	650	4827	Total	631	68	3462	649	4810	Total	632	68	3447	646	4793	Total	631	68	3480	650	4829	

Summary: TD%: [24/32]; TD%VoI: [14/16]; TD%N: [10/16]

Overview Analysis. Referencing Lusk & Halperin, the tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: C>D≈B>A. The TD Null testing information is: **Summary: TD%: [24/32]; TD%VoI: [14/16]; TD%N: [10/16]**. This indicates that there are Twenty-Four of Thirty-Two [24/32] that are TD:

Significant. Further, for the VoL {A & C} 14 of a possible 16 are TD significant. Finally, for the Neutral variables: N: {B & D} there are 10 TD for the possible 16.

3. **MPV Capital Investing Strategy** [A=Betting on Future, B=Maintenance Mode, C=Milking the Business, D=Supporting Growth]

CPLA	A	B	C	D	0.01	SESAL	A	B	C	D	*	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	*
Low	248	528	406	516	1698	Low	229	478	425	527	1659	Low	264	739	351	472	1826	Low	251	644	380	555	1830
	51	51	54	46	50%		47	46	56	47	49%		54	73	47	42	54%		51	62	50	49	54%
	0.1	0.1	2	4			0.1	1	9	1			0.1	69	8	31			0.1	15	2	5	
High	240	502	349	616	1707	High	254	553	328	592	1727	High	224	272	403	659	1558	High	237	387	375	576	1575
	49	49	46	54	50%		53	54	44	53	51%		46	27	53	58	46%		49	38	50	51	46%
	0.1	0.1	2	4			0.1	1	8	1			0.1	80	9	37			0.1	17	2	5	
Total	488	1030	755	1132	3405	Total	483	1031	753	1119	3386	Total	488	1011	754	1131	3384	Total	488	1031	755	1131	3405

Summary: TD%: [22/32]; TD%VoI: [16/16]; TD%N: [6/16]

Overview Analysis. Referencing Lusk & Halperin, the tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: D>A≈B>C. The TD Null testing information is: **Summary: TD%: [22/32]; TD%VoI: [16/16]; TD%N: [6/16]**. This indicates that there are Twenty-To of Thirty-Two [22/32] that are TD: Significant. Further, for the VoL: {D & C} all are TD significant. Finally, for the Neutral variables: N: {A & B} there are Six TD of the possible 16.

4. **MVP: Dividend Coverage** [A= Moderate, B=Strong, C=Weak]

CPLA	A	B	C	0.1	SESAL	A	B	C	*	PDCPL	A	B	C	*	CCPL	A	B	C	*
Low	693	1432	503	2628	Low	763	1431	530	2724	Low	581	1421	392	2394	Low	582	1366	414	2362
	49	49	53	50%		54	50	56	52%		42	49	42	46%		41	47	44	45%
	0.1	0.1	2			1	3	3			6	7	3			4	3	0.1	
High	717	1468	442	2627	High	645	1456	413	2514	High	816	1458	546	2820	High	831	1533	531	2895
	51	51	47	50%		46	50	44	48%		58	51	58	54%		59	53	56	55%
	0.1	0.1	2			1	4	3			5	6	3			4	3	0.1	
Total	1410	2900	945	5255	Total	1408	2887	943	5238	Total	1397	2879	938	5214	Total	1413	2899	945	5257

Summary: TD%: [18/24]; TD%VoI: [12/16]; TD%N: [6/8]

Overview Analysis. Referencing Lusk & Halperin, the tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: B>A>C. The TD Null testing information is: **Summary: TD%: [18/24]; TD%VoI: [12/16]; TD%N: [6/8]**. This indicates that there are Eighteen of Thirty-Two [18/32] that are TD: Significant. Further, for the VoL: {A & B} 12 of a possible 16 are TD significant. Finally, for the Neutral variable: N: {A} there are Six TD for the possible eight.

5. **MPV: Dividend Quality** [A=High, B=Low, C=Medium]

CPLA	A	B	C	0.3	SESAL	A	B	C	0.2	PDCPL	A	B	C	*	CCPL	A	B	C	*
Low	1405	589	634	2628	Low	1516	576	632	2724	Low	1317	565	512	2394	Low	1282	573	507	2362
	49	51	52	50%		53	50	52	52%		46	49	42	46%		45	49	41	45%
	1	0.1	1			1	1	0.1			0.1	3	4			0.1	5	4	
High	1463	570	594	2627	High	1343	578	593	2514	High	1529	582	709	2820	High	1588	587	720	2895
	51	49	48	50%		47	50	48	48%		54	51	58	54%		55	51	59	55%
	1	0.1	1			1	1	0.1			0.1	2	4			0.1	4	3	
Total	2868	1159	1228	5255	Total	2859	1154	1225	5238	Total	2846	1147	1221	5214	Total	2870	1160	1227	5257

Summary: TD%: [16/24]; TD%VoI: [10/16]; TD%N: [6/8]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: $A > C > B$. The TD Null testing information is: **Summary: TD%: [16/32]; TD%VoI: [10/16]; TD%N: [6/8]**. This indicates that there are Sixteen of Twenty-Four [16/24] that are TD: Significant. Further, for the VoL: {A & B} Ten of a possible 16 are TD significant. Finally, for the Neutral variable: N: {C} there are Six TD for the possible eight.

6. **MPV: Earnings Coverage** [A = Earnings Focus, B= Laggard, C= Leader, D= Revenue Focus]

CPLA	A	B	C	D	*	SESAL	A	B	C	D	*	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	0.06
Low	389	720	655	294	2058	Low	376	741	664	283	2064	Low	464	736	623	297	2120	Low	422	674	666	325	2087
	50	53	48	47	50%		49	55	49	46	50%		61	55	46	48	52%		54	50	49	52	51%
	0.1	3	1	1			0.1	5	0.1	3			12	2	9	2			2	0.1	1	0.1	
High	385	637	710	331	2063	High	398	614	688	336	2036	High	301	614	736	325	1976	High	353	683	699	301	2036
	50	47	52	53	50%		51	45	51	54	50%		39	45	54	52	48%		46	50	51	48	49%
	0.1	3	1	1			0.1	5	0.1	3			13	2	10	2			2	0.1	1	0.1	
Total	774	1357	1365	625	4121	Total	774	1355	1352	619	4100	Total	765	1350	1359	622	4096	Total	775	1357	1365	626	4123

Summary: TD%: [22/32]; TD%VoI: % [12/16]; TD%N: [10/16]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: $C > A \approx D > B$. The TD Null testing information is: **Summary: TD%: [22/32]; TD%VoI: [12/16]; TD%N: [10/16]**. This indicates that there are Twenty-Two of Thirty-Two [22/32] that are TD: Significant. Further, for the VoL: {B & C} 12 of a possible 16 are TD significant. Finally, for the Neutral variables: N: {A & D} there are 10 TD for the possible 16.

7. **MPV Growth Expectations** [A= Expected Decline, B= Strategic Play, C = Substandard, D= Superior]

CPLA	A	B	C	D	0.3	SESAL	A	B	C	D	0.9	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	*
Low	259	309	303	347	1218	Low	248	310	285	350	1193	Low	323	335	377	281	1316	Low	256	411	278	390	1335
	50	49	53	48	50%		49	49	50	48	49%		63	53	68	39	54%		49	65	48	54	54%
	0.1	0.1	1	0			0.1	0.1	0.1	0.1			8	0	18	33			2	13	4	0	
High	259	324	272	382	1237	High	263	322	290	373	1248	High	186	297	181	448	1112	High	263	222	298	339	1122
	50	51	47	52	50%		51	51	50	52	51%		37	47	32	61	46%		51	35	52	47	46%
	0.1	0.1	1	0			0.1	0.1	0.1	0.1			10	0	22	39			3	16	5	0	
Total	518	633	575	729	2455	Total	511	632	575	723	2441	Total	509	632	558	729	2428	Total	519	633	576	729	2457

Summary: TD%: [14/32]; TD%VoI: % [8/16]; TD%N: [6/16]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: $D > A \approx B > C$. The TD Null testing information is: **Summary: TD%: [14/32]; TD%VoI: [8/16]; TD%N: [6/16]**. This indicates that there are Fourteen of Thirty-Two [14/32] that are TD: Significant. For the VoL: {A & D} eight of a possible 16 are TD significant; for the Neutral variables: N: {A & B} there are six TD of the possible 16.

8. **MPV: M&A Action** [A= Acquirer, B= Target]

CPLA	A	B	*	SESAL	A	B	*	PDCPL	A	B	*	CCPL	A	B	*
Low	866	389	1255	Low	945	234	1179	Low	708	325	1033	Low	734	275	1009
	47	99	56%		51	60	53%		38	83	46%		40	70	45%
	28	133			1	4			25	117			12	56	
High	989	2	991	High	903	153	1056	High	1147	66	1213	High	1121	116	1237
	53	1	44%		49	40	47%		62	17	54%		60	30	55%
	36	169			1	5			21	100			10	46	
Total	1855	391	2246	Total	1848	387	2235	Total	1855	391	2246	Total	1855	391	2246

Summary: TD%: [16/16]; TD%VoI: % [16/16]; TD%N: [N/A]

Overview Analysis. The tacit LQ order is obvious: $A > B$. Clear is that A :Acquirers dominate in every relevant economic context relative to B :Target firms—this is often characterized metaphorically as the “Food Chain”. The TD Null testing information is: **Summary: TD%: [16/16]; TD%VoI: [16/16]; TD%N: [N/A]**. This indicates that all of the LQ are sensitive.

9. **MPV: Management of Reserves** [A=Modest Buildup, B=Modest Drain, C=Strong Buildup, D=Strong Drain]

CPLA	A	B	C	D	0.3	SESAL	A	B	C	D	.0.2	PDCPL	A	B	C	D	0.9	CCPL	A	B	C	D	0.5
Low	425	1	1151	198	1775	Low	472	2	1092	205	1771	Low	459	1	1153	209	1822	Low	456	2	1134	223	1815
	48	33	51	49			53	67	49	51			52	33	52	51			51	67	51	55	
	1	0.1	1	0.1			1	0.1	1	0			0.1	0.1	0.1	0.1			0.1	0.1	0.1	1	
High	463	2	1087	209	1761	High	415	1	1129	199	1744	High	427	2	1061	197	1687	High	432	1	1105	184	1722
	52	67	49	51			47	33	51	49			48	67	48	49			49	33	49	45	
	1	0.1	1	0.1			1	0.1	1	0			0.1	0.1	0.1	0.1			0.1	0.1	0.1	1	
Total	888	3	2238	407	3536	Total	887	3	2221	404	3515	Total	886	3	2214	406	3509	Total	888	3	2239	407	3537

Summary: TD%: [10/32]; TD%VoI: % [6/16]; TD%N: [4/16]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: $C > A \approx B > D$. The TD Null testing information is: **Summary: TD%: [10/32]; TD%VoI: [6/16]; TD%N: [4/16]**. This indicates that there are Ten of Thirty-Two [10/32] that are TD: Significant. For the VoL: {C & D} six of a possible 16 are TD significant. For the Neutral variables: N: {A & B} there are 4 TD for the possible 16.

10. **MPV: Share Price Performance** [A= Fading, B= Lagging, C= Leading, D= Rising]

CPLA	A	B	C	D	*	SESAL	A	B	C	D	*	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	0.04
Low	332	902	313	432	1979	Low	357	702	550	382	1991	Low	339	670	598	383	1990	Low	375	621	653	364	2013
	43	79	24	61	51%		46	62	43	54	51%		44	59	47	55	51%		48	55	51	52	52%
	9	184	178	16			3	25	17	1			8	14	6	2			2	2	0.1	0.1	
High	442	236	979	272	1929	High	412	432	731	319	1894	High	428	461	688	315	1892	High	400	512	639	340	1891
	57	21	76	39	49%		54	38	57	46	49%		56	41	54	45	49%		52	45	49	48	48%
	9	189	183	16			4	26	18	2			8	15	6	2			2	3	0.1	0.1	
Total	774	1138	1292	704	3908	Total	769	1134	1281	701	3885	Total	767	1131	1286	698	3882	Total	775	1133	1292	704	3904

Summary: TD%: [28/32]; TD%VoI: % [22/24]; TD%N: [6/8]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: $C > D > A \approx B$. The TD Null testing information is: **Summary: TD%: [28/32]; TD%VoI: [22/24]; TD%N: [6/8]**. This indicates that Twenty-Eight of Thirty-Two [28/32] are TD: Significant. For the VoL: {A,B & C} 22 of a possible 24 are TD significant. Finally, for the Neutral variable: N: {D} there are 6 TD of the possible 8.

11. **MPV: Sustainability of Returns** [A= Eroding, B=Improving, C=Questionable, D= Sustainable]

CPLA	A	B	C	D	*	SESAL	A	B	C	D	*	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	*
Low	80	84	644	886	1694	Low	69	65	587	940	1661	Low	76	112	846	752	1786	Low	89	102	750	897	1838
	58	42	55	48	51%		51	33	50	52	50%		55	56	72	41	54%		65	51	64	49	55%
	2	3	4	2			0.1	12	0.1	1			0.1	0.1	76	53			3	0.1	17	12	
High	57	115	530	949	1651	High	66	134	586	880	1666	High	61	87	326	1078	1552	High	48	97	424	937	1506
	42	58	45	52	49%		49	67	50	48	50%		45	44	28	59	46%		35	49	36	51	45%
	2	3	4	2			0.1	12	0.1	1			0.1	0.1	88	61			3	0.1	21	15	
Total	137	199	1174	1835	3345	Total	135	199	1173	1820	3327	Total	137	199	1172	1830	3338	Total	137	199	1174	1834	3344

Summary: TD%: [22/32]; TD%VoI: % [12/16]; TD%N: [10/16]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: $D > B \approx C > A$. The TD Null testing information is: **Summary: TD%: [22/32]; TD%VoI: % [12/16]; TD%N: [10/16]**. This indicates that there are Twenty-Two of Thirty-Two [22/32] that are TD: Significant. Further, for the

VoL: {A & D} 12 of a possible 16 are TD significant. Finally, for the Neutral variables: N: {B & C} there are 10 TD of the possible 16.

12. **MPV: Valuation Characteristics** [A= **Challenged**, B= **Harvesting**, C= **Outperforming**, D= **Turnaround**]

CPLA	A	B	C	D	0.03	SESAL	A	B	C	D	0.13	PDCPL	A	B	C	D	*	CCPL	A	B	C	D	*
Low	110	417	227	374	1128	Low	102	395	235	347	1079	Low	120	539	159	383	1201	Low	69	454	265	458	1246
	59	48	48	51	50%		55	46	49	48	48%		71	63	33	53	54%		37	52	56	63	55%
	3	0.1	0.1	0.1			2	1	0.1	0.1			9	13	37	0.1			11	1	0.1	8	
High	75	449	250	355	1129	High	82	461	240	379	1162	High	50	322	318	345	1035	High	116	411	212	271	1010
	41	52	52	49	50%		45	54	51	52	52%		29	37	67	47	46%		63	48	44	37	45%
	3	0.1	0.1	0.1			2	1	0.1	0.1			11	15	43	0.1			13	1	0.1	9	
Total	185	866	477	729	2257	Total	184	856	475	726	2241	Total	170	861	477	728	2236	Total	185	865	477	729	2256

Summary: TD%: [18/32]; TD%VoI: % [10/16]; TD%N: [8/16]

Overview Analysis. The tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: C>B≈D>A. The TD Null testing information is: Summary: TD%: [18/32]; TD%VoI: % [10/16]; TD%N: [8/16]. This indicates that Eighteen of Thirty-Two [18/32] are TD: Significant. For the VoL: {A & C} ten of a possible 16 are TD significant. Finally, for the Neutral variables: N: {B & D} there are eight TD of the possible 16.

13. **MPV: Relative Evaluation** P/B Valuation Characteristics [A= **P/B Above Peers**, B= **P/B Below Peers**]

CPLA	A	B	*	SESAL	A	B	0.9	PDCPL	A	B	*	CCPL	A	B	0.7
Low	1234	835	2069	Low	1310	766	2076	Low	958	1055	2013	Low	1311	778	2089
	47	54	50%		50	50	50%		36	70	49%		50	50	50%
	4	7			0.1	0.1			82	144			0.1	0.1	
High	1394	704	2098	High	1306	765	2071	High	1670	443	2113	High	1317	764	2081
	53	46	50%		50	50	50%		64	30	51%		50	50	50%
	4	6			0.1	0.1			78	137			0.1	0.1	
Total	2628	1539	4167	Total	2616	1531	4147	Total	2628	1498	4126	Total	2628	1542	4170

Summary: TD%: [8/16]; TD%VoI: % [8/16]; TD%N: [N/A]

Overview Analysis. The obvious tacit LQ order from positive to negative indications of market performance of this Linguistic codex is: A>B. The TD Null testing information is: Summary: TD%: [8/16]; TD%VoI: % [8/16]; TD%N: [N/A]. This indicates that for the CPLA & PDCPL there are TD effects. However, for the SESAL & the CCPL such sensitivity is not in evidence.

5. **Inferential Analysis: The Results**

5.1 *Panel Sensitivity Test H1*

In this case, we used the differential frequency counts as measured by the TD significance tests as a measure of interaction of the firms over of the Panel relative to the LQ for each MPV aggregated over the Panel. Assuming the Null is the State of Nature, there should be very few TD Chi2-significant LQ-Median Partition effects. Due to possible, but unlikely biasing inter-MPV interaction effects over the LQ-set, the Null was set at five times the reasonable FPE non-directional test of 5% or in our case: 25% for our testing purposes. Therefore, we will use the number of TD differentials observed overall as the computation of the sensitivity association percentage and then test the 95% directional CI for the inclusion of 25%. If the very conservative expectation, 25%, is in the observed 95%CI then this will not provide the rationale for rejecting the Null of no association. *Results:* There were 242 effects which were TD indications of the 368 possible in the Panel or 65.8%. The lower limit for the directional test of the 95%CI was 61.7%. The proffered expectation was 25% and is clearly outside the observed interval rationalizing rejecting the Null in favor of the support for sensitivity for the CapitalCube MPV for the LQ of the Panel tested. The p-value of this result is <0.0001.

5.2 *Panel Specificity: H2 iff: H1* is founded—i.e., the Null of *H1* is rejected—then the operational conditional Null for *H2* offered is that the number TD effects over the {**Unfavorable** or *Favorable*} indications would be no different in percentage terms compared to such TD effects for the Neutral indications. Clearly, the specificity directional test for our study related to this Null is that there will be more TD indications for the avowed non-Neutral indications.

Results: There were 160 indications that were classified as {Unfavorable or *Favorable*} of the 216 possible or 74.1%. For the Neutral LQ, there were 82 with TD indications classified as Neutral for the 152 Possible or 53.9%. The z-calculated for the difference was 4.00 which had a directional p-value of < 0.00003 thereby rationalizing rejecting the Null of H_2 with the proffered expectation.

5.3 Conclusion The summary of this testing is:

There is no evidence that the imbedded Variable Set: MPV[LQ[DMV]] used by CapitalCube overall or for any of its partitions as expressed in the S&P500 dataset has at its generating function(s) random process(es). Therefore, the CapitalCube Market Navigation Platform for the imbedded variable set tested seems Sensitive & Specific and so could provide market relevant decision-making information.

The testing of the directional market effects for this *Sensitive & Specific* imbedded variable set is the next research investigation to be undertaken.

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, *Deloitte Touche LLP*, Audit & Assurance Services, Frankfurt [Main], Germany and the anonymous reviewers of the *Journal of Accounting and Finance Research* for the for their helpful suggestions.

References

- Allen, F. & Carletti, E. (2010). An overview of the crisis: Causes, consequences, and solutions. *International Review of Finance*, 10, 1-26. <https://doi.org/10.1111/j.1468-2443.2009.01103.x>
- Cho, W.K.T. & Gaines, B.J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *American Statistician*, 61, 218-223. <https://doi.org/10.1198/000313007X223496>
- Jorion, P. & Zhang, G. (2009). Credit Contagion from Counterparty Risk. *Journal of Finance*, 64, 2053-2087. <https://doi.org/10.1111/j.1540-6261.2009.01494.x>
- Lusk, E & Halperin, M. (2014). Detecting Newcomb-Benford Digital Frequency Anomalies in the Audit Context: Suggested Chi2 Test Possibilities. *Journal of Accounting and Finance Research*, 3, 191-205. <https://doi.org/10.5430/afr.v3n2p191>
- Lusk, E. & Halperin, M. (2015). An Empirical Contextual Validation of the CapitalCube™ 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>
- Lusk, E. & Halperin, M. (2016). A Linguistic Examination of the CapitalCube™ Market Effect Variables. *Accounting and Finance Research*, 6. <https://doi.org/10.5430/afr.v5n2p89>
- Mishkin, F. (2011) Over the Cliff: From the Subprime to the Global Financial Crisis. *Journal of Economic Perspectives*, 25, 49-70. <https://doi.org/10.1257/jep.25.1.49>
- Pachamanova, D & Fabozzi, F. (2014). Recent trends in equity portfolio construction analytics. *Journal of Portfolio Management*, Spring, 137-151. <https://doi.org/10.3905/jpm.2014.40.3.137>
- Petter, K., Tütüncü R. & Fabozzi, F. (2014). 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234, 356–371. <https://doi.org/10.1016/j.ejor.2013.10.060>
- Sevastjanov, P & Dymova, L. (2009). Stock screening with use of multiple criteria decision making and optimization. *Omega*, 37, 659 – 671. <https://doi.org/10.1016/j.omega.2008.04.002>
- Tamhane, A & Dunlop, D. (2000). *Statistics and Data Analysis*. Prentice-Hall: Upper-Saddle River NJ. USA ISBN: 0-13-744426-5

Notes

Note 1. However, we did randomly fill the Panel categories and when we randomly tested the percentage of TDs we found that there were 7 in the 100 exposures. Assuming that this is the case, then 25% as the Null test seems conservative.

Note 2. We have used the research report of Lusk & Halperin (2014) where they suggest a Chi2 calibration range of [315 to 440] on an instance basis for Benford profiling for Chi2 profile inference testing. Therefore we will use the most extreme case of: 220 [440/2] instances for each triage label as the initial basis.