Intercorrelations of Measures of Forecasting Accuracy and a Recommendation

John R. Dickinson
University of Windsor
MExperiences@bell.net

ABSTRACT

Forecasting plays several roles in business games. Regarding sales forecasting, and more specifically the accuracy of sales forecasting, several definitions have been conceptualized and applied. Despite their conceptual differences, if the measures yield high intercorrelations then, empirically at least, the choice among them may be a matter of indifference. The present research summarizes numerous forecasting accuracy measures, estimates their intercorrelations, and concludes with some considerations for working toward a consensus and a recommendation.

INTRODUCTION

Forecasting is prominent in business games. Of specific interest here is sales forecasting, although forecasts of several other criteria have been employed. (More exhaustive lists of criteria that have been forecasted may be found in deSouza, Bernard, & Cannon (2010, 20 criteria), Hand & Sims (1975, 6 criteria), and Teach (2007, 7 criteria).)

Forecasting of unit sales has been used by Anderson & Lawton (1992, 1990), de Souza, Bernard, & Cannon (2010), Gosen & Washbush (2001), Hand & Sims (1975), Palia (2011, 2004), Peach & Platt (2000), Teach (2007, 1990, 1989, 1987), Washbush (2003), and Wolfe (1993). Forecasting of market share has been used by Anderson & Lawton (1988), de Souza, Bernard, & Cannon (2010), Gosen & Washbush (2002), Teach (2007, 1990, 1989, 1987), and Wolf (1993).

Often, the focus of forecasting-related research is not the act of forecasting, but the accuracy of forecasts. Across the many works just cited and others, operationally defining accuracy has taken several forms. None of those several forms has drawn a consensus from educators and researchers. This diversity of measures of forecasting accuracy compromises the accumulation of knowledge across the many business gaming studies using different forms. Their being different by definition it cannot be expected that the various measures yield identical results. However, little research in business gaming has estimated the degree of differences in values the various measures yield. More specifically, if those results are not very highly intercorrelated then the issue of which is the measure of choice–and thus coming to gain some consensus–becomes prominent. The present study compares the several forms, also adding the dimension of demand versus sales, and concludes with conceptual and definitional arguments toward arriving at a consensus measure.

DEMAND VERSUS SALES

Dickinson (2013, p. 102) offers the following explanation:

> ...it is useful to clarify that "demand" refers to product units that consumers seek to purchase. This is the demand attributable to the desirability to consumers of the company's offering (along with other factors, e.g., the game environment structure and competitors' strategies). In contrast, "sales" refers to units actually sold. The difference between demand and sales is stockouts. Stockouts reflect the opportunity loss of sales that could have could have been made, but were not due to lack of availability. It is demand that is influenced by company strategy. Sales equals demand where sufficient units are available.

The incidence of stockouts may not be inconsequential. Dickinson (2006b, pp. 234-235) reported that, "Across the (four) competitions between 20.9 and 27.3 percent of approximately 6,000 inventories stocked out."
Several studies do not make clear whether the actual sales variable used takes into account stockouts (Anderson and Lawton 1990, de Souza, Bernard, & Cannon 2010, Gosen & Washbush 2001, Hand & Sims 1975, Teach 1989, Wolfe 1993).

While it may be nominally termed a "sales" forecast, it is demand that is forecasted. Simple as the schema is, it is critically important where sales forecasting accuracy is involved. Suppose, for example, that a manager forecasts unit sales of 100. A stockout occurs, however, and actual unit sales are 80. The manager's forecast will *appear* to be in error by 20 units. If demand is, say, 95, though, the manager's forecast is only in error by 5 units. The bulk of the "error" derives from there not being available sufficient inventory, not in the manager's forecasting ability. This in itself calls into question the validity of using unit sales, instead of unit demand, in measuring forecasting accuracy.

The foregoing is a general argument for the use of demand, rather than sales, in measuring forecast accuracy. The example assumes an explicit forecast made by the manager. The present study, though, uses an implicit forecast. The number of units the manager makes available for sale–beginning inventory plus units ordered for resale–is taken to be his or her forecast. (There is nothing in the configuration of the specific game, e.g., supply disruption, inflation in supplied unit price, etc., that would support this *not* being the case.) Where a stockout occurs, the forecast error is the amount of the stockout. Where a stockout does not occur, the forecast error is the amount of ending inventory.

Comparisons made in this study include measures based both on unit demand and on unit sales.

OPERATIONALIZATIONS OF FORECASTING ACCURACY (ERROR)

Sales forecasting error has been operationalized in several ways.

[i] The absolute deviation of forecasted unit sales from actual unit sales divided by actual sales was used by de Souza, Bernard, & Cannon (2010, p. 72). That formula was adapted from Teach (1989, p. 104) who applied the formula to both unit sales and market share.

[ii] "The percent deviation of actual sales from forecasted sales." was used by Hand and Sims (1975, p. 711). In contrast to the immediately above, the divisor was forecasted sales. "Each of the six forecast errors [of which unit sales was one] is computed by dividing the measure by the team's forecast." (p. 710)

[iii] Peach & Platt (2000, p. 245) "...divided actual demand by projected demand...Numbers closer to one hundred percent were interpreted as indicating more learning."

[iv] Gosen & Washbush (2001, p. 93) used an absolute, rather than relative, definition comprising the difference "...between predicted sales in units and actual sales in units..."

[v] Washbush (2003, p. 251) first calculated a "total demand" for each company in a given period "...by summing actual sales and lost sales for each area. Forecast error for each round of play was calculated by subtracting forecast sales from actual [sic, presumably total] demand and converting to the absolute value." Note that Washbush (2003) accounts for stockouts, i.e., lost sales.

An additional study uses a relative or percentage measure, but does not make clear what is the divisor. Wolfe (1993) does not directly define "company demand" forecasting error, but a table footnote presenting forecasting errors states, "All numbers are percentages..." (p. 57)

In sum, the various operationalizations differ with respect to:

- Is the measure in units or percent, i.e., relative, form? [iv] and [v] are in units; [i], [ii], and [iii] are in percents.
- If the measure is a relative one, relative to what? I.e., what is the divisor? For [i] the divisor is unit sales and for [ii] and [iii] the divisor is the unit forecast value.
- What is the value of the measure when there is no error? Zero for [i], [ii], [iv], and [v]; 100 percent for [iii].
- Are stockouts taken into account? [i]-[iv] no; [v] yes.

There appears to be no conventional labeling of the basic sales forecasting accuracy components. Here, then, let D indicate unit demand (i.e., before stockouts), S indicate unit sales, and F indicate unit forecast. The respective calculations of forecasting accuracy (most often actually forecasting error) for the above operationalizations, then, are:

[i] $|S-F|/S$ [iv] $|S-F|$
[ii] $|S-F|/F$ [v] $|D-F|$
[iii] $|1-(S/F)|$

It may be seen that [ii] and [iii] are perfectly linearly correlated.

Per Dickinson (2013) the occurrence of a stockout renders unit sales an inaccurate indicator of the consequences of the company's strategy which, in turn, renders inaccurate the sales~forecast relationship as a measure of forecast accuracy. In this condition it is demand that is the more accurate indicator than sales. Accordingly, the extant operationalizations of forecasting accuracy are here augmented with the incorporation of unit demand. The mix of measures investigated, then, is:

[1] $|S-F|/S$ [5] $|S-F|$
[2] $|D-F|/D$ [6] $|D-F|$
[3] $|S-F|/F$
[4] $|D-F|/F$

Since the various operationalizations comprise different calculations, obviously different values for forecasting accuracy (or error) will result. This study investigates the extent of those differences; specifically the degree of intercorrelation among the measures.

IMPLICIT MEASURE OF FORECASTING ACCURACY

Unlike virtually all other studies of forecasting accuracy, the present research does not rely on self-reported forecasts made by game participants. Rather, an implicit forecast is made as follows (Dickinson 2013): Consider games where managers order product units for resale or produce units for sale. The implied unit sales forecast is the number of product units ordered or produced plus any inventory available at the beginning of the competition period. When a manager does not order/produce units, presumably the manager anticipates, i.e., forecasts, beginning inventory to be sufficient. "Sufficient" is an inexact forecast, of course, and in such cases no implicit forecast value is available. Dickinson (2013), though, reports that in one competition where 805 product orders might have

TABLE 1
NUMBERS OF FORECASTS AND STOCKOUTS

Region-Product Segment	Number of Forecasts	Number of Stockouts	Stockouts as % of Forecasts
1	381	80	21.00
2	321	32	9.97
3	401	127	31.67
4	379	72	19.00

been placed, in only 6.5 percent of them was an order not placed. Such instances are omitted from the data analyzed below. In the present study, where 1608 product orders might have been placed, in only 5.3 percent of them was an order not placed.

Above, the critical distinction between demand and sales is explained. Where a stockout occurs, unit sales are not a valid indicator of the consequences of a company's (marketing) strategy. Essentially, in such cases, the consequences are understated by sales and the calculation of forecasting accuracy (error) using unit sales is erroneous.

In the present study, both sales and demand are present among the measures of forecasting accuracy. Where a stockout does not occur, unit demand is equal to unit sales. Where a stockout does occur, demand is equal to unit sales plus the unit amount of the stockout.

DATA

(Implicit) Forecast, sales, and demand data were obtained using *The Marketing Management Experience* (*MME*, Dickinson 2006a). In the *MME*, companies may operate in either or both of two geographic regions and may market either or both of two products (a digital still camera and a digital video camera), giving rise to four region-product segments. For each of the four segments, an order for products to be resold may be placed with an inventory being maintained in each. For each competition period, then, there are potentially four implicit forecasts for each company. (*MME* managers need not necessarily operate in all four segments in any given period, though most choose to do so.)

Data were potentially available for 51 companies in each of the four region-product segments. Companies competed within 12 industries for nine periods, following a single trial period.

TABLE 2
MEAN INTERCORRELATIONS AMONG MEASURES
MEASURES OF FORECASTING ACCURACY

	Measure	[1] |S-F|/S	[2] |D-F|/D	[3] |S-F|/F	[4] |D-F|/F	[5] |S-F|
[2]	|D-F|/D	.894 .975 .756 .945				
[3]	|S-F|/F	.969 .971 .987 .972	.839 .939 .715 .901			
[4]	|D-F|/F	.596 .916 .417 .781	.813 .962 .850 .906	.565 .926 .384 .769		
[5]	|S-F|	.942 .921 .956 .917	.828 .894 .698 .847	.958 .920 .963 .930	.564 .858 .375 .715	
[6]	|D-F|	.734 .891 .592 .787	.892 .915 .912 .880	.720 .879 .561 .777	.870 .908 .881 .894	.785 .967 .618 .854

Entries are mean correlations for:

	Segment 1	n=45 companies
	Segment 2	n=43 companies
	Segment 3	n=46 companies
	Segment 4	n=45 companies

As noted above, an implicit forecast cannot be calculated when a product order is not placed in a given competition period. Too, again, managers may choose to not operate in some segments. The actual number of data points, then, was reduced accordingly.

Table1 presents the number of implicit sales forecasts for each *MME* region-product segment. Also in Table 1 are the number of instances where an inventory stockout occurred. As explained above, stockouts are a critical consideration in operationalizing the consequences of a company's strategy; where a stockout occurs, sales is not an accurate measure of the effect of strategy.

ANALYSIS

For each company and each region-product segment and each competition period an implicit sales forecast was calculated as described earlier. Corresponding unit sales (affected by stockouts) and unit demand (unaffected by stockouts) were known. For a given company and segment, then, for each period a measure of forecasting accuracy was calculated. This was done for each of the six measures defined above. For any two measures of forecasting accuracy, a correlation was calculated over the nine competition periods. This was done "within company." (To ensure stability, a correlation was only calculated when data were available for at least six of the nine periods.)

In sum, for each company correlations for all pairs of the forecasting accuracy measures were calculated. Means of the correlations were calculated across the companies; 43-46 companies depending on the specific region-product segment. Those mean intercorrelations are presented in Table 2.

RESULTS

Table 2 presents (mean) intercorrelations among the six operationalizations of sales forecasting accuracy.

The distribution of values in Table 2 is wide; that distribution is summarized in Table 3. While the modal range is greater than .90, 60 percent of the correlations are less than 0.90 and 13 percent are less than 0.60.

Interpretation of the many correlation coefficients may be facilitated by making comparisons within meaningful subsets of them.

Relative versus Absolute

Two of the measures of forecasting accuracy are absolute: |S-F| and |D-F|. The former may be made relative by dividing by S or by dividing by F. This is not merely dividing by a constant; S and F take on different values as a competition progresses over periods. All eight correlations between the absolute and relative versions (two divisors x four segments) are greater than 0.91. Similarly, for |D-F| being divided by D or F, all eight correlations are greater than 0.87. Between themselves, there is not a great difference whether the measures are absolute or relative. Despite these high correlations, though, the absolute and relative measures are not interchangeable.

|S-F| correlated with other relative measures–|D-F|/D and |D-F|/F–finds four of eight correlations less than 0.80. |D-F| correlated with |S-F|/S and |S-F|/F yields six of eight correlations less than 0.80. That is, there are marked differences.

Relative to Demand~Sales or Relative to Forecast

Relative measures of forecasting accuracy are relative to either the actual result–S or D–or to the forecasted value, F. Within each of a pair of measures, this comparison can be made:

[1] |S-F|/S	[2] |D-F|/D
[3] |S-F|/F	[4] |D-F|/F

The four (i.e., four segments) |S-F|/S~|S-F|/F correlations are all greater than 0.96. All of the four |D-F|/D~|D-F|/F correlations are 0.962 or less, with two of them being 0.85 or less. It is the latter set of correlations that is the more informative since, as has been explained, unit sales where a stockout has occurred is an inaccurate indicator of the effect of (marketing) strategy. The difference between using D or F as a divisor is not inconsequential.

Demand Versus Sales

Three pairs of forecasting accuracy measures differ only in that one member of the pair is based on sales and the other member is based on demand:

[1 |S-F|/S	[3] |S-F|/F	[5] |S-F|
[2] |D-F|/D	[4] |D-F|/F	[6] |D-F|

Inspecting the relevant cells of Table 2 reveals a consistent pattern of correlations. In the cells, the Segment 2 correlation is the highest, Segment 4 is second highest, Segment 1 is third highest, and Segment 3 is fourth highest. This is no coincidence. The corresponding percents of stocked out inventories are 9.97%, 19.00%. 21.00%, and 31.67%. That is,

TABLE 3
DISTRIBUTION OF MEAN CORRELATIONS

Range	Count	Percent
$.90 \le r < 1.0$	24	40.00
$.80 \le r < .90$	16	26.67
$.70 \le r < .80$	10	16.67
$.60 \le r < .70$	2	3.33
$.50 \le r < .60$	5	8.33
$.40 \le r < .50$	1	1.67
$.30 \le r < .40$	2	3.33
Total	60	100.00

understandably, correlations between measures are highest when the difference between unit sales and unit demand is least and the correlation is lowest when the difference is greatest. Cautioning is that across the 12 correlations in those three cells of Table 2, half are less than 0.80. Ignoring stockouts when calculating forecast accuracy can constitute a material inaccuracy.

DISCUSSION

In one sense, all of the six measures investigated here are "correct," just as the mean and median are both "correct" measures of location. Their definitions are straightforward and the researcher might, based on their definitions, chose one that he or she deems best suits his or her research purpose. There are additional considerations, though, that narrow the field.

Demand versus Sales

Conceptually, it seems clear that the effect of (marketing) strategy should be in terms of unit demand, rather than unit sales. (As noted at the beginning of this report, there may, of course, be additional effect criteria beyond sales.) This study has demonstrated that measures of forecasting accuracy based on demand versus sales may differ markedly, i.e., they are considerably less than perfectly correlated. This phenomenon is due to stockouts (and not due to the forecast measure being an implicit one). In some games, supply may be unlimited making stockouts moot.

There is also the practicality that while unit sales is an integral element of competition results/feedback, unit demand may not always be provided. That information is surely generated by a game's software; it just may not be made available to the game administrator or participants. This limitation can be easily remedied by the game designer.

It may be (probably correctly) maintained that where students are required to submit self-reported forecasts those forecasts may be more thoughtful or otherwise more well-founded than the implicit forecast values used in this study. Such explicit forecasts, then, may be more accurate than implicit forecasts. However, there is no particular basis for any such greater accuracy affecting the (in)comparability of the various operationalizations of forecasting accuracy.

Relative versus Absolute

The two absolute measures–|S-F| and |D-F|–may be valid for any given application. However, those results are not comparable across forecasts where levels of demand vary. For example, in the *MME* the market potential for Region 2 (12,000 and 6,000 units for the two products, respectively) is approximately twice the size of the potential for Region 1 (6,000 and 3,000). Whether it be |S-F| or |D-F|, an absolute error of, say, 500 units is clearly not comparable between the two regions. International business games may have much larger demand levels than games of more local scope. To facilitate the accumulation of knowledge regarding forecasting accuracy across different researches, use of absolute measures should be discouraged.

Absolute errors might be of greater variation at higher levels of demand than at lower levels. In keeping with this, interestingly the |S-F|~|D-F| correlations for Region 2 (i.e., Segments 3 and 4, .618 and .854) are considerably lower than their respective counterparts for Region 1 (Segments 1 and 2, .785 and .967).

|D-F|/D

In light of the above two considerations, there remain these measures: |D-F|/D and |D-F|/F. Correlations between the two for the four region-product segments are .813, .962, .850, and .906, respectively (Table 2), indicating that the two are not interchangeable. Two rationales favor the former. First, it is demand that is the object of the forecast, not *vice versa*; demand is the criterion. Second, demand is founded in the (simulated) market, usually common to all competitors and seemingly a more fundamental basis than the more idiosyncratic forecasts of individual managers.

RECOMMENDATION

To repeat, all of the six measures of forecasting accuracy investigated here are "correct." Though the argument for using demand rather than sales is compelling, there may be counter considerations. Too, in light of varying philosophies and circumstances researchers may persist in using a variety of measures. Granting this, it would facilitate adding to the store of knowledge if researchers, if not adopting, at least reported one common measure. With the rationales and analyses presented here, it is recommended that |D-F|/D be considered for that common measure.

REFERENCES

Anderson, Philip H. & Lawton, Leigh (1992). The relationship between financial performance and other measures of learning on a simulation exercise. *Simulation & Gaming*, Vol. 23 (September), 326-340.

Anderson, Philip H. & Lawton, Leigh (1990). The relationship between financial performance and other measures of learning on a simulation exercise. In Wingender, John R. & Wheatley, Walt J. (Eds.), *Developments in business simulation & experiential exercises*, Vol. 17, 6-10. (Reprinted from *The Bernie Keys Library*.)

Anderson, Philip H. & Lawton, Leigh (1988). Assessing student performance on a business simulation exercise. In Sanders, Patricia & Pray, Thomas (Eds.), *Developments in business simulation and experiential exercises*, Vol. 15, 241 -245. (Reprinted from *The Bernie Keys Library*.).

de Souza, Moises Pacheco, Bernard, Ricardo R. S., & Cannon, Hugh M. (2010). Another look at the use of forecasting accuracy on the assessment of management performance in business simulation games. In Smith, J. Alexander (Ed.), *Developments in business simulation and experiential learning*, Vol. 37, 68-78. (Reprinted from *The Bernie Keys Library*.)

Dickinson, John R. (2013). An implicit measure of forecasting accuracy. In Boscia, Marian (Ed.), *Developments in business simulation and experiential learning*, Vol. 40, 102 -106. (Reprinted from *The Bernie Keys Library*.)

Dickinson, John R. (2006a). *The marketing management experience*. Windsor, Ontario: Management Experiences. ISBN: 0-9691231-1-6

Dickinson, John R. (2006b). The invalidity of profit=f(market share) PIMS "validation" of marketing games. In Smith, J. Alexander (Ed.), *Developments in business simulation and experiential learning*, Vol. 33, 232-237. (Reprinted from *The Bernie Keys Library*.)

Gosen, Jerry & Washbush, John (2002), The validity investigation of a test assessing total enterprise simulation learning. In Vaughan, Mary Jo & Pillutla, Sharma (Eds.), *Developments in business simulation and experiential learning*, Vol. 29, 247-251. (Reprinted from *The Bernie Keys Library*.)

Gosen, Jerry & Washbush, John (2001). An initial validity investigation of a test assessing total enterprise simulation learning. In Pittenger, Khushwant S. & Vaughan, Mary Jo (Eds.), *Developments in business simulation and experiential learning*, Vol. 28, 92-95. (Reprinted from *The Bernie Keys Library*.)

Hand, Herbert H. & Sims, Henry P., Jr. (1975). Statistical evaluation of complex gaming performance. *Management Science*, Vol. 21, No. 6 (February), 708-717.

Palia, Aspy (2011). Tracking forecast error type, frequency and magnitude with the forecast error package. In Murph, Elizabeth (Ed.), *Developments in business simulation and experiential learning*, Vol. 31, 45-58. (Reprinted from *The Bernie Keys Library*.)

Palia, Aspy (2004). Online sales forecasting with the multiple regression analysis data matrices package. In Feinstein, Andrew Hale & Potosky, Denise (Eds.), *Developments in business simulation and experiential learning*, Vol. 31, 180 -182. (Reprinted from *The Bernie Keys Library*.)

Peach, E. Brian & Platt, Richard G. (2000). Total enterprise simulations and optimizing the decision set: assessing student learning across decision periods. In Page, Diana & Snyder, LT (Eds), *Developments in business simulation and experiential learning*, Vol. 27, 242-247. (Reprinted from *The Bernie Keys Library*.)

Teach, Richard D. (2007). Forecasting accuracy and learning: a key to measuring simulation performance. In Smith, J. Alexander (Ed.), *Developments in business simulation and experiential learning*, Vol. 34, 57-66. (Reprinted from *The Bernie Keys Library*.)

Teach, Richard D. (1990). Profits: the false prophet in business gaming. *Simulation & Gaming*, Vol. 21 (March), 12-26.

Teach, Richard D. (1989). Using forecasting accuracy as a measure of success in business simulations. In Pray, Tom & Wingender, John (Eds.), *Developments in business simulation & experiential exercises*, Vol. 16, 103-107. (Reprinted from *The Bernie Keys Library*.)

Teach, Richard D. (1987). Profits: the false prophet. In Kelly, Lane & Sanders, Patricia (Eds.), *Developments in business simulation and experiential exercises*, Vol. 14, 205-207. (Reprinted from *The Bernie Keys Library*.)

Washbush, John B. (2003). Simulation performance and forecast accuracy–is that all? In Pillutla, Sharma & Feinstein, Andrew Hale (Eds.), *Developments in business simulation and experiential learning*, Vol. 30, 250-253. (Reprinted from *The Bernie Keys Library*.)

Wolfe, Joseph (1993). "On the propriety of forecasting accuracy as a measure of team management ability: a preliminary investigation. *Simulation & Gaming*, Vol. 24 (March), 47-62.