SIMULATIONS OF THE RELATIONSHIP BETWEEN AN INFORMATION SYSTEM'S INPUT ACCURACY AND ITS OUTPUT ACCURACY

(Research- in-progress)

Irit Askira Gelman

University of Arizona, Tucson, AZ 85721-0108

askira@eller.arizona.edu

Abstract: The economic consequences of data quality are partly determined by the relationship between the quality of the input data and the quality of the information that an information system outputs. However, the relationship between an information system's data accuracy and its output information accuracy is hard to assess. A popular belief on this issue is reflected by the saying "garbage in garbage out", namely, the accuracy of the output of an information system is positively and tightly linked to the accuracy of its input. Yet, this belief has not been validated.

The relationship between an information system's input accuracy and its output accuracy is the subject of this paper. A traditional assumption in research of the relationship between input accuracy and output accuracy is that errors are independent. Motivated by a belief that dependence between errors is, in fact, common, this research examines the potential effect of the dependence factor on the relationship between an information system's input accuracy and its output accuracy. The main research questions are: (1) How does dependence between errors affect the relationship between input accuracy and output accuracy? (2) Is the relationship between input accuracy and output accuracy positive? An earlier paper [1] analyzed these questions applying the information structure model, rooted in statistical decision theory. The inquiry in [1] was limited, however, to an information system that uses a single input for calculating the output. This paper extends the scope of the investigation to situations in which an information system produces output using multiple rather than a single input. The inquiry applies simulations for that purpose.

Keywords: Data and information accuracy, dependence between errors, cost benefit analysis

INTRODUCTION

The economic consequences of data quality are partly determined by the relationship between the quality of the input data and the quality of the information that an information system outputs. This is because data often undergo various processing before any actual use, such that quality may change. However, the

relationship between an information system's data accuracy and its output information accuracy is hard to assess. The popular belief is reflected by the saying "garbage in garbage out", namely, the accuracy of the output of an information system is positively and tightly linked to the accuracy of its input. Yet, this belief has not been validated.

The relationship between an information system's input accuracy and its output accuracy is the subject of this paper. A traditional assumption in research of the relationship between input accuracy and output accuracy is that errors are independent. This assumption has the advantage that it greatly simplifies statistical-mathematical approaches. However, there is evidence that this assumption may not always be true in practice. In fact, there are numerous reasons why errors would not be independent in many cases. When the source of information is human, lack of knowledge can cause dependence between errors, reflecting the person's knowledge or skill gaps. People, as well as organizations, may choose strategies that would direct the provision of false information on numerous details. Computer input devices may suffer from defects that would cause dependence between errors in different data items. Conversion tables between information systems may have "bugs" such that when one output is incorrect it would go together with errors in many other outputs. And so on.

Motivated by a belief that dependence between errors is, in fact, common, my research examines the potential effect of the dependence factor on the relationship between an information system's input accuracy and its output accuracy. The main research questions are: (1) How does dependence between errors affect the relationship between input accuracy and output accuracy? (2) Is the relationship between input accuracy and output accuracy positive? The belief in a positive link will be called next the "monotonicity" assumption.

An earlier paper [1] analyzed these questions applying the statistical decision theory-based information structure model [7][12]. The analysis in [1] showed that in situations in which the relationship between the input and the output cannot be captured by a deterministic function (e.g., forecasts), dependence between errors in the input and "unexplained errors" can have powerful effect on the relationship between input accuracy and output accuracy. In particular, the relationship between input accuracy and output accuracy is not necessarily positive. That paper [1] also illustrated conditions of dependence between errors in the input and unexplained errors through a series of scenarios that referred to practical settings.

The inquiry in [1] is limited, however, to an information system that uses a single input for calculating the output. The current paper illustrates the results in [1], and extends the scope of the inquiry to situations in which an information system produces its output using multiple inputs. An additional class of dependence between errors that is studied in this context is dependence between errors in different inputs. This paper applies simulations for that purpose. The simulations portray simple information systems whose inputs are price, sale quantity, and/or cost data. If, for example, due to a general data timeliness issue, an occasional human error, or the like, input data are based on older customer orders instead of the most current ones, errors in different inputs may not be independent, because price, sale quantities, and cost data are often related.

The following section reviews related literature. Then, I describe the conceptual framework that underlies this study. The simulation scenario and method are described next. A subsequent section introduces the results. The final section discussed the results and future research directions.

LITERATURE REVIEW

A common understanding of the term accuracy in MIS research views a record as accurate if it is in accord with the actual value [2], [14].

Data quality research has traditionally assumed that errors are random or independent. Ballou and Pazer

[1985] propose a model for tracking numeric data errors through a system, to assist with estimates of the impact of these errors on the output. Their model takes into account both processing errors and errors due to input inaccuracy, though the emphasis is on how processing magnifies or dampens data errors. Ballou et al.[1998] present a methodology for estimating various quality dimensions of the output information. Output quality is estimated using a model that is similar to that in [Ballou and Pazer, 1985]. Parssian et al. [2004] present a methodology to help with assessments of the accuracy and completeness of outputs of relational algebra operations—selection, projection, and Cartesian product—given measures of the quality of source data. Ballou and Pazer [1990] analyze the impact of errors in estimates of decision criteria on decision accuracy. They focus on binary decisions in decision processes that apply a multi-criteria, conjunctive, satisficing decision rule. The outcome is a theory that supplies decision-makers with understanding as to what features of the decision problem magnify the likelihood of an incorrect decision. Mukhopadhyay and Cooper [13] analyze the relationship between data accuracy and decision accuracy in an inventory control decision-making problem. Their results confirm the microeconomic production-theoretic view that such relationship is positive, with diminishing marginal influence.

A series of studies explored the relationship between input accuracy and output accuracy empirically, assuming various prediction problems and model-building paradigms [8][9][10][11][5]. Errors were generated such that they were random. Nonetheless, results did not confirm monotonicity consistently, apparently due to sub-optimalities of model building paradigms. Klein and Rossin [10] investigated the influence of errors on the prediction accuracy of linear regression models in forecasting the net asset value of mutual funds. They found significant negative influence of both error rate and error magnitude in test data. Surprisingly, a higher error rate and higher error magnitude in training data increased the predictive accuracy compared to error-free data. Similar experiments with back-propagation neural network models [11] showed that error rate and error magnitude in test data are negatively related to prediction accuracy, however, a moderate error rate in training data had positive effect on the predictive accuracy compared to error-free data. Bansal et al. [5] compared the effect of errors in test data on the accuracy of linear regression and neural network models when forecasting the prepayment rate of mortgage-backed security portfolios. Error magnitude had significant negative effect on the predictive accuracy of both the linear regression and the neural network models, and on the payoff measure of the linear regression model. Error rate had significant negative effect on the predictive accuracy and payoff measure of the linear regression, but no effect on those of the neural network model. Hwarng [9] studied the effect of noise in training data on the accuracy of a back-propagation neural network model in time-series forecasts. All three possible relationships were observed: positive relationship, negative relationship, and no relationship. A noise level that is consistent with the magnitude of the standard deviation of the time series appeared to have positive effect.

Askira Gelman et al. [1] examined the effect of dependence between errors on the relationship between input accuracy and output accuracy based on a general model of information that is often called *information structure* [12]. In this context accuracy and informativeness were defined utilizing Blackwell's sufficiency criterion [6], [7]. This theoretical basis enabled to account for stochastic elements, and, in addition, to avoid a difficulty of earlier empirical studies. In those studies the bias, or specific properties, of the chosen forecasting methodology (or the information system application in general) affected the results. Therefore, the distinct relationship between input accuracy and output accuracy could not be clarified. The analysis in [1] concentrated instead on the maximal accuracy that can be achieved, independent of the choice of methodology and information processing in general, when using a given input data source for predicting the value of some designated output variable. The inquiry in [1] was limited, however, to the condition in which an information system uses a single input for calculating the output. This paper extends the scope of the inquiry to situations in which an information system produces output using multiple rather than a single input.

CONCEPTUAL FRAMEWORK

The question of the relationship between input accuracy and output accuracy is addressed under the assumption that information processing amount to a function that maps the values of a random variable (one-dimensional or multi-dimensional) to another random variable. Ideally, the input to an information system is error-free—such input is denoted S^i_j, j=1,..,n, in Figure 1. The output of the information system, $h(S^i_1..\ S^i_n)$, is an estimate of the value of a variable of interest S^o. Our analysis will correspond to such $h(S^i_1..\ S^i_n)$ that is the best possible estimate of S^o based on S^i_j, j=1,..,n. This estimate may, nonetheless, be inaccurate if the relationship between S^i_j, j=1,..,n, and S^o can not be captured by a deterministic function. The error term is viewed as a random variable that is classified in this case as an "unexplained error". Such error term will be denoted by ε.

In practice, however, S^i_j, j=1,..,n, are often not available. The available inputs, denoted Y_j, j=1,..n, are estimate of S^i_j, j=1,..,n, respectively, that may not be error-free; Y_j, j=1,..n, include an error term, denoted

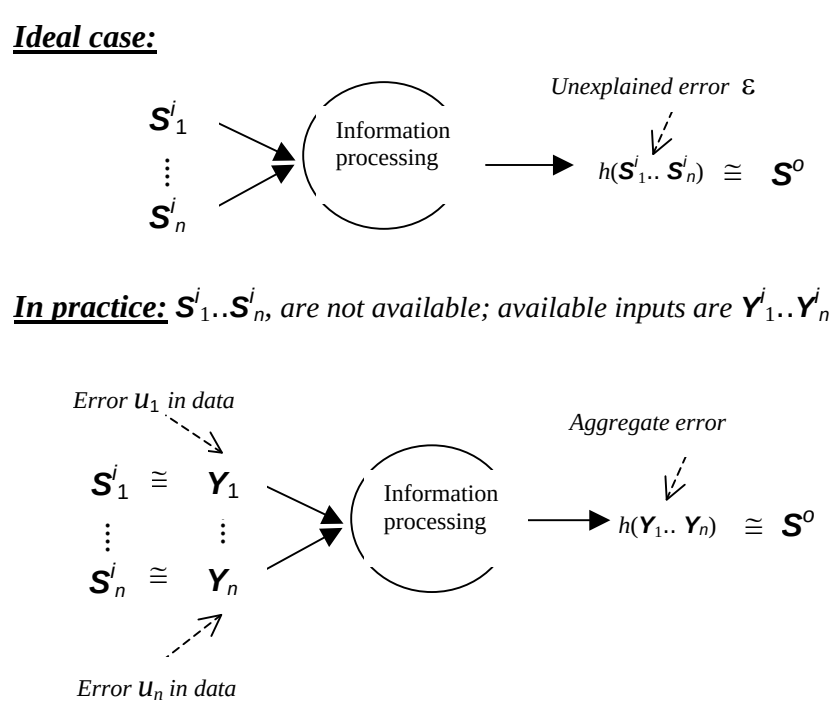

Figure 1: Conceptual framework

u_j, j=1,..n, respectively. The output, $h(Y_1..\ Y_n)$, will be defined such that it is, again, the optimal estimate of S^o based on $Y_1..\ Y_n$.

SIMULATIONS

It is common understanding that prices and demands are not independent. Marketing specialists apply this understanding on a daily basis—they manipulate demand through variations in price. Similarly, prices and costs are not independent in general.

Suppose, for example, that an application calculates revenues from customers, but a human mistake leads to a situation in which this application uses older records instead of basing its calculations on the most recent purchases of customers. As a result, when the number of units of a product that a customer bought one month ago is input to the system instead of the number of units that he or she bought today, such number might be incorrect. Moreover, the link that exists between quantity and price suggests that this error could be accompanied by a related error in price per unit data.

When older records are used in forecasting, e.g., when past sales data are unintentionally used in a forecasting task instead of more current data, then, again, it may actually happen that the resulting data inaccuracies will not be independent of unexplained errors that would occur when using the "correct" data. For example, in a seasonal market an error in the data that reflects the change of seasons might in fact be more indicative of a future trend than current data (see time-series).

The ensuing simulations refer to three simple information systems, one conducting a calculation based on price and cost data, and two others involving demand and revenue forecasting. The dependence between errors that is examined by the simulations can be assumed to have occurred in conditions like the ones just described.

Simulation Method

The relationship between input accuracy and output accuracy is approached through 146 simulations using quantitative data. The simulations were performed with GAUSS Light, a mathematical and statistical programming language.

Examined factors: The simulations explore the effect of two variables—input accuracy, and dependence between errors, on output accuracy. Input accuracy is operationalized through error magnitude; errors were generated from normal distributions with mean zero, such that their magnitude is measured by the standard deviation of such distributions. Dependence is operationalized by the correlation coefficient.

The dependent variable: Output accuracy is measured using Root Mean Squared Error (RMSE):

$$\text{RMSE} = [1/n\textstyle\sum_i(\boldsymbol{w}_i - \hat{w}_i)^2]^{1/2}$$

n is the number of data items, $\boldsymbol{w}_i$ denotes the correct value of a variable, $\hat{w}_i$ denotes the predicted value of such variable, $i = 1..\ n$. RMSE provides information about the magnitude of the error; higher RMSE implies lower accuracy.

Simulation design: 56 simulations center on the dependence between the error in the input and an unexplained error. These simulations relate to a single-input system, which forecasts costomers' future demand based on past sales. Specifically, the relationship between $\boldsymbol{S}^i_D$ and $\boldsymbol{S}^o_D$ is assumed to be captured by the model:

$$\boldsymbol{S}^o_D = \boldsymbol{S}^i_D + \varepsilon_D\ .$$

56 additional simulations center on the dependence between errors in different inputs. In these simulations the information system is captured by a deterministic function that calculates profit on

individual products. Given $\boldsymbol{S}^i_P$, designating the price charged, and $\boldsymbol{S}^i_C$—the overall cost of the product, this function outputs the difference:

$$\boldsymbol{S}^o_\Pi = 0.5*\boldsymbol{S}^i_P - \boldsymbol{S}^i_C .$$

The coefficient (0.5) may be taken to match a situation in which only part of the charge is transferred back to the producer—the purpose of this coefficient purpose in these simulations is to provide some insight about the role of the information system.

The remaining 34 simulations examine the combined effect of dependence between errors in different inputs and dependence between errors in the input and unexplained errors with a system that predicts future revenues. The relationship between current and future revenue is taken to be:

$$\boldsymbol{S}^o_R = 0.5*\boldsymbol{S}^i_P * \boldsymbol{S}^i_D + \varepsilon_R.$$

For each of these three simulation classes, 1000 values of the error-free input sources were generated under the assumption that $\boldsymbol{S}^i_P \sim U[300,400]$, and $\boldsymbol{S}^i_C, \boldsymbol{S}^i_D \sim U[100,200]$. The same values were used in all the simulations in a particular class.

Values of the unexplained error, ε, were created under the assumption that $\varepsilon_R, \varepsilon_D \sim N(0, \sigma_\varepsilon^2)$; where $\sigma_{\varepsilon D}$=6 (one input system), $\sigma_{\varepsilon R}$=800 (stochastic two-input system). Again, the same 1000 values of the unexplained error were used in all the simulations in a particular class. The values of $\boldsymbol{S}^o_\Pi$, $\boldsymbol{S}^o_R$, $\boldsymbol{S}^o_D$, were calculated based on the generated values of the inputs and, in the first and third simulation classes, the matching unexplained error values.

Errors in the input were produced from normal distributions, with zero means, according to the following rules (see also Table 1). In the simulations of the single input system, error magnitude was manipulated through the choice of the respective standard deviation, denoted next σ_{uD}, such that five values were tried: σ_{uD} =3,6,9,12,15. In simulations of a two-input system, the values of an error u_P were drawn with a fixed standard deviation, σ_{u_P}=12. Values of u_C were created such that five standard deviation values were tried: σ_{uC}=3,6,9,12,15. In simulations of a two-input stochastic system, the value of u_D corresponded to a fixed standard deviation, σ_{u_D}=2. Values of u_P were created such that three standard deviation values were tried: σ_{u_P}=2,6,12.

The simulations targeted diverse dependence levels. In the simulations of the single input system and the deterministic two-input system, 11 different dependence levels were created for each error magnitude which were distributed evenly, more or less, in the range -1 to +1 of correlation coefficients. In the remaining simulations, simulations corresponded to 11 dependence types, as follows:

- Dependence type 1: u_P, u_D, are each maximally negatively correlated with ε_R (i.e., correlation coefficients are each equal -1)
- Dependence type 2: u_P is maximally negatively correlated with ε_R.
- Dependence type 3: u_P, u_D, are maximally positively correlated (correlation coefficient is +1).
- Dependence type 4: u_D is maximally negatively correlated with ε_R.
- Dependence type 5: u_P is maximally negatively correlated with ε_R. u_D is maximally positively correlated with ε_R.
- Dependence type 6: u_P, u_D and ε_R are independent.
- Dependence type 7: u_D is maximally positively correlated with ε_R.
- Dependence type 8: u_P, u_D, are maximally negatively correlated.

- Dependence type 9: u_P, u_D, are each maximally positively correlated with ε_R.
- Dependence type 10: u_P is maximally positively correlated with ε_R.
- Dependence type 11: u_P is maximally positively correlated with ε_R. u_D is maximally negatively correlated with ε_R.

One-input system	**Error rate** 0%	1 simulation
	Error magnitude: σ_{uD} =3,6,9,12,15 ($\sigma_{\varepsilon D}$ =6) **Dependence levels:** 11 levels	5 **x** 11 = 55 simulations
Two-input system (deterministic	**Error-free data**	1 simulation
	Error magnitude: σ_{u_C}=3,6,9,12,15 (σ_{u_P}=12) **Dependence levels:** 11 levels	5 **x** 11 = 55 simulations
Two-input system (stochastic setting)	**Error free data**	1 simulation
	Error magnitude: σ_{u_P}=2,6,12 (σ_{u_D}=2, $\sigma_{\varepsilon R}$=800) **Dependence levels:** 11 types	3 **x** 11 = 33 simulations
		Total: 146 simulations

Table 1: Simulation design

The values of $\boldsymbol{Y}_C$, $\boldsymbol{Y}_D$, and $\boldsymbol{Y}_P$ were calculated based on the values of $\boldsymbol{S}^I_C$, $\boldsymbol{S}^I_D$, $\boldsymbol{S}^I_P$, and the matching error values that were created. These calculated values were applied to the information system models in place of $\boldsymbol{S}^I_C$, $\boldsymbol{S}^I_D$, $\boldsymbol{S}^I_P$, respectively, to derive estimates of $\boldsymbol{S}^I_P$, $\boldsymbol{S}^o_\Pi$, $\boldsymbol{S}^o_R$. The optimality of such estimates in terms of prediction error bias and variance can be easily proved. These estimates were used in conjunction with the "true" values of $\boldsymbol{S}^I_P$, $\boldsymbol{S}^o_\Pi$, $\boldsymbol{S}^o_R$ for calculating the RMSE values.

RESULTS

The results of the simulations are depicted in Figure 2, Figure 3, and Figure 4, corresponding to the three types of information systems that were simulated. Figure 2 represents the simulations of a one-input

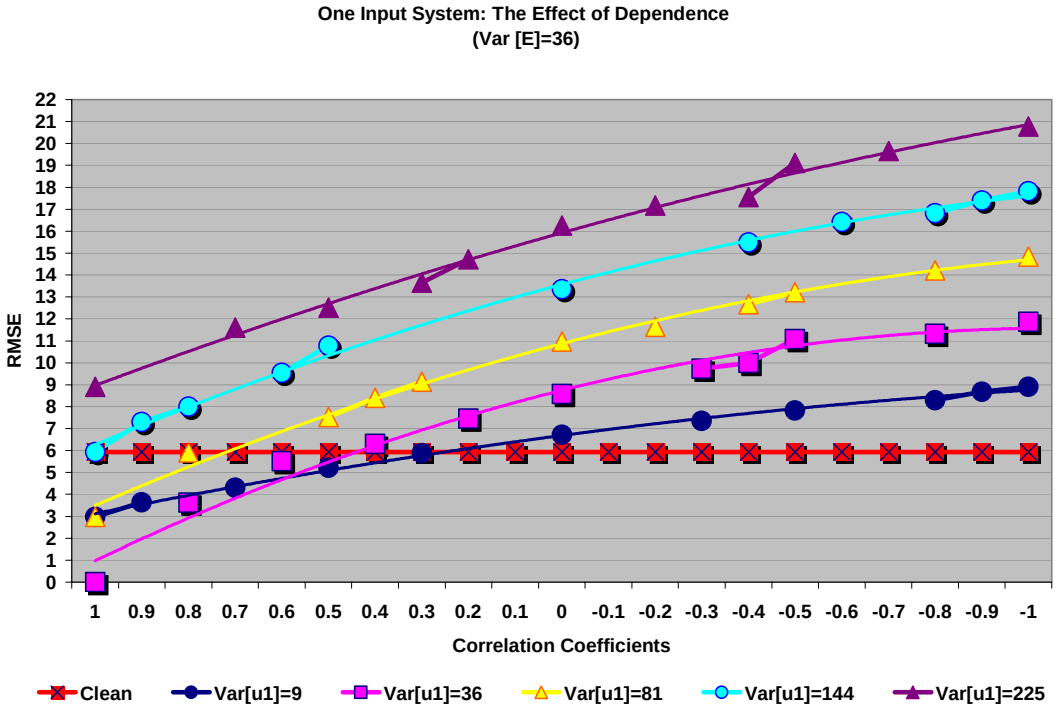

Figure 2: The effect of dependence between errors in data and unexplained errors in a one-input system

system, Figure 3 represents the simulations of a two-input system in deterministic settings, and Figure 4 matches the simulations of a two-input system in stochastic settings. Each distinct line depicts the results under specific assumptions on error magnitudes. The Y-axis shows RMSE values. The X-axis portrays dependence between errors. In Figure 2 and Figure 3 dependence between errors is represented by correlation coefficients, while in Figure 4 dependence between errors matches an earlier specification. Unlike Figure 2 and Figure 3 that aim to show how prediction accuracy varies with dependence along a continuum of dependence levels, Figure 4 provides an outline through focus on extreme points, where one or more dependence values are maximal.

Dependence between errors has a strong effect: The strong influence of dependence between errors on the relationship between input accuracy and output accuracy is demonstrated in all the charts. Given any error magnitude and system, output accuracy varies widely together with the correlation between the errors.

Dependence versus independence: Given any system and error magnitude, dependence between errors can generate output accuracy values that are far superior to output accuracy when the errors are independent. On the other hand, output accuracy can be substantially inferior to that when such system applies inputs where errors are independent.

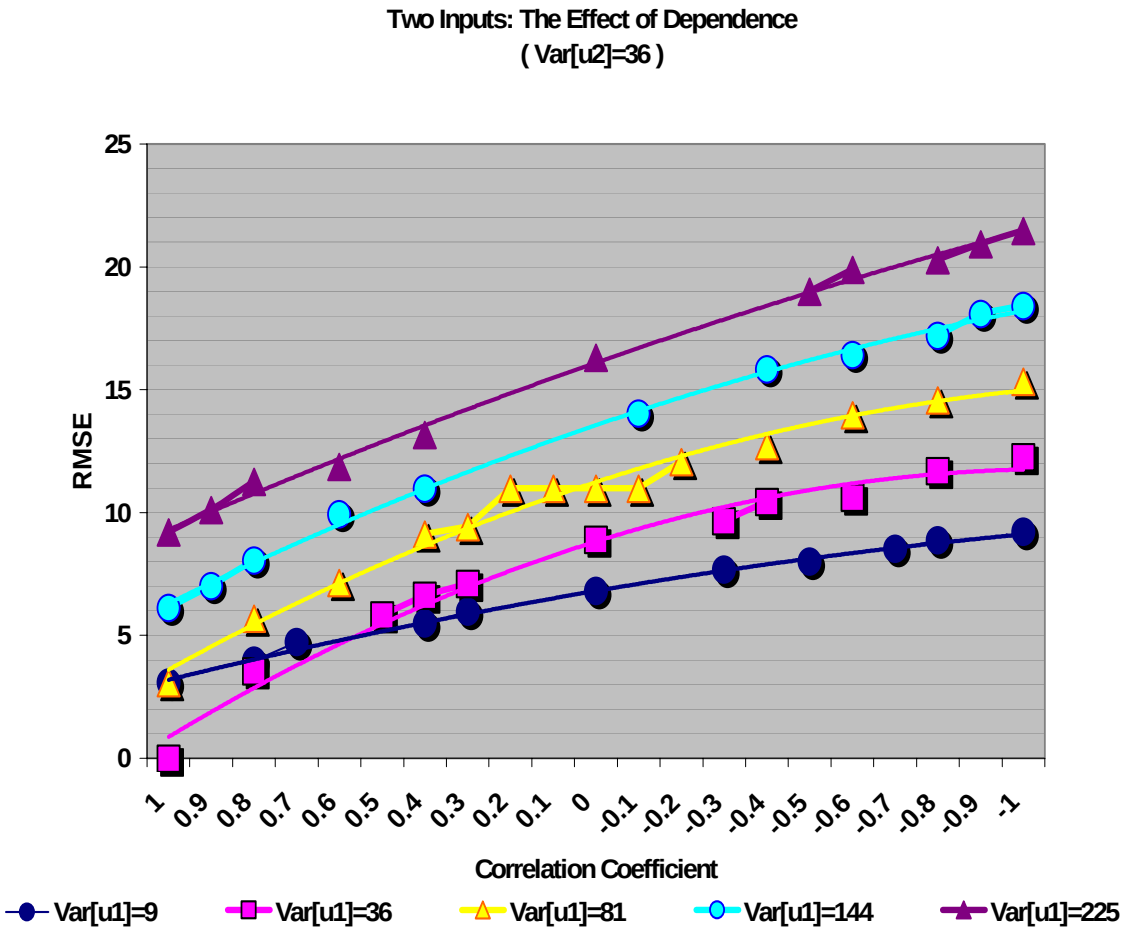

Figure 3 The effect of dependence between input errors in a two-input system

Dependence between errors versus error-free data: Dependence between errors can generate output that is just as accurate as output that was created by processing error-free data. In the forecasting applications (Figure 2 and Figure 4), it is shown that output created through the use of inaccurate data can yield results that are superior to results with error-free data (notice that graphs reach below the horizontal lines that represent clean data in these two charts).

Data quality improvements: Dependence between errors is, at least partially, an independent factor, such that higher input accuracy can go together with a concurrent change in dependence, either way. Hence, the outcome in terms of output accuracy may amount to improvement in some cases, or decline in others. For example, consider Figure 3. Assume that a decrease in error magnitude where error variance drops from 81 to 36 is accompanied by a shift from positive dependence 0.4 to independence, i.e., correlation is zero. In this case output accuracy would not increase and may even decrease somehow. On the other hand, output accuracy will be boosted if the same improvement in input accuracy goes together with higher dependence between errors.

The effect of similarity in error magnitudes: Similarity in magnitude generates the most dramatic improvements in output accuracy. In Figure 2 this power is demonstrated in the graph that matches error variance 36. This error variance is the same as the error variance of the unexplained error in that application. When the dependence between the errors is maximal they cancel each other completely and produce perfectly accurate forecasts. A complicating factor is that information processing can dampen, or, alternatively, augment error magnitude drastically [Ballou, 1985]. Figure 4 shows a dramatic amplification as a result of the multiplication operation that it involves. Consider also Figure 3: the best results are achieved when error variance is 36, despite the fact that this value is different from the variance of the error in the second input, which is 144 (σ_{u_p}=12). This result is due to the moderating effect of the application, which lowers the magnitude of the error when it multiplies it by 0.5 (which implies that error variance is now 36.)

The effect of direction of correlation: In a similar way, information processing can reverse the direction of dependence between errors. Positive correlation can transform to negative correlation and vice versa. The model that underlies Figure 3 demonstrates this capacity when cost is subtracted from price in the function there.

In sum, while dependence between errors can have a powerful effect on output accuracy, assessing the actual impact can be challenging.

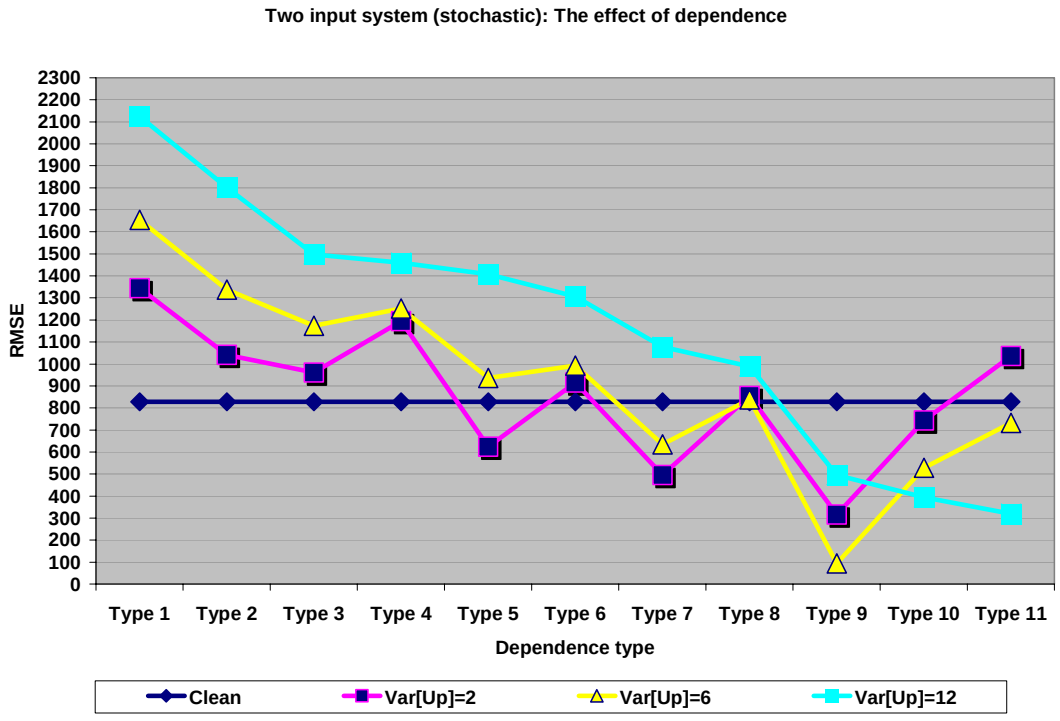

Figure 4: The effect of dependence between errors in a two input (stochastic) system

CONCLUSIONS

Users of information systems might be familiar with instances of dependence between errors due to human lack of knowledge, faulty input devices, "buggy" interfaces between information systems, or other causes. The analysis in this paper takes likewise scenarios one step further by assuming that errors moderate or, alternatively, supplement, each other, *regularly*. Furthermore, I assume that such relationship between errors is possible not only for errors in different inputs, but also for errors in inputs and unexplained errors.

The simulations in this paper demonstrate the important effect of dependence between errors in quantitative data on the relationship between input accuracy and output accuracy. The findings suggest that dependence between errors can reinforce an increase in data accuracy or moderate and even outweigh it, such that the outcome can vary greatly due to effects of dependence, and, in particular, higher output accuracy is not guaranteed.

While earlier work demonstrated such results for the case of dependence between errors in input data and unexplained errors in non-deterministic settings, this paper uncovered another kind of dependence between errors that can have influence both when the information system implements a deterministic function, and in non-deterministic settings. This is dependence between errors in different inputs of the information system.

From a research viewpoint, the accumulated results highlight the sensitivity of discoveries about the relationship between input accuracy and output accuracy to assumptions on the dependence factor. Since MIS research has been conducted till now under a narrow set of assumptions that have only been partly justified in terms of their validity in real-world settings, the findings imply that future research can gain from a more careful examination of the validity of traditional assumptions in practical settings. To the extent that dependence between errors is indeed common, it would be useful to develop understanding of recurrent dependence patterns and ground future analyses in realistic assumptions.

From a practical standpoint, the findings suggest that cost benefit analyses of data quality projects may benefit from studies of the dependence between errors. Nonetheless, the applicability of the new understanding to practical situations is limited at this stage. Translation of such understanding to practical approaches requires further study. There is a need to achieve better understanding of recurrent circumstances that encourage actual dependence between errors, common error dependence patterns in typical applications, and effects of such dependence patterns on output accuracy and economic value. In addition, a methodological approach for the identification, assessment, and resolution of dependence between errors can be useful for addressing the issue in practical situations. For example, a search for factors that have the potential to create actual dependence may be part of such methodology. It might help focus on specific applications where dependence is an issue, and provide further insight about the shape of such dependence and its effect on output accuracy. Still, more precise assessments of the effect of dependence may turn, in the general case, to be a challenging task.

REFERENCES

[1] Askira Gelman, I., D. Pingry, and D. Zeng, "*Does Higher Data Accuracy Produce Higher Prediction Accuracy?*" International Workshop on Data and Information Quality in conjunction with CAISE'04, Riga, Latvia, 2004.

[2] Ballou, D. P. and Pazer, H. L., "Modeling Data and Process Quality in Multi-input, Multi-output Information Systems." *Management Science*, Vol. 31, No. 2, 1985, pp. 150-162

[3] Ballou, D. P., Pazer, H. L., Belardo, S., and Klein, B.,"Implication of Data Quality for Spreadsheet Analysis." *DATA BASE*, Vol. 18, No. 3, 1987, pp. 13-19.

[4] Ballou, D. P., Wang, R. Y., Pazer, H. L., and Tayi, G. K., "Modeling Information Manufacturing Systems to Determine Information Product Quality." *Management Science*, Vol. 44, No. 4, 1998, pp. 462-484.

[5] Bansal, A., Kauffman, R. J. and Weitz, R. R., "Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: A Business Value Approach.", *Journal of Management Information Systems*, Vol. 10, No. 1, 1993, pp. 11 - 32.

[6] Blackwell, D. "Comparison of Experiments." *Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 93-102.

[7] Blackwell, D. "Equivalent comparisons of experiments." *Annals of Mathematical Statistics* Vol. 24, No. 2, 1953, pp. 265-272.

[8] Haughton, D., Robbert, A., Senne, L.P., and Gada, V.,"Effect of Dirty Data on Analysis Results." 8th *International Conference on Information Quality* (ICIQ-2003), 2003, pp. 64-79.

[9] Hwarng, H.B., "Insights into Neural-Network Forecasting of Time Series Corresponding to ARMA(p, q) Structures." *Omega*, Vol. 29, No. 3, 2001, pp. 273-289.

[10] Klein, B.D., and Rossin, D.F., "Data Quality in Linear Regression Models: Effect of Errors in Test Data and Errors in Training Data on Predictive Accuracy." *Informing Science*, Vol. 2, No. 2, 1999.

[11] Klein, B.D., and Rossin, D.F., "Data Quality in Neural Network Models: Effect of Error Rate and Magnitude of Error on Predictive Accuracy." *Omega*, Vol. 27, No. 5, 1999, pp. 569-582

[12] Marschak, J., "Economics of Information Systems." *Journal of American Statistical Association*, Vol 66, No. 333, March 1971, pp. 192-219.

[13] Mukhopadhyay, T., and Cooper, R.B., "The Impact of Management Information System on Decisions", *OMEGA (International Journal of Management Science)*, Vol. 20, No. 1, 1992, pp. 37-49.

[14] Wand, Y. and Wang, R. Y., "Anchoring Data Quality Dimensions in Ontological Foundations.", *Communications of the ACM* , Vol. 39 , No. 11 , November 1996 , pp. 86 -95.