Inferring Transactions from Financial Statements*

ANIL ARYA, *The Ohio State University*

JOHN C. FELLINGHAM, *The Ohio State University*

JONATHAN C. GLOVER, *Carnegie-Mellon University*

DOUGLAS A. SCHROEDER, *The Ohio State University*

GILBERT STRANG, *Massachusetts Institute of Technology*

Abstract

In this paper, we embed the double entry accounting structure in a simple belief revision (estimation) problem. We ask the following question: Presented with a set of financial statements (and priors), what is the reader's "best guess" of the underlying transactions that generated these statements? Two properties of accounting information facilitate a particularly simple closed form solution to this estimation problem. First, accounting information is the outcome of a linear aggregation process. Second, the aggregation rule is double entry.

Keywords Double entry; Aggregation; Subspaces; Estimation

Condensé

Les auteurs enchâssent la structure de la comptabilité en partie double dans un problème simple de révision d'opinion (estimation). Ils posent la question suivante : quelle sera la meilleure estimation du lecteur à qui l'on présente un jeu d'états financiers (et de préalables) en ce qui a trait aux opérations sous-jacentes qui ont généré ces états financiers ? Deux propriétés de l'information comptable sont propices à une solution de nature fermée particulièrement simple. Premièrement, l'information comptable est le résultat d'un processus d'agrégation linéaire. Deuxièmement, la règle d'agrégation est la consignation en partie double.

La résolution intégrale du problème d'estimation comporte quatre étapes.

1. La construction d'une représentation algébrique linéaire du processus comptable. Cette première étape nécessite la définition d'une matrice à double entrée permettant de transformer un grand nombre d'opérations en un petit nombre de soldes de comptes.
2. La définition des sous-espaces fondamentaux (c'est-à-dire l'espace rangée, l'espace nul, l'espace colonne et l'espace nul gauche) de la matrice à double entrée. Dans un problème d'algèbre linéaire caractéristique, le choix d'une base pour l'espace nul peut

* Accepted by Jerry Feltham. We thank Rick Antle, Bill Beaver, Joel Demski, Ron Dye, Jerry Feltham (editor), Karl Hackenbrack, Chuck Horngren, Yuji Ijiri, Richard Lambert, Brian Mittendorf, Paul Newman, Robert Swieringa, Rick Young, David Ziebart, two anonymous referees, and workshop participants at Case Western, Florida, Ohio State, Purdue, and Yale for helpful comments.

être un exercice fastidieux. Heureusement, le problème comptable ne soulève pas ce genre de difficulté puisque la base peut être définie au premier coup d'œil. Cela découle directement du principe de double entrée de la comptabilité en vertu duquel la matrice linéaire de transformation comptable est une matrice d'incidence. La matrice d'incidence, à son tour, permet la représentation géométrique des relations comptables sous forme de diagramme orienté.

3. À partir des états financiers, la résolution du problème pour tous les vecteurs d'opérations susceptibles d'avoir généré les états financiers. Il s'agit en d'autres termes de remonter des états financiers aux opérations. C'est là que les sous-espaces fondamentaux ont leur utilité. Tous les vecteurs d'opérations cohérents présentent le même élément espace rangée. Ils ne diffèrent que dans le poids accordé aux vecteurs d'espaces nuls.
4. À partir des vecteurs d'opérations possibles définis à l'étape précédente, le choix du vecteur d'opérations qui minimise la fonction de perte du lecteur. La décomposition orthogonale des vecteurs d'opérations cohérents en leurs éléments espace rangée et espace nul se révèle appropriée dans le calcul des opinions postérieures du lecteur. Dans le contexte de l'analyse, les convictions préalables du lecteur garantissent l'indépendance des deux éléments. Cela, ajouté au fait que toute l'information de l'espace nul est perdue en cours d'agrégation, suppose que la conviction postérieure du lecteur de l'élément espace nul est la même que sa conviction préalable — aucune mise à jour ne touche l'espace nul. D'autre part, une mise à jour explicite touche l'espace rangée. Étant donné un ensemble particulier de soldes de comptes, il existe un et un seul élément espace rangée possible. La somme des deux éléments donne la moyenne postérieure du lecteur. En supposant une fonction de perte quadratique, la moyenne postérieure correspond à la meilleure estimation du lecteur.

Les auteurs étendent brièvement la caractérisation algébrique linéaire à l'exploration d'un problème de classification dans lequel la tâche du lecteur consiste à classer l'entreprise dans l'une de deux catégories, à partir des états financiers observés. Ce problème est quelque peu délicat du fait que la comptabilité en partie double induit une interdépendance entre les comptes, même lorsque les opérations sont indépendantes. Les plates-formes comptables sont comparées sur le plan de la capacité de discrimination. Il est possible de mesurer le pouvoir de classification d'une plate-forme en calculant R^2 à partir d'une régression de l'écart des moyennes des opérations pour les types d'entreprises des rangées de la plate-forme comptable. La plate-forme qui présente un R^2 plus élevé a un pouvoir de classification supérieur pour le problème particulier à l'étude. De plus, si l'espace nul d'une plate-forme est un sous-espace de l'espace nul de l'autre plate-forme, la première plate-forme est (légèrement) préférable à la seconde dans tous les contextes où l'on souhaite davantage d'information.

1. Introduction

Linear representations have proved illuminating in analyses of many fundamental accounting problems. The use of local linear approximations to represent cost functions, the formulation of constrained profit maximization problems, the equivalence of reciprocal cost allocation, and a linear program that does not require any allocation are some examples that come to mind.[1] The most basic linear transformation met by accountants is the double entry system that converts transaction amounts into financial statements.

The paper takes its cue from Demski 1992. In discussing the information content school, Demski (1992, 4–5) writes, "In broad terms, this school is based on the economics of uncertainty. Accounting enters as a source of information. ... Use of the accountant's services is endogenous here. In a single person theory, for example, Bayesian revision, given exogenous specification of beliefs, provides a more endogenous view of accounting. ... This legacy of the information content school does not come without cost. The modeling is abstract (as it should be) and largely without accounting (as it should not be)."

Our aim is modest. In this paper, we embed the double entry accounting structure in a simple belief revision (estimation) problem. We ask the following question: Presented with a set of financial statements (and priors), what is the reader's "best guess" of the underlying transactions that generated these statements?[2] Two properties of accounting information facilitate a particularly simple closed form solution to this estimation problem. First, accounting information is the outcome of a linear aggregation process. Second, the aggregation rule is double entry.[3]

The complete solution to the estimation problem is determined in four steps.

1. Construct a linear algebraic representation of the accounting process. Specify a double entry matrix that transforms a large number of transactions into a small number of account balances.
2. Identify the fundamental subspaces (i.e., the row space, nullspace, column space, and left nullspace) of the double entry matrix. In a typical linear algebra problem finding a basis for the nullspace can be a tedious exercise. Happily, in the accounting problem this is not the case. A basis is available at a glance. This result follows directly from the double entry property of accounting, which ensures that the accounting linear transformation matrix is an incidence matrix. The incidence matrix, in turn, allows a geometric representation of the accounting relationships in the form of a directed graph.
3. Given financial statements, solve for all possible transactions vectors that could have generated the statements. That is, invert from financial statements to transactions. This is where the fundamental subspaces come in handy. All consistent transaction vectors have the same row space component. They differ only in the weight they place on the nullspace vectors.
4. From the feasible transaction vectors identified in the previous step, pick the transaction vector that minimizes the reader's loss function. The orthogonal decomposition of consistent transaction vectors into their row space and nullspace components proves convenient in computing the reader's posterior beliefs. In our setting, the reader's priors guarantee that the two components are independent. This, and the fact that all information in the nullspace is lost during aggregation, implies that the reader's posterior belief of the nullspace component is the same as her or his prior — no updating occurs in the nullspace. On the other hand, crisp updating occurs in the row space. Given a particular set of account balances, there is one and only one possible row space component.

The sum of the two components yields the reader's posterior mean. Assuming a quadratic loss function, the posterior mean is the reader's best guess.

The linear algebra characterization is extended briefly to explore a classification problem in which the reader's task is to classify the firm as one of two firm types based on the observed financial statements. This problem is a bit delicate in that double entry accounting induces interdependence in accounts even when transactions are independent. Accounting platforms are compared based on their ability to discriminate. A platform's classification power can be measured by computing the R^2 from a regression of the difference in transactions means for the firm types on the rows of the accounting platform. The platform with a larger R^2 has greater classification power for the specific problem at hand. Further, if the nullspace of one platform is a subspace of the nullspace of the other platform, the former platform is (weakly) preferred to the latter platform in all settings where more information is desired.

There is a long tradition in accounting of formally modeling the accounting system (e.g., Butterworth 1972; Mattessich 1964; Williams and Griffin 1964a, b; Ijiri 1971). Such analysis is inherently abstract and parsimonious, and we believe, as have other authors before us, has far-reaching impact. For example, in discussing the application of "matrix theory" to cost allocation, Williams and Griffin (1964a, 678) remark, "To the extent that matrix formulations of accounting analysis contribute to more rigorous and logical models of accounting theory, additional support is adduced for such methods. Surely this use provides an appropriate complement to the more practical benefits immediately available to the accounting practitioner."

Ijiri (1971, 780) makes a broader connection between the study of aggregation and scientific investigation: "The significance of aggregation theory lies not just in its formal theorems but in its implicit capacity to help theories evolve." While the estimation problem is not explicitly dealt with in Ijiri 1967, 1968, 1971, the language (and tools) developed in these papers help in the estimation exercise. Representing the accounting system as an incidence matrix, the use of pseudoinverses, and the use of a linear aggregation coefficient as a measure of information retained (lost) are ideas present in Ijiri's work that we employ.

The remainder of the paper is organized as follows. Section 2 presents the estimation problem. Section 3 presents the solution. Characterizing the invertibility solution in terms of the fundamental subspaces of the double entry matrix is the critical aid in deriving the reader's best guess for transactions. Section 4 extends the discussion to a classification problem. Section 5 concludes the paper.

2. The question

The problem we study in this paper is as follows. Presented with a set of financial statements, $\boldsymbol{x}$, what is the reader's "best guess" of the underlying (unknown) transactions, $\boldsymbol{y}$? We assume the reader's best guess is one that satisfies the following two criteria. First, it must be consistent with the observed financial statements. Second, from all consistent transaction vectors, it minimizes the expected loss (assuming the loss function is quadratic).

We assume the reader's priors (before observing $\boldsymbol{x}$) are that $\boldsymbol{y}$ is normally distributed with mean $\bar{\boldsymbol{y}}$ and identity variance–covariance matrix.[4]

The solution to the problem is determined in four steps.

1. Construct a linear algebraic representation of the accounting process: $\boldsymbol{Ay} = \boldsymbol{x}$.
2. Identify the fundamental subspaces of the double entry matrix $\boldsymbol{A}$.
3. Given $\boldsymbol{x}$, solve for all consistent $\boldsymbol{y}$ vectors — that is, $\boldsymbol{y}$ vectors satisfying $\boldsymbol{Ay} = \boldsymbol{x}$.
4. Given the reader's priors over $\boldsymbol{y}$, derive the conditional mean of the consistent $\boldsymbol{y}$ vectors identified in the previous step.

The first step is possible because accounting aggregates using a linear transformation: several transactions are summarized by a few account balances using a double entry rule. The double entry rule also eases the second step. It allows for a directed graph representation of the accounting system from which the fundamental subspaces can be easily derived. The fundamental subspaces, in turn, prove crucial in step 3. Information about the transactions residing in one of the fundamental subspaces of $\boldsymbol{A}$, the nullspace, is lost during aggregation. In contrast, transaction information residing in the row space of $\boldsymbol{A}$ is retained. The conditional (posterior) mean in step 4 then reflects (1) that the nullspace component of transactions is lost during aggregation and (2) that the nullspace and the row space components are statistically independent. For any quadratic loss function, the conditional mean is the reader's best guess (i.e., the reader's Bayesian point estimate).[5]

We will derive a general proof and use the following numerical example for illustration purposes. A firm undertakes the following seven transactions:

1. purchase inventory for cash
2. plant acquisition for cash
3. cash expenses
4. cash sales
5. cost of goods sold
6. depreciation (product cost)
7. depreciation (period cost)

Assume the reader's priors (before financial statements are presented) are that the expected value of the seven transaction amounts are 7, 9, 1, 10, 5, 1, and 2, respectively. At the end of the period, the reader is presented with the firm's account balances as follows:

Balance sheet	Ending balance	Beginning balance
Cash	2	10
Inventory	4	0
Plant	6	0
Owners' equity	12	10

Income statement

Sales	10
Cost of goods sold	5
Gen'l and admin.	3
Income	2

The question we ask is the following: What is the reader's best guess regarding the transaction amounts that created these financial statements?

Clearly, the reader's priors are not consistent with the observed financial statements. For example, the priors would imply that the ending cash balance should be 3.[6] However, the ending cash balance is 2. Hence, the reader's aim is to determine all $\boldsymbol{y}$ vectors that are consistent with the given financial statement and, from among all such feasible $\boldsymbol{y}$'s, to pick the $\boldsymbol{y}$ vector that minimizes her or his expected loss.

3. The answer

Step 1: Linear algebraic representation of the double entry process

The core of a linear representation of a double entry system is a transformation matrix denoted by $\boldsymbol{A}$. The matrix $\boldsymbol{A}$ has m rows and n columns, where m is the number of accounts and n is the number of transactions. Accounting aggregates so that $m < n$. There are two non-zero entries in each column; the non-zero entries denote the accounts that are "connected" by the transaction journal entry.We adopt the following sign convention for the non-zero entries: debits to an account are denoted by +1, and credits are denoted by −1. The $\boldsymbol{A}$ matrix for our example is

$$\begin{array}{r} \text{Cash} \\ \text{Inventory} \\ \text{Plant} \\ \text{Sales} \\ \text{Cost of goods sold} \\ \text{Gen'l and admin.} \end{array} \begin{bmatrix} -1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Denote by $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$, the vector of changes in account balances. The convention we use is to multiply accounts that have a credit balance by −1 and those with a debit balance by +1. (This is the same convention used when recording the $\boldsymbol{A}$ matrix.) For our example, $\boldsymbol{x}$ is (−8, 4, 6, −10, 5, 3). That is, during the period, cash declined (net credit) by 8, inventory increased (net debit) by 4, and so forth. The elements in $\boldsymbol{x}$ sum to zero. This represents the basic accounting identity that assets equals liabilities plus owners' equity.

Denote by $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$, the vector of transaction amounts. Vector $\boldsymbol{x}$ is prepared by aggregating $\boldsymbol{y}$ using the transformation $\boldsymbol{A}$: $\boldsymbol{Ay} = \boldsymbol{x}$. For example, if $\boldsymbol{y}$ = (8, 9, 1, 10, 5, 1, 2), then $\boldsymbol{Ay}$ = (−8, 4, 6, −10, 5, 3) = $\boldsymbol{x}$. The matrix representation is possible because bookkeeping is a linear process.

The double entry matrix, as is true of any matrix, is associated with four fundamental subspaces: the row space, the left nullspace, the column space, and the nullspace. What is special about the double entry matrix is that each column consists of +1 and −1. This incidence property allows for a directed graph representation of the double entry process. The graph makes the task of identifying the fundamental subspaces of the matrix straightforward.

Step 2: The fundamental subspaces of A

The double entry matrix can be represented by a graph of edges connecting nodes. The nodes correspond to accounts, and edges between two nodes correspond to transactions. The graph is "directed" when we assign an arrow to each edge. We adopt the convention that if an edge goes from node j to node k, then the column has −1 in row j and +1 in row k. In other words, the arrow goes away from the account credited and points toward the account debited. For our example the directed graph is as shown in Figure 1.

Left nullspace

The left nullspace of a matrix $\boldsymbol{A}$ consists of all vectors $\boldsymbol{w}$ orthogonal to each column in $\boldsymbol{A}$. This means that $\boldsymbol{w^T}$ appears on the left of $\boldsymbol{A}$: $\boldsymbol{w^T A} = 0$. For an incidence matrix, adding all the rows in $\boldsymbol{A}$ (i.e., weighting each row by 1) yields the zero vector. Hence, the vector of all 1's resides in the left nullspace. The vector of all 1s represents balancing in accounting: assets equal liabilities plus owners' equity. Furthermore, if the graph is connected, there is one and only one independent vector in the left nullspace. A graph is connected when starting from any node (account) there is a path (sequence of transactions) to any other node. Hence, in our example, the dimension of the left nullspace is 1 and the vector of 1s (the balancing vector) is a basis.

Row space

In the language of accountants, knowing the balances in any $m - 1$ accounts and using the accounting identity allows calculation of the balance in the remaining account. In the language of linear algebra, any $m - 1$ rows of the matrix $\boldsymbol{A}$ are independent, but all m rows sum to the zero row. Hence, the dimension of the row space is $m - 1$, and any $m - 1$ rows of $\boldsymbol{A}$ form a basis.

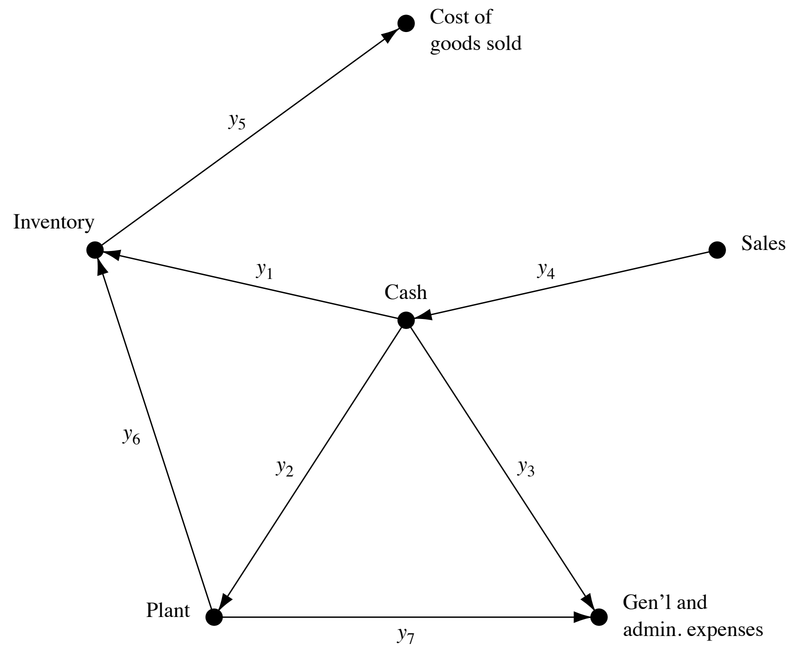
To create the T account for account i, one can simply multiply the ith row of the $\boldsymbol{A}$ matrix with the vector of transaction amounts. Hence, we refer to the space spanned by the rows as the T account space.

Nullspace

To solve the equation $\boldsymbol{Ay} = \boldsymbol{x}$ for $\boldsymbol{y}$, we will need to know all the solutions to $\boldsymbol{Ay} = 0$. The nullspace of $\boldsymbol{A}$ consists of vectors, $\boldsymbol{y}$, that solve the equation $\boldsymbol{Ay} = 0$. That is, the nullspace consists of transaction vectors that produce zero changes in the account balances. The loops of the directed graph form a basis for the nullspace.

In the example there are two independent loops. One loop is to start from the cash node, move along transaction 2 to Plant, then move along transaction 6 to Inventory and then proceed along (minus) transaction 1 back to the starting node.

Figure 1 A directed graph for the 6 by 7 double entry matrix



The seven-element vector representing this loop has 1 in the second position, 1 in the sixth position, −1 in the first position, and zero elsewhere: (−1, 1, 0, 0, 0, 1, 0). Similarly, a second loop connecting Cash – Plant – G & A can be expressed as (0, 1, −1, 0, 0, 0, 1). Note that $\boldsymbol{A}$ times these vectors (or any linear combination of these vectors) is zero. Denote the nullspace matrix as $\boldsymbol{N}$, where the rows of $\boldsymbol{N}$ are a basis of the nullspace. Hence, $\boldsymbol{AN^T} = 0$.

The dimension of the nullspace (the number of loops) is $n - m + 1$, where 1 is the dimension of the left nullspace (recall, the vector of 1s is the basis for the left nullspace). The dimension of the nullspace follows from the fundamental theorem of linear algebra or, alternatively, it can be viewed as an application of Euler's theorem (see, for example, Strang 1998, 363).[7]

Column space

The column space of $\boldsymbol{A}$ consists of all valid journal entries. Each column of $\boldsymbol{A}$ is itself a simple journal entry. Compound entries also lie in this space, since they can be created as a linear combination of the simple journal entries in the columns.

From the fundamental theorem of linear algebra, the dimension of the column space is the same as the dimension of the row space. The only question is to identify $m - 1$ independent columns of $\boldsymbol{A}$ that form a basis. There are many possible choices, and all of them correspond to a "spanning tree" of the directed graph. Edges (columns) that do not contain any loops are termed "trees", and trees that touch all the nodes are termed "spanning". A spanning tree consists of columns 4 and 5 along with any three columns chosen from 1, 2, 3, 6, 7 such that neither of the two loops (loop 1, 2, 6 or loop 2, 3, 7) is retained. The five columns 1, 2, 3, 5, 6 would not be a basis because the loop (1, 2, 6) is included. Columns 1, 2, 3, 4, 5 give a basis for the column space because the five edges touch all the nodes but do not contain any loops.

The following observation summarizes the above discussion.

OBSERVATION 1. *The composition and dimensions of the four fundamental spaces of the double entry matrix are*

Left nullspace:	*Dimension 1, the balancing vector.*
Column space:	*Dimension* $m - 1$*, the space of valid journal entries.*
Nullspace:	*Dimension* $n - m + 1$*, the space of looping transactions.*
Row space:	*Dimension* $m - 1$*, the space of T accounts.*

Step 3: The invertibility solution

The question addressed in this section deals with the inverse transformation from financial statements back to transactions. In particular, given financial statements, to what extent can we disaggregate to find the primitive transaction data? Equivalently, if two distinct sets of transactions yield identical financial statements, is there a systematic relationship linking the two transaction vectors?

The degree to which we can (or cannot) invert uniquely to $\boldsymbol{y}$ is determined by the dimension of the nullspace. If the dimension of the nullspace is zero (it contains only the zero vector), then we can determine $\boldsymbol{y}$ uniquely. In accounting, the dimension of the nullspace exceeds zero, since the number of transactions typically exceeds the number of accounts. Hence, an infinite number of $\boldsymbol{y}$ vectors can produce the same $\boldsymbol{x}$ vector. However, all such $\boldsymbol{y}$ vectors differ only in the weights they place on the loops of the directed graph (the nullspace vectors).

Given an $\boldsymbol{x}$ vector, the first step is to compute any vector of transactions, $\boldsymbol{y}$, such that $\boldsymbol{Ay} = \boldsymbol{x}$. Call this particular solution $\boldsymbol{y^p}$. All $\boldsymbol{y}$s that produce the same $\boldsymbol{x}$ can be written as the sum of $\boldsymbol{y^p}$ and any linear combination of the loops. Recall that $\boldsymbol{A}$ times $\boldsymbol{y^p}$ is $\boldsymbol{x}$, and $\boldsymbol{A}$ times the loops (nullspace vectors) is 0.

One easy choice for $\boldsymbol{y^p}$ makes use of the spanning tree as follows. Set the transaction amounts for edges not included in the spanning tree equal to 0. For our example, recall that columns 1–5 form a spanning tree. Hence, set $y_6 = y_7 = 0$. Given $\boldsymbol{x} = (-8, 4, 6, -10, 5, 3)$, the first five elements of the $\boldsymbol{y}$ vector, y_1 through y_5, are found by solving the set of equations

$$\begin{bmatrix} -1 & -1 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} -8 \\ 4 \\ 6 \\ -10 \\ 5 \\ 3 \end{bmatrix}.$$

We are assured that a solution exists to the above problem, since $\boldsymbol{x}$ resides in the column space of $\boldsymbol{A}$ and the first five columns of $\boldsymbol{A}$ are a basis for that space. In this case, $\boldsymbol{y^p}$ is (9, 6, 3, 10, 5, 0, 0). Hence, all $\boldsymbol{y}$ vectors consistent with the $\boldsymbol{x}$ vector are of the form

$$\boldsymbol{y} = \begin{bmatrix} 9 \\ 6 \\ 3 \\ 10 \\ 5 \\ 0 \\ 0 \end{bmatrix} + k_1 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \text{ where } k_1 \text{ and } k_2 \text{ are arbitrary constants.}$$

Another common choice for $\boldsymbol{y^p}$ is to pick a y vector that resides entirely in the row space of $\boldsymbol{A}$. This choice is unique and is denoted by $\boldsymbol{y^R}$. $\boldsymbol{y^R}$ can be found by multiplying the accounts $\boldsymbol{x}$ by the pseudoinverse of $\boldsymbol{A}$, denoted $\boldsymbol{A^+}$.[8] For the example, $\boldsymbol{y^R}$ = (7.5, 6, 4.5, 10, 5, 1.5, –1.5). Hence, the invertibility solution can also be written as

$$\boldsymbol{y} = \begin{bmatrix} 7.5 \\ 6 \\ 4.5 \\ 10 \\ 5 \\ 1.5 \\ -1.5 \end{bmatrix} + k_1 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \text{ where } k_1 \text{ and } k_2 \text{ are arbitrary constants.}$$

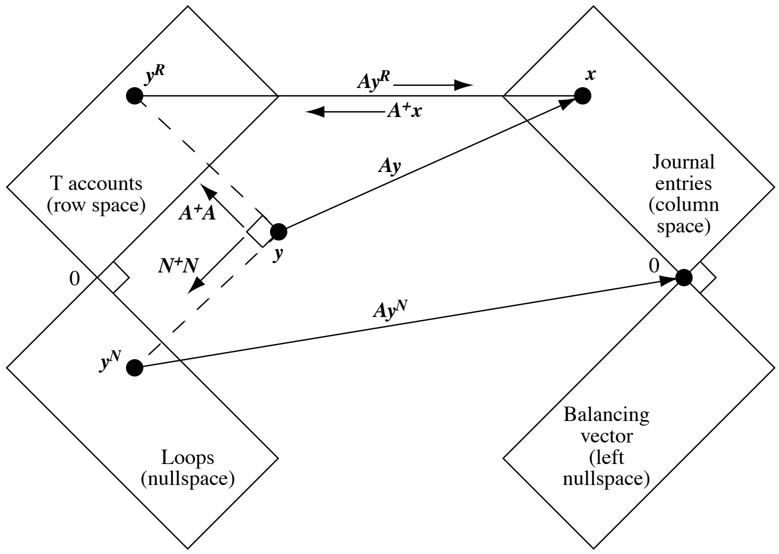
It is easy to verify that $\boldsymbol{y^R}$ resides entirely in the row space of $\boldsymbol{A}$: $\boldsymbol{y^R}$ is orthogonal to the two nullspace vectors. Figure 2 summarizes the interrelationships among the four subspaces, as well as the operation of $\boldsymbol{A}$ and $\boldsymbol{A^+}$. Any $\boldsymbol{y}$ vector can be projected into the row space of $\boldsymbol{A}$ by the compound matrix operation $\boldsymbol{A^+A}$ and into the nullspace of $\boldsymbol{A}$ by $\boldsymbol{N^+N}$. The two projections are denoted by $\boldsymbol{y^R}$ and $\boldsymbol{y^N}$, respectively. Since the sum of $\boldsymbol{y^R}$ and $\boldsymbol{y^N}$ is $\boldsymbol{y}$, $\boldsymbol{N^+N} = \boldsymbol{I} - \boldsymbol{A^+A}$, where $\boldsymbol{I}$ is the $n \times n$ identity matrix.

Observation 2 presents the invertibility solution.

OBSERVATION 2. *The invertibility solution is* $\boldsymbol{A^+x} + \boldsymbol{N^Tk}$, *where* $\boldsymbol{k}$ *is a vector of arbitrary weights.*

Any weighted combination (where the weights sum to 1) of two consistent transaction vectors is itself consistent. The invertibility solution set is convex. This

Figure 2 The four fundamental subspaces of the double entry matrix



is because consistent vectors have the same row component A^+x (the row component is unique for each x) and differ only on the weights they place on the nullspace vectors. A weighted combination of two such vectors is A^+x plus a weighted average of the two nullspace vectors. This weighted combination is consistent with x, since A times the row component yields x and A times the weighted nullspace component yields zero.

Recall that the invertibility solution can generally be written as $y^p + N^T k$. The choice of $y^p = A^+x$ is made to ensure that the two components of y that make up the invertibility solution are orthogonal; the first piece lies in the row space and the second in the nullspace of A. The orthogonal component representation proves convenient when the reader of financial statements uses the information to update her or his beliefs regarding the transactions vector.[9]

Step 4: Updating priors

The final step is to derive the conditional mean of the consistent y vectors identified in the previous step. As argued earlier, for a Bayesian reader with a quadratic loss function, the best guess is simply the conditional mean. Proposition 1 presents the reader's best guess of the transactions after she or he has had a chance to examine the financial statements.

PROPOSITION 1. *Given financial statements x, the reader's best guess of the unknown transactions is $A^+x + N^+N\bar{y}$.*

PROOF OF PROPOSITION 1. We first check to see that the best guess is consistent with x — that is, $A(A^+x + N^+N\bar{y}) = x$. The transformation AA^+ projects x into the column space of $\boldsymbol{A}$. But since $\boldsymbol{x}$ already resides in the column space of $\boldsymbol{A}$, $\boldsymbol{AA^+x} = \boldsymbol{x}$. Also, $\boldsymbol{AN^+N} = 0$. To see this, substitute $\boldsymbol{I} - \boldsymbol{A^+A}$ for $\boldsymbol{N^+N}$ and note that $\boldsymbol{AA^+A} = \boldsymbol{A}$. Hence, $\boldsymbol{A}(\boldsymbol{A^+x} + \boldsymbol{N^+N}\bar{\boldsymbol{y}}) = \boldsymbol{x} + 0$ — that is, the best guess is consistent.

We next check that from among all consistent $\boldsymbol{y}$ vectors, $\boldsymbol{A^+x} + \boldsymbol{N^+N}\bar{\boldsymbol{y}}$ has the least expected loss. From Observation 2 and the use of the quadratic loss function, $\boldsymbol{A^+x} + \boldsymbol{N^+N}\bar{\boldsymbol{y}}$ is the minimum expected loss estimator if $\mathrm{E}(\tilde{\boldsymbol{y}}^N | \tilde{\boldsymbol{y}}^R = \boldsymbol{A^+x}) = \boldsymbol{N^+N}\bar{\boldsymbol{y}}$. Since $\tilde{\boldsymbol{y}}^R = \boldsymbol{A^+A}\tilde{\boldsymbol{y}}$, $\tilde{\boldsymbol{y}}^N = \boldsymbol{N^+N}\tilde{\boldsymbol{y}}$, and $\tilde{\boldsymbol{y}} \sim N(\bar{\boldsymbol{y}}, \boldsymbol{I})$, it follows that $\tilde{\boldsymbol{y}}^R \sim N(\boldsymbol{A^+A}\bar{\boldsymbol{y}}, \boldsymbol{A^+A})$ and $\tilde{\boldsymbol{y}}^N \sim N(\boldsymbol{N^+N}\bar{\boldsymbol{y}}, \boldsymbol{N^+N})$. Since the two components are normally distributed and $\mathrm{Cov}(\tilde{\boldsymbol{y}}^R, \tilde{\boldsymbol{y}}^N) = \mathrm{Cov}(\boldsymbol{A^+A}\tilde{\boldsymbol{y}}, \boldsymbol{N^+N}\tilde{\boldsymbol{y}}) = \boldsymbol{A^+AIN^+N} = 0$, the row space and the nullspace components are statistically independent. Hence, $\mathrm{E}(\tilde{\boldsymbol{y}}^N | \tilde{\boldsymbol{y}}^R = \boldsymbol{A^+x}) = \mathrm{E}(\tilde{\boldsymbol{y}}^N) = \boldsymbol{N^+N}\bar{\boldsymbol{y}}$. ■

The intuition for the result is straightforward. The unconditional mean is composed of two orthogonal parts, $\boldsymbol{A^+A}\bar{\boldsymbol{y}}$ and $\boldsymbol{N^+N}\bar{\boldsymbol{y}}$. Since all information in the nullspace is lost during the aggregation process of preparing financial statements and the nullspace component is independent of the row space component, the observability of $\boldsymbol{x}$ does not lead to any updating of the reader's beliefs regarding the nullspace component; the best guess of the nullspace component of the transaction vector continues to be $\boldsymbol{N^+N}\bar{\boldsymbol{y}}$. However, the observability of $\boldsymbol{x}$ does result in belief revision of the row component of the transaction vector. This updating is crisp. Every $\boldsymbol{x}$ vector is associated with a unique counterpart $\boldsymbol{A^+x}$ in the row space. That is, $\boldsymbol{A^+x}$ is the only transaction vector residing in the row space that is consistent with the financial statements. Hence, the reader's best guess of the underlying transaction vector is the sum of $\boldsymbol{A^+x}$ and $\boldsymbol{N^+N}\bar{\boldsymbol{y}}$.[10]

Return to the ongoing example with $\boldsymbol{x} = (-8, 4, 6, -10, 5, 3)$ and $\bar{\boldsymbol{y}} = (7, 9, 1, 10, 5, 1, 2)$. On observing $\boldsymbol{x}$, the reader's best guess is

$$\boldsymbol{A^+x} + \boldsymbol{N^+N}\bar{\boldsymbol{y}} = \begin{bmatrix} 7.5 \\ 6 \\ 4.5 \\ 10 \\ 5 \\ 1.5 \\ -1.5 \end{bmatrix} + \begin{bmatrix} 0.125 \\ 3.25 \\ -3.375 \\ 0 \\ 0 \\ -0.125 \\ 3.375 \end{bmatrix} = \begin{bmatrix} 7.625 \\ 9.25 \\ 1.125 \\ 10 \\ 5 \\ 1.375 \\ 1.875 \end{bmatrix}.$$

We next turn to interval estimation. Corollary 1 presents the region in which the transaction vector resides with probability $1 - \alpha$.

COROLLARY 1. *Given financial statements* $\boldsymbol{x}$*, the reader assigns probability* $1 - \alpha$ *that the transaction vector resides in the region* $\boldsymbol{R}$, $\boldsymbol{R} = \{\boldsymbol{A^+x} + \boldsymbol{N^Tk}: [\boldsymbol{k} - (\boldsymbol{NN^T})^+\boldsymbol{N}\bar{\boldsymbol{y}}]^T (\boldsymbol{NN^T}) [\boldsymbol{k} - (\boldsymbol{NN^T})^+\boldsymbol{N}\bar{\boldsymbol{y}} \leq \chi^2(\alpha)\}$*, where* $\chi^2(\alpha)$ *denotes the upper* (100α)*th percentile of the chi square distribution with degrees of freedom* $n - m + 1$.

PROOF OF COROLLARY 1. As shown in Proposition 1, $\tilde{\boldsymbol{y}}^N \sim N(\boldsymbol{N}^+\boldsymbol{N}\bar{\boldsymbol{y}}, \boldsymbol{N}^+\boldsymbol{N})$. $\tilde{\boldsymbol{y}}^N$ can be written as $\boldsymbol{N}^T\tilde{\boldsymbol{k}}$. Hence, $\boldsymbol{N}\boldsymbol{N}^T\tilde{\boldsymbol{k}} = \boldsymbol{N}\tilde{\boldsymbol{y}}^N$ and $\tilde{\boldsymbol{k}} = (\boldsymbol{N}\boldsymbol{N}^T)^+\boldsymbol{N}\tilde{\boldsymbol{y}}^N$. This implies that $\boldsymbol{k}$ is normally distributed with mean $(\boldsymbol{N}\boldsymbol{N}^T)^+\boldsymbol{N}\boldsymbol{N}^+\boldsymbol{N}\bar{\boldsymbol{y}} = (\boldsymbol{N}\boldsymbol{N}^T)^+\boldsymbol{N}\bar{\boldsymbol{y}}$ and variance–covariance matrix $(\boldsymbol{N}\boldsymbol{N}^T)^+\boldsymbol{N}\boldsymbol{N}^T(\boldsymbol{N}\boldsymbol{N}^T)^+ = (\boldsymbol{N}\boldsymbol{N}^T)^+$. (This uses the pseudoinverse property that $\boldsymbol{N}\boldsymbol{N}^+\boldsymbol{N} = \boldsymbol{N}$.) The corollary then follows from the fact that the sum of $n - m + 1$ squared standard normal variables follows a chi square distribution with $n - m + 1$ degrees of freedom. ■

Not surprisingly, the posterior distribution of $\boldsymbol{y}$ has a smaller variance than the prior distribution. Since $(\boldsymbol{N}\boldsymbol{N}^T)^+$ is the variance–covariance matrix associated with $\tilde{\boldsymbol{k}}$, the variance–covariance matrix of $\boldsymbol{A}^+\boldsymbol{x} + \boldsymbol{N}^T\tilde{\boldsymbol{k}}$ is $\boldsymbol{N}^+\boldsymbol{N} \leq \boldsymbol{I}$.[11] Also, as $\boldsymbol{x}$ changes, the conditional mean changes but the conditional variance does not. The conditional variance changes only if information in addition to what is contained in $\boldsymbol{x}$ is provided. Of course, the new information cannot just be a linear combination of elements of $\boldsymbol{x}$. In the language of linear algebra, information is "new" only if it results in the reduction of the nullspace. When that happens, the conditional variance further declines.

With additional information, belief revision occurs in the same manner as in Proposition 1 (and Corollary 1). Return to the example. Say, in addition to the information in $\boldsymbol{x}$, the reader also learns that the total depreciation charge is 4. Total depreciation corresponds to $y_6 + y_7$. Augment the $\boldsymbol{A}$ matrix (the $\boldsymbol{x}$ vector) with an additional row (element) that reflects this information. In particular, [0 0 0 0 0 1 1] is the additional row in the $\boldsymbol{A}$ matrix, and 4 is the additional element in the $\boldsymbol{x}$ vector.

The additional row is independent of the other rows in $\boldsymbol{A}$. This follows from the fact that adding the row reduces the dimension of the nullspace from two to one. A quick way to see this is to use some accounting and examine the impact of the additional information on the directed graph. The Plant T account reveals that if total depreciation is 4, plant acquisition (y_2) must be 10. Reconsider the loops in Figure 1. Since y_2 is known only one loop remains, and it consists of transactions 1, 3, 6, and 7. The nullspace vector is (1, 0, –1, 0, 0, –1, 1).

Using the augmented $\boldsymbol{A}$ matrix, the augmented $\boldsymbol{x}$ vector, and the lone nullspace vector for $\boldsymbol{N}$ results in the best estimate of (7.25, 10, 0.75, 10, 5, 1.75, 2.25).

4. Extension: A classification problem

In this section we embed the accounting conditional estimation problem in an economic decision context. The decision problem we consider is a classification exercise. In our setting, a reader attempts to categorize a firm as being one of two "types" based on the observed financial statements. The two types differ in the mean transactions amounts; the financial statements, of course, capture only the row space components of the transactions. The decision maker uses the financial statements to update her or his priors of the different types. The updating exercise is somewhat delicate because the accounting process introduces a dependency in the accounts even when there is no dependency in the underlying transactions. This is because any time an account is debited, another account is credited. Double entry introduces a negative covariance.

We rank alternative accounting platforms based on an ex ante efficiency criterion of minimizing the total cost of misclassification.[12] The classification problem is characterized by $\bar{y}_d$, the vector of differences in the mean transaction amounts of the two types. The ranking criterion is intuitive. Platform $\boldsymbol{A}$ is preferred to platform $\boldsymbol{B}$ if regressing $\bar{y}_d$ in the rows of $\boldsymbol{A}$ yields a higher R^2 than when $\bar{y}_d$ is regressed in the rows of $\boldsymbol{B}$.

This is analogous to what we did in the previous section. The vector $\bar{y}_d$ can be decomposed into the row space and the nullspace of the aggregation matrix. The former component is retained, while the latter is lost when $\boldsymbol{y}$ is aggregated. The R^2 of our regression measures the proportion of $\bar{y}_d$ that lies in the rows of the aggregation matrix. A higher R^2 means that more of the $\bar{y}_d$ vector resides in the row space and, hence, less of the relevant transaction information is lost in the aggregation process. We note that even if one platform has a row space that is a subset of the row space of another platform, there may be no increase in the expected cost of misclassification if the former platform is used. For example, if $\bar{y}_d$ resides entirely in the smaller row space, then R^2 is 1 irrespective of which of the two platforms is used — the discriminant power under the two platforms is the same and is equal to the discriminant power when transaction information is available.

The above result is contextual — that is, platform choice is made for a particular $\bar{y}_d$ vector. A corollary to this result is that platform $\boldsymbol{A}$ is preferred to platform $\boldsymbol{B}$ for all $\bar{y}_d$ vectors if, and only if, the nullspace of $\boldsymbol{A}$ is a subset of the nullspace of $\boldsymbol{B}$. The nested nullspace condition implies that any information in $\boldsymbol{By}$ can be reproduced by a linear combination of the elements of the vector $\boldsymbol{Ay}$.

The classification problem

A reader of financial statements is interested in discriminating between two types of firms, L and H. Each firm undertakes n transactions. The transaction amount vector for the i-type firm is denoted by $\boldsymbol{y}_i$, $i = L,H$.

Each type of firm is characterized by the mean value of its transactions vector, denoted by $\bar{y}_i$. The difference in the two mean transaction vectors, $\bar{y}_H - \bar{y}_L$, is defined as $\bar{y}_d$. The actual transactions vector is a linear function of the mean transactions vector, $\bar{y}_i$, and a random state of nature $\tilde{e}_i$: $\tilde{y}_i = \bar{y}_i + \tilde{e}_i$. We assume $\tilde{e}_i$ is normally distributed with mean zero and identity variance–covariance matrix.[13]

The reader's priors are that the firm is equally likely to be of L- or of H-type. The decision maker's objective is to use the firm's financial statements, $\boldsymbol{Ay}$, to classify the firm as an L- or H-type such that the ex ante (before $\boldsymbol{x}$ is observed) probability of misclassification is minimized.[14]

The solution to the classification problem

Revisiting the ongoing numerical example, suppose the two equiprobable types are characterized by the following mean transaction vectors: $\bar{y}_L = (8, 9, 1, 10, 5, 1, 2)$ and $\bar{y}_H = (7, 9, 1, 10, 5, 1, 2)$. Suppose a firm's observed financial statement is $\boldsymbol{x} = (-8, 4, 6, -10, 5, 3)$. The question is what is the reader's best guess of the firm's type?

Notice that the types differ, on average, only in transaction 1, the acquisition of inventory. However, as we will shortly see, information about other transactions (i.e., depreciation) will prove to be helpful in discriminating between the two types. The usefulness of other information is due to the covariance introduced by the double entry system. This is stated in the next observation.

OBSERVATION 3. *Even when transactions are independent, a dependency in accounts is introduced because of the double entry process itself.*

Since $\boldsymbol{x} = \boldsymbol{Ay}$, the variance–covariance matrix associated with $\boldsymbol{x}$ is $\boldsymbol{AA^T}$. Further, when $\boldsymbol{A}$ is an incidence matrix (i.e., composed only of journal entries), $\boldsymbol{AA^T}$ takes on an intuitive form. The diagonal elements of the matrix, say a_{ii}, are equal to the number of journal entries affecting account i. The off-diagonal elements, say a_{ij}, are equal to -1 if accounts i and j are connected by a journal entry, and zero otherwise.

In making her or his classification decisions, the reader has to be cognizant of the induced variance–covariance structure. For example, if the variance associated with account i is high, the reader will place less weight on it. Also, an account that on the average is the same for both types may still be valuable in the classification exercise because of the possibility of "learning". That is, negative covariance allows information regarding account i to update the reader's beliefs regarding account j.

Given $\boldsymbol{x}$, denote the (posterior) probability that a firm is of type i by $\Pr(i|\boldsymbol{x})$. Using Bayes's rule,

$$\Pr(i \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x}|L)}{f(\boldsymbol{x}|L) + (\boldsymbol{x}|H)}$$

where $f(\boldsymbol{x}|i) = k\ \mathrm{Exp}[-0.5(\boldsymbol{x} - \boldsymbol{A}\,\bar{\boldsymbol{y}}_i)^T(\boldsymbol{AA^T})^+(\boldsymbol{x} - \boldsymbol{A}\,\bar{\boldsymbol{y}}_i)]$. The function $f(\boldsymbol{x}|i)$ is the probability density function of a multivariate normal distribution, and k is the normalizing constant.[15]

The optimal classification rule and the total probability of misclassification are presented in the following lemma:

LEMMA. (a) *Given* $\boldsymbol{x}$, *a firm is classified as an L-type if* $\Pr(L|\boldsymbol{x}) > \Pr(H|\boldsymbol{x})$; *else it is classified as an H-type.*

(b) *Under the above classification rule, the (ex ante) probability of misclassification is* $F(-\Delta_A/2)$, *where* $\Delta_A^2 = (\boldsymbol{A}\bar{\boldsymbol{y}}_d)^T(\boldsymbol{AA^T})^+(\boldsymbol{A}\,\bar{\boldsymbol{y}}_d)$ *and* F *is the cumulative distribution function of a standard normal distribution.*

The classification rule in part (a) simply compares posterior probabilities. The proof of part (b) can be found in any standard discussion of discriminant analysis (e.g., Johnson and Wichern 1988). The sketch of the proof is as follows. Comparing $\Pr(L|\boldsymbol{x})$ with $\Pr(H|\boldsymbol{x})$ is the same as comparing $f(\boldsymbol{x}|L)$ with $f(\boldsymbol{x}|H)$. Substituting the probability density function of the multivariate normal distribution for f, taking natural logarithms, and simplifying implies that $f(\boldsymbol{x}|L) > f(\boldsymbol{x}|H)$ iff $\boldsymbol{\ell^T x} < \boldsymbol{\ell^T m}$, where $\boldsymbol{\ell} = (\boldsymbol{AA^T})^+(\boldsymbol{A}\,\bar{\boldsymbol{y}}_d)$ and $\boldsymbol{m} = (\boldsymbol{A}\,\bar{\boldsymbol{y}}_L + \boldsymbol{A}\,\bar{\boldsymbol{y}}_H)/2$. That is, a classification rule equivalent to that presented in Lemma (a) is to convert the vector $\boldsymbol{x}$ into a scalar using ℓ and then compare it with the midpoint, $\boldsymbol{\ell^T m}$, of the scalar distribution.[16] The probability of misclassification is $\Pr(\boldsymbol{\ell^T x} < \boldsymbol{\ell^T m} | H)0.5 + \Pr(\boldsymbol{\ell^T x} > \boldsymbol{\ell^T m} | L)0.5$; the first (second) number is the probability that an H-type (L-type) firm is classified as an L-type (H-type) weighted by the probability that the firm is of H-type (L-type). Since $\boldsymbol{\ell^T x}$ is normally distributed, these probabilities can easily be computed and shown to be equal to $F(-\Delta_A/2)$.

Using Lemma (a) for our example implies that the firm should be classified as an L-type, since $\Pr(L|\boldsymbol{x}) = 0.5775$ and $\Pr(H|\boldsymbol{x}) = 0.4225$. To see why the reader needs to be cognizant of the correlation induced by double entry, suppose that, in addition to $\boldsymbol{x}$, the reader learns that depreciation ($y_6 + y_7$) is equal to 4. Since, on average, the two types differ only in y_1, the reader may conclude that this additional information will not affect the classification exercise. Because of the induced correlation effect, this conclusion is incorrect. Using the augmented $\boldsymbol{A}$ matrix (the original $\boldsymbol{A}$ with the additional row [0 0 0 0 0 1 1]) and the augmented $\boldsymbol{x}$ vector (−8, 4, 6, −10, 5, 3, 4) yields $\Pr(L|\boldsymbol{x}) = 0.4688$. Based on $\boldsymbol{x}$ and the depreciation information, the firm is now classified as an H-type.

The link between the classification problem and the nullspace thinking presented earlier is highlighted by the following exercise. For the example, compute the R^2 associated with regressing $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ on the rows of $\boldsymbol{A}$ and the augmented $\boldsymbol{A}$ matrix. The R^2 are 5/8 and 3/4, respectively. Also, using Lemma (b), compute the Δ^2 for the two cases.[17] The Δ^2 are the same as the R^2. This connection between the regression problem and the classification problem is not a coincidence.

In the regression problem, the total sum of squares is $\|\bar{\boldsymbol{y}}_{\boldsymbol{d}}\|^2$. As is the case with any n-length vector, $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ can be decomposed into the row space and the nullspace of $\boldsymbol{A}$. The squared length of these projections is the sum of squares regression and the sum of squares error, respectively. Hence, $R^2 = (\|\boldsymbol{A^+A}\,\bar{\boldsymbol{y}}_{\boldsymbol{d}}\|^2)/(\|\bar{\boldsymbol{y}}_{\boldsymbol{d}}\|^2)$.[18]

In the aggregation process we know that only the row space component is preserved. Intuitively, this suggests that $\|\boldsymbol{A^+A}\,\bar{\boldsymbol{y}}_{\boldsymbol{d}}\|^2$ should be related to the discriminant power when financial statements are used. As we show next, $\Delta_A^2 = \|\boldsymbol{A^+A}\,\bar{\boldsymbol{y}}_{\boldsymbol{d}}\|^2$. Thus, comparing two aggregation platforms, say $\boldsymbol{A}$ and $\boldsymbol{B}$, is simply an exercise in computing and comparing R^2s.

PROPOSITION 2. (a) *Platform **A** is (weakly) preferred to platform **B** for a particular $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ if and only if the R^2 associated with regressing $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ in the rows of **A** is greater than the R^2 associated with regressing $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ in the rows of **B**.*

(b) *Platform **A** is preferred to platform **B** for all $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ if and only if the nullspace of **A** is a subspace of the nullspace of **B**.*

Proof of Proposition 2. (a) From Lemma (b), the probability of misclassification is $F(-\Delta/2)$. For a given $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ vector, $\boldsymbol{A}$ is preferred to $\boldsymbol{B}$ iff $F(-\Delta_A/2) \leq F(-\Delta_B/2)$. Since F is an increasing function (it is a cumulative distribution function), this implies that $\boldsymbol{A}$ is preferred to $\boldsymbol{B}$ iff $\Delta_A^2 \geq \Delta_B^2$. Δ_A^2 (Δ_B^2) can be written as the length of the projection of $\bar{\boldsymbol{y}}_{\boldsymbol{d}}$ in the rows of $\boldsymbol{A}$ ($\boldsymbol{B}$):

$$\Delta_A^2 = (A\,\bar{y}_d)^T(AA^T)^+(A\,\bar{y}_d) = \bar{y}_d{}^TA^T(AA^T)^+A\,\bar{y}_d = \bar{y}_d{}^TA^T(A^T)^+A^+A\,\bar{y}_d$$

$$= \bar{y}_d{}^T(A^+A)^TA^+A\,\bar{y}_d = \bar{y}_d{}^TA^+A\,\bar{y}_d = \|A^+A\,\bar{y}_d\|^2.$$

The second-to-last equality is true because A^+A is a projection matrix; projection matrices are symmetric and idempotent. Since $R^2 = (\|A^+A\,\bar{y}_d\|^2/\|\bar{y}_d\|)^2$, the result follows.

(b) If the nullspace of $\boldsymbol{A}$ is a subset of the nullspace of $\boldsymbol{B}$, then the row space of $\boldsymbol{B}$ is a subset of the row space of $\boldsymbol{A}$. This implies that the length of the projection of any vector into the row space of $\boldsymbol{A}$ is at least as large as its projection into the row space of $\boldsymbol{B}$. Also, if the nullspace of $\boldsymbol{A}$ is not a subset of $\boldsymbol{B}$, then it is always possible to find a vector whose projection in the row space of $\boldsymbol{A}$ is smaller than its projection in the row space of $\boldsymbol{B}$. The result then follows from the fact that the length of the projection of $\bar{y}_d$ into the rows of the aggregating matrix is inversely proportional to the probability of misclassification. ■

Proposition 2(b) is a statement of comparison in the sense of Blackwell 1951: platform $\boldsymbol{A}$ is preferred to platform $\boldsymbol{B}$ for all decisions (not just the discriminant problem). The information produced by the transformation $\boldsymbol{By} = \boldsymbol{x_B}$ can be created by a linear combination of the information in $\boldsymbol{Ay} = \boldsymbol{x_A}$, but not vice versa. That is, $\boldsymbol{x_B} = \boldsymbol{Cx_A}$, where $\boldsymbol{C}$ has a non-empty nullspace. (For the case where $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same dimensions, a non-empty nullspace of $\boldsymbol{C}$ implies that $\boldsymbol{C}$ is not invertible.)

Showing $\boldsymbol{x_B} = \boldsymbol{Cx_A}$ is equivalent to showing $\boldsymbol{By} = \boldsymbol{C}(\boldsymbol{Ay})$ for all $\boldsymbol{y}$. That is, $\boldsymbol{B} = \boldsymbol{CA}$.[19] Choose $\boldsymbol{C} = \boldsymbol{BA}^+$. This choice (1) satisfies $\boldsymbol{B} = \boldsymbol{CA}$ and (2) has a non-empty nullspace. To see (1), note that $\boldsymbol{B} - \boldsymbol{CA} = \boldsymbol{B} - \boldsymbol{BA}^+\boldsymbol{A} = \boldsymbol{B}(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A}) = 0$. The last equality follows because $\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A}$ projects any vector into the nullspace of $\boldsymbol{A}$, and the nullspace of $\boldsymbol{A}$ is contained in the nullspace of $\boldsymbol{B}$ (from Proposition 2(b)). Since $\boldsymbol{A}$ is an incidence matrix, the vector of all 1s lies in the left nullspace of $\boldsymbol{A}$. The left nullspace of $\boldsymbol{A}$ is the nullspace of $\boldsymbol{A}^+$. Hence, $\boldsymbol{BA}^+$ times the vector of ones is 0. $\boldsymbol{C}$ has a non-empty nullspace. This proves (2).

5. Conclusion

We characterize the double entrydouble entry aggregation process as an incidence matrix. An incidence matrix can be represented by a directed graph, which allows a convenient and simple derivation of the fundamental subspaces of the matrix. The fundamental subspaces, in turn, lead to a closed form solution to the following inference problem: Presented with a set of financial statements (and priors), what is the best guess of the underlying transactions that generated the statements?

In practice, a major strength of accounting is its dynamic nature. While it may be possible to manage earnings in the short run, it is much harder to do so in the long run. Sooner or later, accounting catches up. A limitation of this paper is that we studied only static attributes of the accounting system. We conjecture that the ability of linear algebra to model dynamic systems (see, e.g., Strang 1986) may prove beneficial in the study of accounting. Perhaps even the language developed in linear algebra to describe dynamic systems (e.g., eigenvalues and eigenvectors) may have accounting counterparts.

Another possible extension is to incorporate benefits into aggregation. A commonly cited benefit to aggregation is motivated by bounded rationality: limits on information transmission, reception, and processing can make aggregated information desirable. A second potential benefit to aggregation arises when individual items are measured with errors. Aggregation may allow errors in the individual items to cancel out (Datar and Gupta 1994). A third reason is that the aggregation process itself may add information (Sunder 1997). Another reason a firm may wish to disclose aggregated information is if disaggregated information includes proprietary information that may be exploited by competitors (Newman and Sansing 1993). Finally, agency models have studied settings in which the principal's limited ability to commit has led to coarse (aggregated) information being optimal (see, e.g., Cremer 1995 and Demski and Frimor 2000). These models build on the familiar game-theoretic idea that there are games in which a player may gain by limiting his or her own information if the opponents know he or she has done so, because this may induce the opponents to play in a desirable fashion. An interesting question arises: Could accounting structure, and perhaps even double entry, be an optimal way of aggregating information?

Appendix

In this appendix, we discuss the construction and properties of the pseudoinverse of a matrix. Consider any matrix $\boldsymbol{A}$ of rank r. $\boldsymbol{A}$ can be written as the product of two matrices, $\boldsymbol{F}$ and $\boldsymbol{G}$, where $\boldsymbol{F}$ has r independent columns and $\boldsymbol{G}$ has r independent rows. The matrices $\boldsymbol{F}^T\boldsymbol{F}$ and $\boldsymbol{G}\boldsymbol{G}^T$ are $r \times r$, symmetric, and nonsingular. Nonsingular square matrices have regular inverses. Hence, $(\boldsymbol{F}^T\boldsymbol{F})^{-1}(\boldsymbol{F}^T\boldsymbol{F}) = (\boldsymbol{G}\boldsymbol{G}^T)(\boldsymbol{G}\boldsymbol{G}^T)^{-1} = \boldsymbol{I}$.

Define $\boldsymbol{F}^+ = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T$ and let $\boldsymbol{G}^+ = \boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T)^{-1}$. Note that $\boldsymbol{F}^+$ is an inverse of $\boldsymbol{F}$ in the sense that $\boldsymbol{F}^+\boldsymbol{F} = \boldsymbol{I}$ and, similarly, $\boldsymbol{G}\boldsymbol{G}^+ = \boldsymbol{I}$. Define $\boldsymbol{A}^+ = \boldsymbol{G}^+\boldsymbol{F}^+$. The pseudoinverse of $\boldsymbol{A}$ is $\boldsymbol{A}^+$ and has the following properties:

$$\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$$

$$\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+$$

$$\boldsymbol{A}\boldsymbol{A}^+ = (\boldsymbol{A}\boldsymbol{A}^+)^T$$

$$\boldsymbol{A}^+\boldsymbol{A} = (\boldsymbol{A}^+\boldsymbol{A})^T$$

A matrix satisfying the above properties is also called a generalized inverse or a Moore-Penrose inverse. It always exists and is unique (see Penrose 1955).

For the $\boldsymbol{A}$ matrix in our running example, $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{A}^+$ are presented below.

$$
\boldsymbol{F} = \begin{bmatrix} -1 & -1 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}
$$

$$
\boldsymbol{A}^{+} = \boldsymbol{G}^{\boldsymbol{T}} (\boldsymbol{G}\boldsymbol{G}^{\boldsymbol{T}})^{-1} (\boldsymbol{F}^{\boldsymbol{T}}\boldsymbol{F})^{-1} \boldsymbol{F}^{\boldsymbol{T}} = \frac{1}{48} \begin{bmatrix} -13 & 17 & -1 & -13 & 17 & -7 \\ -10 & 2 & 14 & -10 & 2 & 2 \\ -9 & -3 & 3 & -9 & -3 & 21 \\ 8 & 8 & 8 & -40 & 8 & 8 \\ -8 & -8 & -8 & -8 & 40 & -8 \\ -3 & 15 & -15 & -3 & 15 & -9 \\ 1 & -5 & -11 & 1 & -5 & 19 \end{bmatrix}.
$$

Endnotes

1. See, for example, Demski 1994 and Ijiri 1975 for discussions of these linear representations.
2. We thank a referee for suggesting that the formal representation of the accounting system could be used to answer the belief revision problem.
3. Other authors have expressed different views. For example, McCarthy (1982, 559) writes, “It is a primary contention of this paper that the semantic modeling of accounting object systems should not include elements of double entry bookkeeping such as debits, credits, and accounts.”
4. While our result is stated for the identity variance–covariance case, it can be easily extended to the case of a nonidentity variance–covariance matrix. The key is to decompose any variance–covariance matrix into a symmetric factorization $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\boldsymbol{T}}$. The variable $\boldsymbol{\Gamma}^{-1}\boldsymbol{y}$ is associated with identity variance–covariance and can be used instead of $\boldsymbol{y}$ in the paper. This is the same transformation of variable technique that is used to convert a generalized least squares problem into an ordinary least squares problem. See Greene 1997 (507). The transformation is particularly simple for the special case where the dependency in transactions is only through sales (as is assumed when the percentage-of-sales forecasting method is adopted). In particular, say the relationship between transaction i and sales is of the form $(f_i + ei)$ sales, where f_i is a proportionality constant and the error terms (the e_is) are independently distributed. In this case, dividing each transaction by sales creates items that are independently distributed.
5. Denote the quadratic loss function by $(\boldsymbol{w} - \boldsymbol{b})^{\boldsymbol{T}}\boldsymbol{C}(\boldsymbol{w} - \boldsymbol{b})$, where $\boldsymbol{w}$ is an n-length random variable, $\boldsymbol{C}$ is an $n \times n$ matrix of constants, and T denotes a transpose. The expected value of this loss function is minimized when $\boldsymbol{b} = \mathrm{E}(\boldsymbol{w})$. See, for example, Greene 1997 (316).

6. Ending cash balance = beginning cash balance – purchase inventory – plant acquisition – cash expenses + cash sales = 10 – 7 – 9 – 1 + 10 = 3.
7. Euler's theorem can be stated as follows. In a connected graph, the number of loops equals $n - m + 1$, where n is the number of edges and m is the number of nodes.
8. Readers unfamiliar with pseudoinverses can read the appendix for a derivation of $\boldsymbol{A}^+$ and its properties. Also, see Ijiri 1965, 1996.
9. Furthermore, since the pseudoinverse is a commonly found function in standard mathematical packages (e.g., Mathematica and MATLAB), this characterization is particularly easy to operationalize.
10. We were unable to compute the reader's best guess of transactions using only traditional T accounts. The idea of breaking transactions into a minimum (length) component and a component that leaves the financial statements unchanged is not only an elegant way to represent transactions but also an idea that we found extremely useful in solving an accounting problem.
11. $\boldsymbol{A} \leq \boldsymbol{B}$ if $\boldsymbol{B} - \boldsymbol{A}$ is a positive semidefinite matrix.
12. Arya, Fellingham, and Schroeder (2000) compare the cost of using account balances rather than transaction information in the classification exercise.
13. The reasons for this choice are twofold. First, by eliminating any interdependency in transactions we are able to focus on the interdependencies introduced by the double entry matrix. Second, solving the problem when the variance–covariance matrix is not I is straightforward (see note 4).
14. Minimizing the probability of misclassification is equivalent to the reader minimizing the expected cost of misclassification, assuming that the costs of misclassifying an L-type as an H-type firm and an H-type as an L-type firm are equal. The case of unequal costs of misclassification and unequal priors can be handled easily but provides no additional intuition.
15. If $\boldsymbol{AA^T}$ is nonsingular, the normalizing constant is $(2\pi)^{-n/2}|\boldsymbol{AA^T}|^{-1/2}$. Of course, since k appears both in the numerator and the denominator of the expression for $\Pr(i|\boldsymbol{x})$, $\Pr(i|\boldsymbol{x})$ can be calculated without computing k.
16. The choice of $\boldsymbol{\ell} = (\boldsymbol{AA^T})^+(\boldsymbol{A}\,\bar{\boldsymbol{y}}_{\boldsymbol{d}})$ is intuitive. It maximizes the squared distance between the means of the scalar distributions, standardized by the variance. That is, $\boldsymbol{\ell}$ is chosen so as to maximize

    $$[\boldsymbol{\ell}^T(\boldsymbol{A}\,\bar{\boldsymbol{y}}_{\boldsymbol{H}} - \boldsymbol{A}\,\bar{\boldsymbol{y}}_{\boldsymbol{L}})]^2 / \boldsymbol{\ell}^T(\boldsymbol{AA^T})\boldsymbol{\ell}.$$

17. Not surprisingly, the ex ante probability of misclassification with depreciation information, F(–3/8) = 0.333, is less than when depreciation information is unavailable, F(–5/16) = 0.346.
18. R is the linear aggregation coefficient in Ijiri 1968.
19. This is the familiar test for comparing experiments (information systems) in Blackwell 1951.

References

Arya, A., J. Fellingham, and D. Schroeder. 2000. Accounting information, aggregation and discriminant analysis. *Management Science* 46 (6): 790–806.

Blackwell, D. 1951. Comparison of experiments. *Proceedings of second Berkeley symposium on mathematical statistics and probability*, 93–102. Berkeley, CA: University of California Press.

Butterworth, J. E. 1972. The accounting system as an information function. *Journal of Accounting Research* 10 (Spring): 1–27.

Cremer, J. 1995. Arm's length relationships. *Quarterly Journal of Economics* 110 (2): 275–95.

Datar, S., and M. Gupta. 1994. Aggregation, specification and measurement errors in product costing. *Accounting Review* 69 (4): 567–91.

Demski, J. 1992. Accounting theory. Working paper, Yale University.

———. 1994. Managerial uses of accounting information. Boston: Kluwer Academic.

Demski, J., and H. Frimor. 2000. Performance measure garbling under renegotiation in multi-period agencies. *Journal of Accounting Research* 37 (Supplement): 187–214.

Greene, W. 1997. *Econometric analysis*. Englewood Cliffs, NJ: Prentice Hall.

Ijiri, Y. 1965. On the generalized inverse of an incidence matrix. *Journal of Society for Industrial and Applied Mathematics* 13 (3): 827–36.

———. 1967. *The foundations of accounting measurement: A mathematical, economic, and behavioral inquiry*. Englewood Cliffs, NJ: Prentice-Hall.

———. 1968. The linear aggregation coefficient as the dual of the linear correlation coefficient. *Econometrica* 36 (2): 252–59.

———. 1971. Fundamental queries in aggregation theory. *Journal of the American Statistical Association* 66 (December): 766–82.

———. 1975. *Theory of accounting measurement*. Studies in Accounting Research No. 10. Sarasota, FL: American Accounting Association.

———. 1998. On the generalized inverse of network matrices and its applications to accounting networks. In *Operations research: Methods, models and applications: Proceedings of a conference in honor of Gerald L. Thompson on the occasion of his 70th birthday*, eds. S. Zionts and J. Aronson, 241–58. Westport, CT: Quorum Books.

Johnson, R., and D. Wichern. 1988. *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.

Mattessich, R. 1964. *Accounting and analytical methods*. Homewood, IL: Richard D. Irwin.

McCarthy, W. 1982. The REA accounting model: A generalized framework for accounting systems in a shared data environment. *Accounting Review* 57 (3): 554–78.

Newman, P., and R. Sansing. 1993. Disclosure policies with multiple users. *Journal of Accounting Research* 31 (1): 92–112.

Penrose, R. 1955. A generalized inverse for matrices. *Proceedings of Cambridge Philosophical Society* 51: 406–13.

Strang, G. 1986. *Introduction to applied mathematics*. Wellesley, MA: Wellesley-Cambridge Press.

———. 1998. *Introduction to linear algebra*. Wellesley, MA: Wellesley-Cambridge Press.

Sunder, S. 1997. *Theory of accounting and control*. Cincinnati, OH: International Thomson.

Williams, T., and C. Griffin. 1964a. Matrix theory and cost allocation. *Accounting Review* 39 (July): 671–78.

———. 1964b. *The mathematical dimension of accountancy.* Cincinnati, OH: South-Western.