The Information Content of Accounting Reports: An Information Theory Perspective *

Jonathan Ross†

November 23, 2015

Abstract

Is it possible to quantify the information content of earnings? If possible, then how? This study examines accounting as a classical communication system with the purpose of providing a framework with which to approach these fundamentally important questions. Information theory was established in the early-mid 20th century to describe the properties of classical communication systems. Applying concepts from this theory to an accounting context provides insight into the questions asked above. Specifically, a measure of the information content of financial statement numbers is developed from these information theory concepts. The measure is also applied to several large companies’ earnings numbers and aids in predicting their price movements.

Keywords: information theory, accounting information, uncertainty, entropy

*I gratefully acknowledge the helpful comments and suggestions of John Fellingham as well as workshop participants at Binghamton University and at the University of Kentucky.
†School of Management, State University of New York at Binghamton, jross@binghamton.edu.
1 Introduction

Does the financial accounting and reporting process provide non-redundant information to market participants? The over-arching goal of capital markets accounting research is to speak, in some way, to this question. The word “information”, and more pertinently, “accounting information” appears repeatedly in the literature. However, we do not capture “information”, as defined according to classical information theory, with our stock price-based measure of the information content of accounting numbers.

The first half of the 20th century brought about a revolution in how humans think about information. Claude Shannon (the father of modern information theory) was at the forefront of this revolution. His landmark 1948 paper, *A Mathematical Theory of Communication*\(^1\), was the first paper to formally describe a communication system in which information plays a central role. Concepts such as the capacity of an information channel, uncertainty of a source and the optimal rate of information transmission in a noisy environment revolutionized how we think about information. These concepts laid the groundwork for much of the technology we appreciate today (e.g. the computer, cryptography, telecommunications, television etc.). In the first paragraph of his landmark paper Shannon describes the principal problem of communication; ironically, a problem not too distant from the purpose of financial accounting and reporting . . .

*The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point.*

The purpose of this paper is to show why the long-standing criteria for deciding whether accounting numbers contain information does not logically reconcile with Shannon’s description of information in the context of a classical communication system. The financial accounting process can really be formalized as a classical communication system where information is determined to have been transmitted only when uncertainty regarding a future state-probability distribution has reduced. Consequently I develop a measure of the information content of accounting reports based on Shannon’s entropy (uncertainty) measure introduced in 1948. The measure captures the information in an observed earnings realization (for example) as the percentage change uncertainty (relative to maximum) regarding which state earnings will be in next period.

\(^1\)
2 Related Research

A search of three of the top accounting journals over each of their respective lives found 341 articles in The Accounting Review, 210 in Journal of Accounting Research and 99 in Journal of Accounting and Economics in which the word “information” appeared in the title. Clearly, accounting researchers are interested in the concept of information and how to measure it.

These articles represent two streams of literature\(^2\) which examine the information content of accounting numbers. Both streams borrow from information theory concepts originating with \(^2\), and define information as the change in the state-probability distribution regarding a specific event (variable) upon transmission of a message from a source to a user.\(^3\) What has differentiated these two streams is how each operationalizes the information content of an accounting message.

Theoretically, since accounting information falls under the realm of “information”, the definition given above is generalizable and maps well to an accounting context. The researcher quickly faces an immediate obstacle however when attempting to directly operationalize the definition. Specifically, how should the users’ ex ante (before the message is received) and ex post probability distributions be determined for a particular event or state of the world? These, of course, are not directly observable and therefore must be proxied.

\(^2\) was one of the first papers to examine the information content of accounting information; specifically, earnings. He states . . .

*The information content of earnings is an issue of obvious importance and is a focal point for measurement controversies in accounting.*

He looks at investors reaction to earnings announcements, as reflected in the volume and price movements of common stocks in the weeks surrounding the announcement date. He defines the information content of earnings as the *degree of change in investors’ assessments of the probability distribution of future returns (or prices), where this change is proxied for by the degree of change in the equilibrium value of a company’s stock upon their announcement of earnings.* This study laid the groundwork for a flurry of research\(^4\) which uses the stock price reaction to a firm-announcement

\(^2\)One stream currently remains.
\(^3\)The extant stream hardly conceptualizes information explicitly this way anymore. Rather this idea is lurking implicitly in the background.
\(^4\)For example see \(^7\), \(^8\), \(^9\), \(^11\), \(^12\), \(^13\) and \(^14\) to name a few.
as a proxy for the degree of change in investors’ assessments of the probability distribution of future returns and hence, as a proxy for the information content of that particular announcement.

An important distinction is in order. The definition given above does not mention the word “uncertainty”. The definition is rather vague and does not tell us exactly what these investors’ “assessments of the probability distribution of future returns” are. In order for the above definition to reconcile with “information” as defined according to classical information theory, one must introduce the notion of uncertainty. Information is uncertainty reduction. Thus, the information content of earnings is the reduction in uncertainty regarding future earnings. Defined this way however, the long-standing assumption that price equals discounted expected future earnings does not reconcile with this definition. Assuming price equals discounted expected future earnings implies that changes in price (returns) are equal to changes in discounted expected future earnings. As I discuss later, expected earnings does capture the various states that future earnings could take on as well as the probabilities of those respective states but does not capture changes in uncertainty regarding which state that future earnings will be in. Thus returns, viewed purely as a function of the change in discounted expected future earnings, do not capture changes in uncertainty and hence do not capture information. I show later that uncertainty regarding future earnings and expected future earnings are two different concepts. The first depends only on the state probabilities and the number of states and is independent of value. The second depends on the probabilities, the number of states and the value of earnings in each state. Given that information is uncertainty reduction, returns only capture information if one views price as a function of both uncertainty regarding future earnings and discounted expected future earnings.

Concurrent with Beaver’s idea regarding the information content of accounting numbers, a small stream of research surfaced whose intent was to try and capture investors state-probability distributions more directly. (? initiated interest in a communication theory approach to accounting information; pointing out that communication theory had been introduced in various other sciences such as experimental psychology, linguistics and biophysics but had not yet found its’ way into accounting. They state . . .

It seems reasonable to assume, however, that viewing accountancy as a communication process may provide a clearer picture of the nature and scope of the accounting function.
in an economic system. The opportunity exists because the underlying structure of communication theory may be used to describe the accounting process. (p. 650)

University of Chicago economist Henry Theil was the first to formally apply concepts borrowed from communication theory, specifically entropy, to an accounting context. and applied Shannons' information theory concept of entropy (uncertainty)\(^5\) to analyze the information content of financial statement items. He realized that every financial statement item can be expressed as some fraction of a total. For example, any particular type of asset can be expressed as a fraction of total assets. These fractions summed over all the assets equal one by construction similar to an individuals’ state probability distribution. He then allowed these fractions, observed in period \(t\), to proxy for an individuals’ ex ante \(t + 1\) state probabilities (i.e. prior probabilities) for assets in period \(t + 1\). Once assets in period \(t + 1\) were realized, he then allowed the observed \(t + 1\) fractions to proxy for the individual posterior probabilities. He then could calculate the entropy attributed to the \(t + 1\) asset realizations (message) and called this the information content of \(t + 1\) assets.

Subsequent to this initial application of information theory concepts to accounting reports, and examined the information loss due to different levels of aggregation in the financial statements. He devised a measure of information loss as the change in entropy induced by varying the level of aggregation. As in Theil, the probabilities used in the entropy calculation were simply ratios of the various respective categories of assets (liabilities) to total assets (liabilities).

The idea of applying information theory in an accounting context is very appealing due to the theoretical similarity that an accounting system has with a communication system (see Figure 1 in the next section). However, the way that the above studies went about this application doomed this stream of research from the start. The most glaring problem with the above applications is the fact that all fractions of totals are not formal state-probabilities. By definition, the probability of the occurrence of a particular state in a state-space is the number of ways that state could occur divided by the total number ways all the states could occur. Accordingly, one cannot let the fraction of cash to total assets in one period proxy for the ex ante probability that cash will be in any state the following period. This fraction is simply not a state-probability.

Second, no specification of a state space was provided in those studies. Such a specification is

\(^5\)The functional form for entropy was originally proposed in physics by J. Willard Gibbs (\(\?\)\)) to measure the amount of uncertainty in a particular set of particles (classical system).
essential to any probabilistic interpretation of information. Information uncertainty, as defined by (?), is a function of an individuals’ state-probability distribution; where probabilities are defined in terms of the states of the world that the individual believes could occur ex ante. The reader of the above studies is left wondering what the “probabilities” really represent. Does the fact that cash is 1/5 of total assets in year t mean that I believe cash will be the same fraction next year with probability 1/5? The point is that uncertainty (and hence information) cannot be thought of absent “true” state-probability distributions.6

This paper purposes to follow Shannon’s concepts as closely as possible to ensure that a theoretically correct application of communication theory to an accounting context is maintained. Along the way a few, seemingly limiting, assumptions must be made to make things tractable. However, one should judge the measure developed in regards to its’ predictive ability; some evidence of which is provided in the empirical analysis.

3 Accounting as a Communication System

The general purpose of accounting is to communicate to interested users regarding events that occurred in the past. We judge this communication to have been successful if these users are able, ex post, to “see” those events that transpired ex ante. To this end, we account for these events through time and at some point summarize this accounting by providing the user a set of summary reports.

As described above, accounting is simply a modified classical communication system. Such a system was the focus of Claude Shannons’ influential research and is depicted in Figure 1.

A communication system must begin with an information source which produces a message or sequence of messages to be communicated to an interested user. Economic events7 are the information source in an accounting context. These events transpire and give rise to information which interested users demand. Next, a transmitter, or encoder, must be present to operate on the message in some way to produce a suitable signal for transmission. The double-entry system fulfills this purpose. This linear operator insures that a record is maintained in at least two accounts for

---

6See (?)) who raised similar and additional concerns regarding Lev’s (and hence Theils’) approach.
7I define economic events as transactions which are accounted for within the double-entry system. Of course, the accounting cannot capture the information in macro-level events which are not felt at the firm level. To the extent that the firm-specific economic events are a function of the macro-level effects, this framework holds.
every economic event that transpires. There is some noise in this process however as GAAP are inherently subjective in their prescriptions. Furthermore, people are subject to error as well. Thus, the double-entry system operates on the message in a noisy way. This noise being a function of managerial error, bias and/or subjective interpretation of GAAP. The output from the transmitter is the signal; an encoded version of the message. The financial statements\(^8\) are the signal produced by the double-entry accounting system transmitter. The receiver, an auditor, receives the signal and decodes it. The goal of the decoding process is to try and recover the original message sent from the information source. The auditor performs a series of tests and procedures on the signal to ensure that it is as pure as possible. In a classical communication system, the receiver decodes the signal, recovers the original message with an arbitrarily small level of error and passes it on to its' destination. In an accounting context an auditor attempts to try and understand the original message from the signal but cannot recover it fully. The auditor simply helps to make the signal a better depiction of the underlying message than it was previously. They act as a sort of filter on the signal. The filtering process intends to remove as much of the error and bias in the signal as possible. The auditor then passes the filtered signal on to its destination. The destination is the intended recipient of the original message sent from the information source through the transmitter. The destination in an accounting context would be investors, creditors, regulators and any other interested user of the financial statements. The analogy between a classical communication system and accounting breaks down at this last step as the destination does not receive the original message, rather they receive a filtered signal that they must decode (audited financials).

Visualizing accounting with the framework described is a useful exercise as it helps us to realize that accounting is an example of a classical communication system. Of course, there are aspects of accounting that do not map well to the above framework as pointed out above. Despite this fact, visualizing accounting with this framework enables us to more easily identify important questions regarding the accounting process. For example, we see that the transmitter fulfills an important role in the process as it is the first stop the message makes towards the destination. The ability of the destination user to extract the original message depends first on certain properties of the transmitter. What are the important properties of a transmitter? It seems reasonable to require that it encode the message in such a way that accurate decoding is possible. That is, the function

\(^8\)The statements themselves absent the footnotes.
that encodes the message should have an inverse. If we perform the inverse operation of the transmitter we should be able to recover the original message. Also, it seems reasonable to require that the transmitter be self-correcting. Self-correcting in the sense that, once a signal is observed, we can immediately tell if there is an error, absent any noise in the encoding process. Double-entry accounting does satisfy this property as it ensures that the accounting equation is always in balance, unless an error has been made.

Another important question that arises from thinking about accounting as a communication system relates to the message itself. **Is there a way to quantify the information content of the message that the destination user receives?** Information theory was established as a way to analyze information content and other properties of classical communication systems. This seemed a natural progression as telegraph, telephony, radio and television had all been invented prior to this time period.

Claude Shannon’s influential work in the 1940’s described certain properties of classical communication systems mathematically and derived ways in which to ensure these properties hold. He analyzed both noiseless and noisy communication systems and proved many results relating to the optimal encoding of messages in noisy environments. He proved for example that it is impossible to encode a message in such a way (data compression) that the probability of information loss is arbitrarily small (Noise-less Coding Theorem). He also proved that, given a level of noise, it was possible to encode a message in such a way as to make the error in the resulting signal arbitrarily small (Noisy Coding Theorem).

Arguably Shannon’s most influential contribution however was his formulation of the uncertainty in a given message. His measure of uncertainty, information entropy, was the basis for much of his later work with communication systems (including the Theorems above). Ironically, entropy as a measure of uncertainty already existed in Physics at the time of his work; he just applied it in a classical communication system setting. He didn’t only apply it though. He proved his measure of uncertainty was the only function satisfying certain conditions one would intuitively expect such a measure to satisfy. This measure of uncertainty is the basis of the present paper, and in this measure lies one possible answer to the question bolded earlier. I expound on this in the next section.
4 Information Defined and Measured

Information is a rather elusive concept. As ?) point out, the concept of information is too broad to be captured completely by a single definition. Websters’ Dictionary defines information as follows.

Information – The communication or reception of knowledge or intelligence.

This definition immediately prompts the question: What is knowledge? Webster offers the following definition.

Knowledge – The fact or condition of knowing something with familiarity gained through experience or association.

This definition however seems rather circular. What is “knowing”? How do we know that we know? The study of knowledge is known as epistemology and has a long history dating back to philosophers Plato, Socrates, Confucious and others. The definition of knowledge has been highly debated throughout the centuries. Knowledge is one of those concepts that all of us can, in some sense, intuitively understand, yet when pressed for a general definition, we find it nearly impossible to formalize this intuition.

The primary purpose of this paper is not to offer and defend definitions. The purpose is to examine the concept of information in general and accounting information in particular in a way that hopefully will increase our understanding of this important and elusive construct. I therefore assume that we all adhere to some intuitive notion of knowledge being at the intersection of belief and truth. This is consistent with the overall consensus among philosophers throughout the centuries (e.g. see ?)).

4.1 A Definition of Information

Given the discussion above, I define information as follows\(^9\)

Information – Knowledge, after which one receives and processes, that changes, in an uncertainty changing way, their ex ante probability distribution regarding a set of propositions or states.

\(^9\)This definition is consistent with ?) treatment of the term.
Several points are in order. First, there is a time aspect to information. If no time elapses then information cannot exist. Second, information is a function of probabilities. Third, information is defined in light of certain propositions or states of the world that one has a belief about. Finally, because of these three things, information is often unique to the individual. Information to one person may not be information to another. Figure 2 offers a helpful visual depiction of information.

Prior to making a decision, an individual develops, based on knowledge they possess initially, a state-probability distribution, $P_I$, for each of the variables that could influence their decision. Then, time elapses and the individual either receives:

(a) Knowledge that is informative.

(b) Knowledge that is not informative.

(c) No knowledge.

In the first case, the ex-post probabilities assigned to the states ($P_A$) have changed in a way that changes the individuals’ uncertainty regarding which state the variable of interest will be in. The individuals’ uncertainty $U(\cdot)$ is a function of the probabilities\footnote{The notation $U(P)$ is somewhat unfortunate since it leads to confusion. However, following prior literature I use this notation to represent the uncertainty in, not the state-probabilities themselves, but which state the variable of interest will assume in the future. Therefore I assume that the individual is 100% certain in their probability assignments to these states.} and thus $U(P_I) \neq U(P_A)$. In the second case, regardless of whether the state-probability distribution changes or not, uncertainty has not changed and thus $U(P_I) = U(P_A)$. Finally, in the last case, the state-probability distribution for the variable of interest does not change so neither can the uncertainty and thus $U(P_I) = U(P_A)$.

To illustrate the above discussion, consider the following intuitive example. Suppose you form a discrete state-probability distribution regarding the weather tomorrow. You partition the continuous weather state-space $S$ into two discrete states; $S = \{\text{sunny}, \text{rainy}\}$. You observe that it has been both sunny and rainy for the past few days; alternating between sunshine and rain everyday. You therefore attach the following probabilities to the two states $P_I = (0.5, 0.5)$. After this initial probability assessment you watch the evening news; particularly the weather forecast. Upon receipt of the knowledge in the weather forecast you revise your probabilities appropriately to $P_A = (0.9, 0.1)$. Initially you were very uncertain as to the weather tomorrow; in fact you could
not have been more uncertain. After receiving the weather forecast however you are less certain regarding the weather since you feel that one of the states is much more probable than the other. Thus \( U(P_A) < U(P_f) \) (uncertainty has decreased). Therefore, the knowledge in the weather forecast was informative.

Now assume the same example as above except that, during the weather forecast, the weatherman spent the entire time talking about the incredible basketball game he witnessed the night before and never actually got to the weather forecast for tomorrow. You therefore do not adjust the state-probabilities and \( P_A = (0.5, 0.5) \). In terms of the weather tomorrow, the knowledge provided in the weather forecast had zero information content.

Finally, assume the same original example except that the television station had a blackout during the normal weather forecast time. You therefore assign \( P_A = (0.5, 0.5) \). No knowledge was disseminated during the forecast because there wasn’t one for you to observe. Your uncertainty regarding tomorrow’s weather didn’t change; no information was transmitted.

This example, along with the previous definition and discussion, highlights the distinction between information and knowledge. Information is a subset of knowledge. Information is knowledge that is useful; useful in the sense that probabilities are updated upon receipt of the knowledge in an uncertainty changing way. All information is knowledge but not all knowledge is information. Also, for information to exist, a source (transmitter) must exist. This can be another person, machine, or nature. Knowledge exists independent of a source. Knowledge is transmitted in the form of a message from a source to a user and the user decides on the informativeness of the message.

Finally, and somewhat more subtly, information, as defined, can exist in the absence of a decision context. Appealing to the example above, the individual who observes the weather forecast and updates their state-probability distribution in an uncertainty changing way views the forecast as informative even if none of their decisions depend on tomorrow’s weather.

### 4.2 Information and Financial Accounting

The purpose of accounting is to record, and communicate to interested users, the effect of economic events or transactions on an entity. The details of these events are passed through the double-entry system and summarized in a signal commonly known as the financial statements. This
signal is then operated on by a third, independent\textsuperscript{11} party which filters out noise and error then passes the signal on to the recipient (market participants).\textsuperscript{12} These recipients are assumed to be economically linked, in some way, to the entity and therefore have already formed ex ante probability distributions regarding the future states of the entity. Information therefore plays a central role in the financial accounting and reporting process. If the message (e.g. financial statements) does not change these users’ ex ante probability distributions regarding the future states of the entity, such that uncertainty changes, then the users are no better off after receiving and processing the message than they were before receiving the message. In this case, the user should be indifferent between these reports and a set of blank reports. Thus information plays a critical role in helping one assess if the “accounting” that an entity does, fulfilled its’ purpose. If it did not, then we are hard-pressed to find an economic benefit to offset the costs of doing the accounting.

Up to this point I have introduced a framework which hopefully has persuaded the reader that a direct measure of the information content of a message would be greatly valued; particularly in an accounting context. The definition of information offered earlier provides a clue to help us begin to develop such a measure. From the definition, if uncertainty regarding some state-probability distribution does not change upon receipt of the message, then information does not exist. Thus the information content of the message is zero. Therefore, it seems logical to think in terms of measuring the change in uncertainty of the state-probability distribution. I formalize this concept in the next section.

4.3 A Measure of Information Content

Suppose an individual, which we will label a “user”, has a state-space in mind regarding some variable of interest,\textsuperscript{13} $j$, that may or may not affect a future decision. Denote the state-space as $S_j = \{S; P(S|K_I)\}$ where $S = \{s_1, s_2, \ldots, s_n\}$ is a discrete set of $n$ states that the variable $j$ can take on. $P(S|K_I) = P_I = \{\alpha_{1I}, \alpha_{2I}, \ldots, \alpha_{nI}\}$ is a set of probabilities for each of these states assessed from knowledge possessed initially\textsuperscript{14} by the user. We assume $S$ is exhaustive from the users’ standpoint so that $\sum_{i=1}^{n} \alpha_i = 1$. That is, from the users’ standpoint, $j$ must be in one of the

\textsuperscript{11}Supposedly “independent” in the case of the auditing firms.
\textsuperscript{12}See Figure 1.
\textsuperscript{13}e.g. The “weather” in the example I gave previously.
\textsuperscript{14}That is, $P(S|K_I)$ is the ex ante, state-probability distribution for variable $j$. 
states of $S$. Now, theoretically, $j$ could be continuous and take on infinitely many states. The user however, due to limited cognitive processing ability, does not view $j$ as thus. She partitions the continuous variable $j$ into a set of $n$ states and attaches probabilities to those states. Also, $j$ cannot be in more than one state at a time and I also assume that each of these $n$ states is distinct (non-overlapping).\textsuperscript{15}

Next, a message, $M$, is sent to the user from a source. Upon receipt of the message, the user processes the knowledge contained therein and updates her probability distribution to $P(S|K_A) = P_A = \{\alpha_{1A}, \alpha_{2A}, \ldots, \alpha_{nA}\}$. Based on the discussion in Section 4.1, let the information content of $M$, $IC(M)$, be defined as the absolute value of the change in the users’ uncertainty regarding the state-probability distribution $P(S|K)$ as shown next:

$$IC(M) = |\Delta U| = |U(P_I) - U(P_A)|$$

$$= |U(\{\alpha_{1I}, \alpha_{2I}, \ldots, \alpha_{nI}\}) - U(\{\alpha_{1A}, \alpha_{2A}, \ldots, \alpha_{nA}\})|$$

(1)

The expression in (1) is consistent with the definition of information provided earlier. When uncertainty does not change, $IC(M) = 0$. When it does, $IC(M) > 0$ and we say that the message contained information; information in the sense that, upon receiving the message and processing the knowledge contained therein, the user was able to update their state-probability distribution in a way that changed their uncertainty regarding which state $j$ would take on in the future.

Based on the definition of information given earlier and equation (1), Figure 3 visually depicts the framework that I have set up.

Notice, in Figure 3, that the uncertainty function, $U(\cdot)$, is central to being able to quantify the information content of the message. To further highlight the importance of the role that $U(\cdot)$ plays in the concept of information, consider another example.

Suppose the user is possessed with the following knowledge initially.

$$K_I = \{\text{Urn with 5 balls; 2 red and 3 white, 1 ball is drawn at a time without replacement}\}$$

Also, suppose the variable of interest, $j$, is the color of the ball that will be drawn on the second draw. Thus the state set $S = \{r_2, w_2\}$ where $r_2$ represents a red ball on the second draw and $w_2$

\textsuperscript{15}That is, $s_i \cap s_k = \emptyset$ for all $i, k$. 

represents a white ball on the second draw. Based on the knowledge possessed initially, the user reasons that, if red is drawn on the first draw, then \( P(r_2|K_I) = \frac{1}{4} \). However, if white is drawn on the first draw then \( P(r_2|K_I) = \frac{2}{3} \). Thus, given their initial knowledge, they are uncertain regarding the probability of red being drawn on the second draw. If red is drawn on the second draw then, either \( r_1r_2 \) or \( w_1r_2 \) will have been realized. Of course, given \( K_I \), the user can assess the probabilities of both of these events and add them together to arrive at their assessment of \( P(r_2|K_I) \). Multiple applications of the product rule\(^{16}\) for probabilities can then be applied to arrive at the following:

\[
P(r_2|K_I) = P(r_1r_2|K_I) + P(w_1r_2|K_I)
\]

\[
= P(r_1|K_I) \cdot P(r_2|r_1K_I) + P(w_1|K_I) \cdot P(r_2|w_1K_I)
\]

\[
= \frac{2}{5} \cdot \frac{1}{4} + \frac{3}{5} \cdot \frac{2}{4}
\]

\[
= \frac{2}{5}
\]

Thus, given \( K_I \), the user is uncertain as to which ball will be drawn on the first draw. They therefore, logically, assign the probability of red being drawn on the second draw as \( \frac{2}{5} \).\(^{17}\) Following the above logic, they attach a probability of \( \frac{3}{5} \) to the second state \( w_2 \) and thus their state-probability distribution, based on knowledge possessed initially, is \( P(S|K_I) = \{2/5, 3/5\} \).

Now suppose a message is sent to the user with the knowledge that red was drawn on the first draw. Thus the user has the following knowledge set after the message

\[K_A = \{K_I, M = r_1\}\]

where \( r_1 \) represents the message that a red ball was drawn on the first draw. With this new level of knowledge the user updates their state-probability distribution to \( P(S|K_A) = \{\frac{1}{4}, \frac{3}{4}\} \). Given \( K_A \), the user is more certain regarding the color of the ball that will be drawn on the second draw. Precisely, the message has provided knowledge which eliminates the users’ uncertainty regarding the probability that red or white will be drawn on the second draw.

The preceding example highlights the importance of the uncertainty function \( U(\cdot) \). We need a way to quantify the change in uncertainty resulting from the new knowledge eliminating the uncertainty regarding \( P(r_2, w_2) \). The user, given \( K_A \), is still uncertain as to which ball will be drawn on the first draw.
drawn on the second draw; they just are less uncertain as they were given only $K_I$. When applied to this example, an uncertainty measure that is devised, should show that uncertainty decreased and thus $IC(M) = |\Delta U(\cdot)| > 0$.

At this point I turn to the problem of specifying such a measure of uncertainty, $U(\cdot)$. I was interested in the same problem and set forth three intuitive conditions that such a measure, $U(\cdot)$, should possess. If such a function $U(\cdot)$ exists, it seems reasonable to require of it the following properties:

(1) $U(P)$ should be continuous in the state probabilities $P$. That is, for small changes in $P = \{p_1, p_2, \ldots, p_n\}$ we should observe small changes in $U(P)$ (no jumps in uncertainty). Stated more formally...

$$
\text{For all } \epsilon > 0 \text{ there exists } \delta > 0 \text{ such that when } |P_A - P_I| < \delta \text{ then } |U(P_A) - U(P_I)| < \epsilon \text{ where } P_I \neq P_A \text{ are any probability distributions.}
$$

(2) If all $n$ states have equal probability of occurring (i.e. all the $p_i$ are equal); $p_i = \frac{1}{n}$ for all $i$, then $U(P)$ should be a monotonic increasing function of $n$. That is, $U(P)$ should never decrease when we add more equally probable states. Stated differently; when all states are equally likely, there is more uncertainty regarding which state will be realized when there are more possible states.

(3) The uncertainty $U(P)$ of a compound set of states $S$ should be equal to a weighted average of the uncertainties of any particular, mutually exclusive partitioning of $S$; the uncertainties of these partitions being weighted by their respective probabilities of occurrence. Formally, suppose we are given a compound set of states (i.e. a set of states which, themselves, are sets of states) $S = \{S_1, S_2\} = \{s_{11}, s_{12}, s_{21}, s_{22}\}$ with $S_1 = \{s_{11}, s_{12}\}$ and $S_2 = \{s_{21}, s_{22}\}$. Also suppose $P(S) = \{p_1, p_2\}$ with $P(S_1) = \{p_{11}, p_{12}\}$ and $P(S_2) = \{p_{21}, p_{22}\}$. Finally, suppose the probabilities $P(S_1) = p_1 \ast T_1 = \{p_1 \ast t_{11}, p_1 \ast t_{12}\}$ and $P(S_2) = p_2 \ast T_2 = \{p_2 \ast t_{21}, p_2 \ast t_{22}\}$ where $(T_1, T_2) = \{t_{11}, t_{12}, t_{21}, t_{22}\}$ are the transition probabilities of moving from $S_1$ to each of its’ two states respectively or of moving from $S_2$ to each of its’ two states respectively.

Then the following holds:

18 Assumptions, if you will.
\[ U(P(S)) = p_1 \ast U(T_1) + p_2 \ast U(T_2) \]  \hspace{1cm} (2)

Condition (2) is illustrated with the following example. Suppose we are given \( S \), \( S_1 \) and \( S_2 \) with \( P(S) = \{ \frac{1}{2}, \frac{1}{2} \} \), \( P(S_1) = \{ \frac{1}{3}, \frac{1}{6} \} \) and \( P(S_2) = \{ \frac{1}{4}, \frac{1}{4} \} \). Thus the following probability tree diagrams in Figure 4 are apparent:

Then we have, using equation (2):

\[
U(P(S)) = U(P(S_1, S_2)) \\
= U(p_{11}, p_{12}, p_{21}, p_{22}) \\
= U\left(\frac{1}{3}, \frac{1}{6}, \frac{1}{4}, \frac{1}{4}\right) \\
= p_1 \ast U(T_1) + p_2 \ast U(T_2) \\
= \frac{1}{2} \ast U\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{1}{2} \ast U\left(\frac{1}{2}, \frac{1}{2}\right)
\]

Shannon demonstrated that the only function satisfying the above three conditions was of the following form:\(^{19}\)

\[ U(P) = -C \sum_{i=1}^{n} p_i \ast \log(p_i) \]  \hspace{1cm} (3)

where the constant \( C > 0 \) simply amounts to a choice of a unit of measure\(^{20}\) and \( \log(p_i) \) is the base 2 logarithm of the \( i \)'th state probability.

As pointed out by Shannon, quantities of the form \( -\sum_{i=1}^{n} p_i \ast \log(p_i) \) play a central role in information theory as measures of information, choice and uncertainty. Quantities of this form also played a central role in statistical mechanics before Shannon’s independent discovery of their importance and usefulness in information theory. In fact J. Willard Gibbs was one of the founders of “statistical mechanics”, a branch of theoretical physics concerned with investigating the thermodynamic properties of systems using mathematical/statistical analysis of the large ensembles of particles of which these systems are comprised. Gibbs described entropy as a measure of uncertainty, or “mixedupness”, that remains about a system after its observable macroscopic properties

\(^{19}\)See 7), Appendix 2.

\(^{20}\)For simplicity, and without loss of generality, I set \( C = 1 \) throughout the rest of this paper.
such as temperature, pressure and volume, have been taken into account. He used (3) to quantify the entropy of a system which can be in any one of \( n \) possible microstates where \( p_i \) is the probability of the system being in the \( i \)th microstate of its phase space.

Now, appealing to equations (1) and (3) and using the setup before with \( S = \{s_1, s_2, \ldots, s_n\} \), \( P_I = \{\alpha_{1I}, \alpha_{2I}, \ldots, \alpha_{nI}\} \) and \( P_A = \{\alpha_{1A}, \alpha_{2A}, \ldots, \alpha_{nA}\} \), the information content, \( IC(M) \), of a message, \( M \), given to a user whose interest is in the state that variable \( j \) will take on in the future is:

\[
IC(M) = |\Delta U| = |U(P_I) - U(P_A)| = \left| -\sum_{i=1}^{n} \alpha_{iI} \ast \log(\alpha_{iI}) - \left( -\sum_{i=1}^{n} \alpha_{iA} \ast \log(\alpha_{iA}) \right) \right| \quad (4)
\]

At this point we now have a functional form for \( U(P) \) (and hence \( IC(M) \)) and some reasonable assumptions that this measure of uncertainty satisfies. Next I will examine optimization of \( U(P) \).

I will also illustrate \( IC(M) \) with two useful examples.

### 4.4 Maximizing \( U(P) \)

It turns out that \( U(P) \), and hence \( IC(M) \), is a formal mathematical measure since it satisfies all of the properties that such a measure should possess. That is, \( U(P) \) is non-negative for all \( P \), is countably additive (by condition (2)) earlier and is zero when \( P = \emptyset \). Additionally, as pointed out in condition (1), \( U(P) \) is continuous in the probabilities and is transitive. Also, \( U(P) \) attains it’s maximum when the states in \( S \) are equiprobable; something we would intuitively expect. That is, we are most uncertain regarding which state in \( S \) will be realized when the states in \( S \) have equal probability of occurrence. To see this one simply maximizes \( U(P) \) subject to the constraint that the state-probabilities must sum to one.\(^{21}\)

The information content measure given earlier in equation (4) may be difficult to interpret because it is a raw measure which depends on the initial uncertainty and this uncertainty is bounded above by \( \log(n) \) and hence clearly depends on the number of states. To address this, I will now utilize the optimal value of \( U(P) \) to put an upper bound on \( IC(M) \) such that we can interpret and

\(^{21}\)I omit the proof for the sake of space.
make more meaningful comparisons of the information contained in given messages. If one is faced
with some initial level of uncertainty \(U(P_I)\) and they receive a message, they could move toward
a higher level of uncertainty (uncertainty increases), or they could move toward a lower level of
uncertainty (uncertainty decreases). If we think of a continuum of uncertainty with zero at one end
and \(\log(n)\) at the other end then both \(U(P_I)\) and \(U(P_A)\) lie on this continuum. If \(U(P_I) > U(P_A)\)
then uncertainty decreased upon receipt of the message. The information content of the message
could be thought of as the percentage decrease relative to the maximum decrease that could be
obtained. Conversely, if \(U(P_I) < U(P_A)\) then uncertainty increased and the information content of
the message could be thought of as the percentage increase relative to the maximum increase that
could be obtained. These scenarios are illustrated in Figures 5 and 6.

In Figure 5, \(U(P_I) < U(P_A)\) thus uncertainty increases by \(|a - b|\) and the maximum increase is
\(\log(n) - a\) where \(n\) is the number of states in the perceived state-space. Therefore the information
content of the message is \(\frac{|a - b|}{\log(n) - a} = \frac{|U(P_I) - U(P_A)|}{\log(n) - U(P_I)}\). In Figure 6, \(U(P_I) > U(P_A)\) thus uncertainty
decreases by \(|a - b|\) and the maximum decrease is \(a\). Therefore the information content of the
message is \(\frac{|a - b|}{a} = \frac{|U(P_I) - U(P_A)|}{U(P_I)}\). Of course when \(U(P_I) = U(P_A)\) uncertainty does not change
and I assume that the information content of the message is zero.\(^{22}\)

The information content of a message is summarized below:

\[
IC(M) = \begin{cases} 
\frac{|U(P_I) - U(P_A)|}{U(P_I)} & \text{if } U(P_I) > U(P_A) \\
\frac{|U(P_I) - U(P_A)|}{\log(n) - U(P_I)} & \text{if } U(P_I) < U(P_A) \\
0 & \text{if } U(P_I) = U(P_A)
\end{cases}
\]  

(5)

where \(U(P_I)\) and \(U(P_A)\) are defined as in equation (4). The function in equation (5) measures
the information content of a message in general. Now \(IC(M) \in [0, 1]\) and can be stated in percentage
terms. This is possible due to the fact that \(U(P) = \sum_{i=1}^{n} p_i \log(p_i)\) is bounded above by \(\log(n)\). I
now provide two examples to illustrate \(IC(M)\).

\(^{22}\)I discuss this assumption in more detail in the “Limitations” section.
4.5 Information Content Examples

Revisiting the urn example given in section 4.3, we know intuitively that, given the individual wishes to know the color of the ball that will be drawn on the second draw, there should be information content in the message that a red ball was drawn on the first draw (call this \( IC(M_1) \)). That is, upon receipt of such a message, she is more certain regarding the color that will be drawn on the second draw. \( IC(M_1) \) simply quantifies this reduction in uncertainty and is given below.

Keep in mind that \( P_I = \{ \alpha_{1I}, \alpha_{2I} \} = \{ \frac{2}{5}, \frac{3}{5} \} \) and \( P_A = \{ \alpha_{1A}, \alpha_{2A} \} = \{ \frac{1}{4}, \frac{3}{4} \} \). Using equation (5) we have:

\[
IC(M_1) = \frac{|U(P_I) - U(P_A)|}{U(P_I)} = \frac{\sum_{i=1}^{2} \alpha_{ii} \log(\alpha_{ii}) - \sum_{i=1}^{2} \alpha_{iA} \log(\alpha_{iA})}{\sum_{i=1}^{2} \alpha_{ii} \log(\alpha_{ii})}
\]

\[
= \frac{|\left[-\left(\frac{2}{5} \log \left(\frac{2}{5}\right) + \frac{3}{5} \log \left(\frac{3}{5}\right)\right)\right] - \left[-\left(\frac{1}{4} \log \left(\frac{1}{4}\right) + \frac{3}{4} \log \left(\frac{3}{4}\right)\right)\right]|}{\left[-\left(\frac{2}{5} \log \left(\frac{2}{5}\right) + \frac{3}{5} \log \left(\frac{3}{5}\right)\right)\right] - \left[\frac{2}{5} \log \left(\frac{2}{5}\right) + \frac{3}{5} \log \left(\frac{3}{5}\right)\right]}
\]

\[
\approx 0.1644
\]

Thus, the information content of a message telling her that a red ball was drawn on the first draw, is 0.1644. Analogously, the message decreases her uncertainty by 16.44% of the maximum that it could have decreased it by. Therefore, qualitatively, this message does not have very much information content.

Now, suppose she wants to know the color of the ball that will be drawn on the third draw and we give her a message that a red ball was also drawn on the second draw. Prior to this message she has the following state-probability distribution \( P_I = \{ \frac{1}{4}, \frac{3}{4} \} \). After receipt of this message, she knows there are no red balls left and a white ball must be drawn on the third draw. She revises her probabilities to \( P_A = \{ 0, 1 \} \). The information content of this message, \( IC(M_2) \) is therefore:
\[ IC(M_2) = \frac{|U(P_I) - U(P_A)|}{U(P_I)} \]

\[ = \left| \left[ -\left( \frac{1}{4} \log \left( \frac{1}{4} \right) + \frac{3}{4} \log \left( \frac{3}{4} \right) \right) \right] - \left[ 0 \log(0) + 1 \log(1) \right] \right| \]

\[ = 1 \]

Thus, \( M_2 \) reduces her uncertainty by the maximum amount that it could have such that she is left with zero uncertainty regarding the color of the ball that will be drawn on the third draw. Therefore, of all the messages she could have received that would influence her state-probability distribution for the color of the third draw, \( M_2 \) contained the most information; consistent with her intuition.

The next example illustrates the information content of a message that increases uncertainty. Suppose you are a shareholder of Apple Inc. stock in 2011. The primary variable of interest to you is the future profitability (2012 and beyond) of Apple. You form an initial probability distribution \( P_I = \text{(high, medium, low)} = \{ \frac{3}{4}, \frac{2}{16}, \frac{1}{16} \} \) regarding the future profitability of Apple. You assign these probabilities based on the recent success of Apple and the history of Apple being an innovative leader in the consumer electronics industry. In October of 2011 you receive the message that Steve Jobs has passed away. You still feel Apple has a higher likelihood of high profitability than of medium or low profitability but you feel that the loss of Steve Jobs could hurt the company’s future profitability, particularly due to the intense competition existing in the consumer electronics industry. Your uncertainty increases\(^{23}\) regarding the future of Apple and therefore you revise your state-probability distribution accordingly to \( P_A = \{ \frac{1}{2}, \frac{1}{3}, \frac{1}{3} \} \). The information content of the message is thus:

\[^{23}\text{One should check that } U(P_I) < U(P_A).\]
\[ IC(M) = \frac{|U(P_T) - U(P_A)|}{\log(n) - U(P_T)} \]

\[ = \frac{|\left[-\left(\frac{3}{4}\log\left(\frac{3}{4}\right) + \frac{3}{16}\log\left(\frac{3}{16}\right) + \frac{1}{16}\log\left(\frac{1}{16}\right)\right)\right] - \left[-\left(\frac{1}{2}\log\left(\frac{1}{2}\right) + \frac{1}{4}\log\left(\frac{1}{4}\right) + \frac{1}{4}\log\left(\frac{1}{4}\right)\right)\right]|}{\log(3) - \left[-\left(\frac{3}{4}\log\left(\frac{3}{4}\right) + \frac{3}{16}\log\left(\frac{3}{16}\right) + \frac{1}{16}\log\left(\frac{1}{16}\right)\right)\right]} \]

\[ \approx 0.85 \]

Thus the message that Steve Jobs has died increases your uncertainty by 85% of the maximum possible amount it could have. Qualitatively, of all the messages you could have received, this particular message has relatively high information content. Depending on the amount of risk you are willing to accept, you may now be “informed” enough to want to sell your shares. \( IC(M) \) assigns a number between zero and one to the “information content” of the message that Steve Jobs has passed and \( IC(M) \) is directly a function of the change in uncertainty surrounding your state-probability distribution of a particular variable of interest; the future probability of Apple Inc.

The above examples help to illustrate how \( IC(M) \) measures information content and also, how \( IC(M) \) is quantitatively consistent with our qualitative intuition regarding the information content of a particular message.

In the next section I apply \( IC(M) \), as given in equation (5) to the quantitative financial statement information. Note that the measure introduced in section 4.4 can measure the information content of any time-series variable; not just those variables typically found in the financial statements.

5 Applying the Measure to Financial Statements

5.1 Mapping the Quantitative Financial Information to States

Consider a set of financial statements.\(^{24}\) These financial statements are each made up of a finite set of \( k \) variables, \( V = \{1, 2, \ldots, j, \ldots, k\} \), that are reported on each period. Without loss of generality, consider one of these variables, say \( j = \text{earnings} \). Finally, consider a representative

\(^{24}\)i.e. A balance sheet, income statement, statement of cash flows and statement of retained earnings.
individual. That is, consider an individual whose beliefs represent those of the market of interested users as a whole. I wish to apply $IC(M)$ to determine the information content of the number that is reported for $j$ each period to this representative user. Any discussion of information must begin with a state set; a set of states that the user feels that variable of interest $j$ could possibly take on. In the context of the financial statements, $j$ is continuous and does not admit a discrete state-probability distribution by itself. Rather, a function must be developed to map the variable of interest $j$ to a discrete state-set.\footnote{Unless we are privy to the users’ true, perceived, continuous state-probability distribution for $j$. If we know this then the uncertainty measure in equation (3) can easily be modified to the continuous case.} This immediately gives rise to the problem of which function to choose and how many states of the world does one consider variable $j$ could take on. Suppose I choose a function, $f(j_t)$ which assigns each realization of $j$ to one of four possible states where $j_t$ is the value of $j$ reported for period $t$. The function I have in mind takes the following form:

$$f(j_t) = \begin{cases} 
H & \text{for } j_t > \mu_{j_t} + \sqrt{2}\sigma_{j_t} \\
HM & \text{for } \mu_{j_t} \leq j_t \leq \mu_{j_t} + \sqrt{2}\sigma_{j_t} \\
LM & \text{for } \mu_{j_t} - \sqrt{2}\sigma_{j_t} \leq j_t < \mu_{j_t} \\
L & \text{for } j_t < \mu_{j_t} - \sqrt{2}\sigma_{j_t} 
\end{cases}$$

(6)

where $H$, $HM$, $LM$ and $L$ represent that $j_t$ is in a high, high-medium, low-medium or low state respectively. Notice that $f(j_t)$ is not defined for $t = 1$. The idea is that I observe the reporting of $j_t$ and then map $j_t$ to one, and only one, state in my perceived state-space $S = \{H, HM, LM, L\}$ based on the mean, $\mu_{j_t}$, and standard deviation, $\sigma_{j_t}$, of the values of $j$ that have been realized through time period $t$.\footnote{I assume $\mu_j$ and $\sigma_j$ both exist and are finite for $t \geq 2$.} Of course this is not feasible initially for $j_1$. Instead, I wait until $j_2$ is realized and then map $j_1$ to one of the four states in $S$ using the function in equation (6).

The function $f(j_t)$ maps each earnings realization $j_t$ to one of four possible states in the state set $S = \{H, HM, LM, L\}$. Let the set $X_T = \{f(j_1), f(j_2), \cdots, f(j_T)\}$ be the set of all values $f(j_i)$, where $T$ is the total number of realizations of $j$.\footnote{For example, if five earnings realizations have occurred, a possible scenario could be $X_T = \{HM, LM, LM, H, HM\}$.} Figure 7 illustrates $f(j_t)$.

At this point, one may feel that $f(j_t)$, as specified in equation (6), has been picked out of thin air and is rather arbitrary. Two sources of arbitrariness are perceived to be present. The
first is: why choose four states? I admit that there is some arbitrariness in choosing four states. Mathematically, $U(P)$, and hence $IC(M)$, depends on the number of states. This is implied by condition (2) set forth earlier. Given equally probable states, the more states we add, the more uncertainty. Although the magnitude of $IC(M)$ depends on the number of states, the interval and ratio properties of $IC(M)$ do not. We will often want to compare the information contents of earnings releases within-firm across time and across firms. These comparisons do not depend on the number of states chosen, as long as we remain consistent in our choice. Finally, four states seems reasonably intuitively as well. An individual may view a variable as being high, medium or low but then wonder on which side of medium the variable is: closer to low or closer to high? Therefore I model the individual as thinking of the continuous variable $j$ as being in a high, high-medium, low-medium or low state.

The second source of arbitrariness lies in the specific choice of $f(j_t)$. It turns out that the choice of $f(j_t)$ in equation (6) is the only unbiased choice\textsuperscript{28} of function, given the function depends on the mean and standard deviation of the variable $j$. Here, I appeal to Chebyshev’s theorem. Precisely stated, let $c$ be any number greater than one. Then, for any sample of data, the proportion of observations lying fewer than $c$ standard deviations from the sample mean is at least $1 - \frac{1}{c^2}$. If $c = \sqrt{2}$ then Chebyshev implies that at least 50% of the observations of $j$ lie within $\sqrt{2}$ standard deviations of $\mu$.\textsuperscript{29} As defined in equation (6), these values of $j$ are mapped to two of the four states ($HM$ and $LM$) respectively. Thus, no more than 50% of the values of $j$ will be mapped to $H$ and $L$. Note that Chebyshev doesn’t say anything about the proportion of observations between $\mu$ and $\mu + c\sigma$ for example; the underlying distribution of $j$ will determine this. The theorem only provides bounds (rather loose ones admittedly) on the proportion of observations that lie in the interval $[\mu - c\sigma, \mu + c\sigma]$. Thus, Chebyshev directly implies the four-state mapping function as specified in equation (6) if we want to partition the state-set $S$ in such a way to make it possible for an equal proportion of $j$ values to be mapped to each half of the state set. This ensures, as much as possible, that a particular value of $j$ will not, mechanically, be more likely to be assigned to one of the states.\textsuperscript{30}

\textsuperscript{28}Unbiased in the sense that $f(j_t)$ is the only function depending on $\mu_{j_t}$ and $\sigma_{j_t}$ that creates a mutually exclusive partition of $S$ which allows the possibility of assigning an equal proportion of $j$ values to each half of the state set.

\textsuperscript{29}To see this, simply set $1 - \frac{1}{c^2} = 0.5$ and solve for $c$.

\textsuperscript{30}As will be seen later, the function chosen in equation (6) doesn’t remove all determinism. There still remains the mechanical assigning of early observations of $j$ to the states. This determinism, however, decreases rapidly over
5.2 Forming the State-Probability Distributions \( P_I \) and \( P_A \)

With each value of \( j_t \) now mapped to a particular state, I now can define the probability-state distributions, \( P_{It} \) and \( P_{At} \), she perceives for \( j_t \) at the beginning and end of period \( t \) respectively as follows:\(^{31}\)

\[
P_{I1} = \left\{ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right\} = P_{I2} = P_{A1}
\]

\[
P_{At} = P_{It+1} = \left\{ \frac{a}{T}, \frac{b}{T}, \frac{c}{T}, \frac{d}{T} \right\} \quad t \geq 2
\]

where \( a = \text{count}(X_T = L), \ b = \text{count}(X_T = LM) \)
\[
c = \text{count}(X_T = HM), \ d = \text{count}(X_T = H)
\]

Equation (7) reflects the assumption that, initially, she is completely uncertain as to which state \( j_1 \) will be in\(^ {32} \) and therefore she perceives the four states of \( S \) as being equiprobable. Upon receipt of the second message, \( j_2 \) (e.g., earnings in period two), she revises \( P_{I2} \) accordingly to \( P_{A2} \). This probability distribution is also her initial probability distribution regarding the state that \( j_3 \) will be in and so forth. Thus \( P_{At} = P_{It+1} \) for all \( t \geq 2 \). In forming \( P_{At} \) I apply \( f(j_t) \) to each of the prior realizations of \( j \). This produces the set, \( X_T = \{ f(j_1), f(j_2), \cdots, f(j_T) \}, \ T \geq 2 \), described in the previous section. I simply count the number of observations that were mapped to each state respectively and divide this frequency by the number of periods that have passed. This produces a frequency of occurrence for each state through time period \( T \). I use this frequency to proxy for the ex ante probabilities that she forms for the states of \( S \). An investor likely looks to past earnings in forming a state-probability distribution for future earnings. For example, suppose three out of the first five earnings announcements have been “high” relative to the mean. I model the user as thinking that the earnings in the sixth year will be “high” with 60% probability.

\(^{31}\)We will never be able to know individuals’ true probability assignments, assuming they consciously form them. To apply an information theory approach, the best we can do is to proxy for these assignments in an intuitive way. \( f(j_t), P_I \) and \( P_A \) intend to do this.

\(^{32}\)Intuitive with the notion that uncertainty is at its highest surrounding a firm’s first couple of earnings announcements due to the fact that we have little-to-no, prior, firm-specific history to utilize in forming an expectation.
5.3 The Information Content, \( IC(j) \), of Financial Statement Variable \( j \)

Following equation (5), the information content, \( IC(j) \), of financial statement variable \( j \) at time \( t \) is as follows:

\[
IC(j_t) = \begin{cases} 
\frac{|U_j(P_{It})-U_j(P_{At})|}{U_j(P_{It})} & \text{if } U_j(P_{It}) > U_j(P_{At}) \\
\frac{|U_j(P_{It})-U_j(P_{At})|}{\log(4)-U_j(P_{It})} & \text{if } U_j(P_{It}) < U_j(P_{At}) \\
0 & \text{if } U_j(P_{It}) = U_j(P_{At})
\end{cases}
\] (9)

where \( P_{It} \) and \( P_{At} \) are as defined in equations (7) and (8).

To illustrate, consider the annual earnings, \( j \), of Apple Inc. from the date it went public through the present time (1981-2012).\(^{33}\) Table 1 reports the results from applying equations (6)-(9) to calculate \( IC(j_t) \) for each of these thirty-three earnings realizations. The information content values, \( IC(j_t) \), are superscripted with “−” if \( j_t \) led to an uncertainty decrease or “+” if \( j_t \) led to an uncertainty increase regarding future earnings realizations.

To interpret \( IC(j_t) \) from Table 1, observe the information content of earnings in 1981 (0.5−). This means that the earnings “message” for 1981 decreased our uncertainty\(^{34}\) by 50% of the maximum it could have decreased it by.

One important point illustrated in Table 1 is that \( IC(j_1) \) and \( IC(j_2) \) are mechanical realizations. Since \( P_{I1} = P_{A1} \) by construction, the information content of first-period earnings will always be zero, independent of the company analyzed. Also, the information content of second-period earnings will always be 0.5− independent of the company analyzed. This is due to the mathematical fact that, given any two real numbers \( a > b \), the largest of the numbers \( a = \mu + \frac{\sqrt{2}}{2}\sigma \) and \( b = \mu - \frac{\sqrt{2}}{2}\sigma \) where \( \mu \) and \( \sigma \) are the mean and standard deviation of the two numbers respectively. This, along with equations (6) and (8) immediately imply that the LM state and HM state will be assigned \( \frac{1}{2} \) probability. Once three earnings realizations have been realized, however, equations (6) and (8) allow for a little more variation in uncertainty and hence variation in \( IC(j_t) \). As the number of earnings “messages” released by the company increases, \( IC(j_t) \) becomes less and less mechanical. In essence, \( j_t \) begins to determine \( IC(j_t) \) rather than the mechanical construction set up in equations (6) and (8). Thus, over time, \( IC(j_t) \) better reflects the information content of \( j_t \).\(^{35}\)

---

\(^{33}\)Earnings is the number reported annually for Compustat variable \( NI \).

\(^{34}\)Our uncertainty regarding future realizations of earnings upon receiving the earnings “message” for 1981.

\(^{35}\)That is, as time increase, \( IC(j_t) \) better reflects the intuitive definition of information offered in section 4.1.
Keeping in mind equations (6)-(9), Table 1 illustrates the idea that the earnings realizations, \( j_t \), or “messages” if you will, help us to update our priors. \( IC(j_t) \) captures this updating and hence the information contained in each earnings release. In essence, \( IC(j_t) \) is a mathematical way to extract as much information as possible out of the earnings realizations. Of course, \( IC(j_t) \), as constructed, is only a function of the earnings realizations themselves and does not take into account any qualitative information surrounding an earnings release (such as contained in the footnotes or a press release for example). Therefore, one could view \( IC(j_t) \) as forming a lower bound on the information content of a given earnings release. I discuss this limitation in more detail later.

Appendix A illustrates how to use the measure developed in equations (6)-(9) to assess the information content of the financial statements as a whole rather than one particular variable within those statements. This simply becomes an exercise in aggregation, however, one must pay attention to the fact that part of (if not all of) the information content of one variable may already be subsumed by another variable due to inter-variable dependencies within the financial statements.

### 6 An Empirical Application

Theoretically \( IC(v_{jt}) \) is a perfectly valid measure of the information content of a given financial statement variable \( j \) at time period \( t \). The question however becomes; does the framework adhered to in this paper hold in the real world? I acknowledge that it is unlikely that investors form state-probability distributions exactly the way specified in this paper. In fact, investors likely do not even consciously form state-probability distributions! Investors also likely do not consciously develop utility functions and seek to maximize them. Many economic studies, however, provide evidence consistent with the utility maximization assumption. The primary purpose of this paper is not to test the validity of the measures contained therein; I leave that to a future paper. That being said, there is one, interesting, potential application of \( IC(v_{jt}) \) that I will offer.\(^{36}\)

You will notice from Table 1 that there have been 13 uncertainty-increasing annual earnings’ releases and 19 uncertainty-decreasing earnings’ releases over Apples’ thirty-three year history through 2012.\(^{37}\) Ceteris paribus, this should be interpreted as a good thing, given one has a

\(^{36}\)Although there may be many applications once the measure is subjected to various empirical tests.

\(^{37}\)The average information content of the uncertainty-increasing releases is 0.1529 and the average information content of the uncertainty-decreasing releases is 0.0980.
predisposition to favoring an uncertainty-decreasing information release over the opposite alternative. Over time, Apples’ earnings messages themselves are not neutral regarding uncertainty. That is, they tend to release earnings which decrease shareholders’ uncertainty regarding future earnings. Maybe insight could be obtained by examining the pattern of uncertainty-decreasing and uncertainty-increasing earnings’ releases over time within a firm and across firms. These patterns could provide insight into a given firms’ information environment relative to itself over time and relative to other companies. The patterns might also speak to the relative quality of earnings within a firm across time or across firms.

I have the following idea in mind. Consider the function:

$$q(t) = \frac{\gamma(t)}{\theta(t)}$$

where

$$\gamma(t) = \# \text{ of uncertainty-increasing earnings’ realizations through time period } t$$

$$\theta(t) = \# \text{ of uncertainty-decreasing earnings’ realizations through time period } t$$

An interesting question, is how this function behaves over time for a given firm and across firms. If we plot this function for Apple and Microsoft, interesting picture in Figure 10 emerges. Analyzing Apple, we see that $q(t)$ spikes initially and then fluctuates upwardly through year seventeen (1997) of their history. From year eighteen onward, $q(t)$ follows a generally declining trend. As of now, the proportion of uncertainty-increasing earnings releases relative to uncertainty-decreasing ones is $q_A(2012) = \frac{13}{19} \approx 0.684$. Since 2000, $q(t) < 1$ for all $t$ and thus Apple has been announcing fewer uncertainty-increasing earnings’ “messages” (see Table 1). Microsoft, on the other hand, has experienced a faster increase in $q(t)$ as time has passed.

One interpretation of these observed patterns could be that Microsoft had a better information environment and higher quality earnings from 1985-2002 ceteris paribus. At some time $t^*$ between 2002 and 2003\(^{38}\) the two firms had identical proportions and $q_A(t^*) = q_M(t^*) \approx 0.8$. After this “critical” point, the information environment and earnings quality for Microsoft and Apple respectively has diverged. From the looks of the graph, I would have preferred Microsoft stock to Apple stock until sometime in 2002 when Apples’ stock looks more attractive. This prediction fits the risk

---

\(^{38}\)The time axis of Figure 9 is not depicted at a fine enough level of granularity to enable the user to easily see this.
averse profile of the typical investor. Investors likely prefer companies with a better information environment (less information asymmetry). Thus, given a choice between a company announcing a higher (and generally increasing) ratio of uncertainty-increasing to uncertainty decreasing earnings “messages”, and a company whose same ratio is lower (and generally decreasing), they prefer the more transparent, less risky company.

Based on the previous discussion, and assuming \( q(t) \) adequately captures the strength of the information environment surrounding each company, Figure 10 implies that investors would prefer Microsoft over Apple from 1985-2002 and Apple over Microsoft from 2002-2012. Given earnings “messages” which increase uncertainty regarding future earnings, potential investors shy away from the Apple stock in the former time period in favor of the Microsoft stock. The risk-averse Apple shareholders would also react to the increased uncertainty and seek to sell their shares. They, however, would have difficulty finding a buyer due to reduced demand. Therefore, the price of Apple shares would fall and the value of each shareholders’ investment would diminish. The opposite phenomenon would happen with the Microsoft shareholders during the early time period. These dynamics however would reverse after 2002.

Figure ?? plots the annual return from (1986-2012) for both Microsoft and Apple. Notice how the return of Microsoft is generally greater than Apple from (1986-2002) and then, consistent with the Figure 10 implications, the return drops below Apple and remains there.\(^{39}\)

Figure ?? does not validate the measure of information content introduced in this paper. However, Figure ?? does offer some interesting insights into the potential usefulness of the measures introduced. The consistency of Figure ?? with the predictions implied from Figure 10 provides some evidence in favor of the measure.\(^{40}\)

As a simple test, I also regressed the annual returns of ten, large companies (over their respective lives to-date) on \( q \) with the following OLS, simple linear regression model:

\[
R_{it} = \alpha_0 + \alpha_1 q_{it} + \epsilon_{it} \tag{11}
\]

where \( R_{it} \) is the annual return for firm \( i \) earned over time period \( t \) and \( q_{it} \) is the uncertainty ratio given in equation (21) for firm \( i \) respectively as of the end of time period \( t \).\(^{41}\) Table 2 in Appendix

\(^{39}\)A graph of the annual closing share price for both companies reveals this even clearer.

\(^{40}\)Neither Earnings, EPS nor ROA tell the same story when graphed for both firms. There seems to be “hidden” information in earnings that \( IC(v_{jt}) \) and hence \( q(t) \) capture.

\(^{41}\)Regressing returns on earnings, earnings per share, return on assets or their lagged values respectively for this
B gives the companies used along with the number of observations for each company.

The coefficient on $q_{it}$ was -0.9 (t-stat: -5.31) and the $R^2$ was 9.28%. This is consistent with the reasoning earlier. An increase in $q$ from $t$ to $t+1$ implies that the $t+1$ earnings announcement increased our uncertainty regarding future earnings for firm $i$. Risk-averse shareholders do not like increases in uncertainty and thus seek to sell their shares ceteris paribus. The reduced demand prevents the price from rising much (if at all) and therefore returns fall. Interestingly, the explanatory power of equation (11) for returns is 9.28%.

One should notice at this point that the example given above (i.e. $q(t)$) does not use, in any way, the magnitude of $IC(v_{jt})$. Instead, I simply counted the number of times the measure indicated an increase and decrease in uncertainty respectively. A more direct attempt to validate $IC(v_{jt})$ could be undertaken as follows. Suppose a set of analysts are forecasting earnings, $E_{t+1}$, for period $t+1$ for a given firm. These analysts observe earnings in period $t$ and use this “message” to form their forecasts accordingly. It seems reasonable to assume that if $E_t$ increased their uncertainty regarding $E_{t+1}$ then their subsequent forecasts will be more dispersed than if $E_t$ decreased their uncertainty regarding $E_{t+1}$. This actually is assumed when analyst forecast dispersion is used as a proxy for uncertainty. One could test this assumption with $IC(v_{jt})$ to see if, in fact, the assumption holds.

### 7 Limitations

One limitation surrounding $IC(M)$ as a measure of information content is the fact that I have defined $IC(M) = 0$ when $U(P_I) = U(P_A)$. To illustrate, consider the weather example from section 4.1. Suppose initially $P_I = \{0.9, 0.1\}$. Next, the individual receives a weather forecast $= M$. Upon receiving and processing the knowledge contained in $M$, she updates her prior state-probabilities regarding the weather tomorrow to $P_A = \{0.1, 0.9\}$. It is trivial to check that $U(P_I) = U(P_A)$ and hence, as defined, $IC(M) = 0$. Her overall level of uncertainty regarding which state will be realized has not changed upon receipt of $M$. She is more certain, after $M$, that it will rain tomorrow.

---

42I also regressed returns on lagged $q_{it}$ to see how well the information measure can predict future returns. The coefficient was $-0.39$ (t-stat: -3.01) and the $R^2$ was 3.42%. Also, the correlation between earnings level and $q_{it}$ was -0.06.

43I hope to test this idea and some others in a future paper.

44I revert back to the original notation, without loss of generality, in the interest of simplicity. $M$ is any message at a given point in time. $P_I$ is the individuals’ prior state-probability distribution and $P_A$ is the individuals’ posterior state-probability distribution.
than she was before \( M \). However she is less certain that it will be sunny tomorrow. The increase in certainty regarding the rainy state is exactly offset by the decrease in certainty regarding the sunny state. Thus, quantitatively, her overall uncertainty remains the same. Qualitatively though, one would argue that \( M \) does contain information. \( M \) informs her that she should probably bring an umbrella to the beach and leave her sunscreen at home!

The limitation highlighted above is that \( IC(M) \) captures only the quantitative information content of \( M \).\(^{45}\) There often will be qualitative information contained in \( M \). In this case, \( IC(M) \) does not capture this. This limitation is particularly important in an accounting context if we use \( IC(M) \) to measure the information content of the financial statements. I have been careful to say that \( IC(M) \) captures the information content of the quantitative portions of the financial statements. Other “messages” within the financial statements are the footnotes. It is interesting to think about the problem of quantifying the information content of the qualitative footnotes. Recent advances in textual analysis, including readability measures, and positive/negative tone measures for example, have helped researchers deal with this problem. One could think of \( IC(M) \) as forming a lower bound on the information content of the financial statements. Combining \( IC(M) \) with some of these other measures of footnote information content would provide a more robust measure.

### 8 Conclusion

This paper defines information as that subset of knowledge which changes a users’ uncertainty regarding the state a particular variable of interest will assume in the future. This definition is firmly grounded in an information theory framework. From this definition, I then develop a measure of the information content of a message in general. Shannon entropy, as a measure of uncertainty, is applied to form this measure. Several examples, along the way, illustrate how the measure can be applied.

I also argue that accounting fits within a classical communication system framework. I apply the measure developed to the problem of measuring the information content of “messages” that are transmitted from the accounting and reporting process to interested users. The messages are of interest to users because they contain information regarding future variables of interest.

\(^{45}\) point out that the Shannon entropy function is a measure of uncertainty, “but it is uncertainty when all the information we have consists of just these numbers”.

The Information Content of Accounting Reports: An Information Theory Perspective

(e.g. earnings, cash flows etc.). I modify the general measure to form a specific measure of the information content of accounting “messages” (e.g. earnings, assets, financial statements). I then provide a firm-specific example where the “message” is period $t$ earnings and the variable of interest is period $t+1$ earnings. This example provides some evidence validating the measures’ construct validity.

When applied to an accounting context, the underlying idea is that the quantitative financial statements, in aggregate and at the variable-level, contain information which, at first glance, cannot be observed from the numbers themselves. The measure $IC(v_{jt})$, however, is able to extract this information in a way that is both interesting and insightful.

9 Future Research

This paper only scratches the surface regarding how information theory can be applied to an accounting context. If one accepts the particular way in which I apply information theory, then there are a host of questions surrounding the measures introduced. For example, can the measures predict bankruptcies, earnings restatements or other firm-level calamities? Are the measures consistent with the more traditional earnings-response coefficient information content measure? Do the measures proxy for firm-level risk?46

The measures introduced may also provide an avenue with which to approach deeper, more fundamental questions. For example, what is the optimal reporting period from an investors’ standpoint for a given firm or across firms? In fact, applying the measures introduced to Apple and Microsoft quarterly earnings instead of annual earnings provides earlier evidence that Apple would have been a better investment than Microsoft. This is consistent with the notion that investors would want both companies to report quarterly instead of annually if they had to choose between the two. Of course Microsoft would choose annually over quarterly reporting given the same choice ceteris paribus. This is consistent with a favorable view of the SEC fulfilling its’ duty of protecting investors.

Another fundamental question the measure could render attainable is, at what level of aggregation should the financial statements be presented? If all the information contained in a particular

46Both creditor and shareholder risk. Evidence of a higher cost of capital for those firms with lower information content financial statements, ceteris paribus, may provide evidence that the measures can capture risk.
variable $j$ is already subsumed by some combination of the other variables, then maybe $j$ should not be reported within the statements. The costs of processing the redundant information in $j$ outweigh the benefits!

These questions, particularly the more theoretical ones, are fundamental to understanding accounting as a communication system. Hopefully this paper has made the prospect of providing answers to these questions seem a little more attainable.
References
Appendix A: The Information Content, $IC(F)$, of the Financial Statements

Up to this point I have focused on the information content, $IC(j_t)$, of variable $j$ in time period $t$. Suppose a set of financial statements, $F_t$, is released at time period $t$. Let this set be described as follows:\footnote{I change notation here, by allowing $v_{jt}$ to represent “variable $j_t$”. Note also that $n$ now refers to the number of variables instead of the number of states. The number of states is fixed, by assumption, at four while the number of variables in the financial statements is $n$.}

$$F_t = \{v_{1t}, v_{2t}, \cdots, v_{nt}\}$$ (12)

That is, at time period $t$, we have a set of financial statements comprised of $n$ variables. Now I wish to assess $IC(F_t)$. Simply calculating the sum $\sum_{j=1}^{n} IC(v_{jt})$ does not quite work however since some of the information in any one of the $n$ variables will already be contained in one, or more, of the other variables. Summing in this way will lead to double counting this information.

Figure 8 illustrates what I would like to do when $n = 4$, for example. The shaded regions in Figure 8 begin by shading in the information provided by $v_{1t}$ with blue. Next, I move counterclockwise and shade in (red) the portion of $v_{2t}$ not shaded blue. I then shade in (green) the portion of $v_{3t}$ not already shaded blue or red. Finally, I shade in (purple) the portion of $v_{4t}$ not already shaded blue, red or green. The information content of the financial statements is then equal to the sum of the information contents of each of the colored regions. This process of “sweeping” out the information contents of each financial statement variable one at a time\footnote{Without double counting.} can be accomplished by using the principle of inclusion-exclusion from mathematical set theory. Equation (12) captures the idea for $n = 3$:

$$IC(F_t) = \left( \sum_{j=1}^{3} IC(v_{jt}) \right) - IC(v_{1t}) \cap IC(v_{2t}) - IC(v_{1t}) \cap IC(v_{3t}) - IC(v_{2t}) \cap IC(v_{3t}) + IC(v_{1t}) \cap IC(v_{2t}) \cap IC(v_{3t})$$ (13)

As given in equation (12), $F_t$ is a set consisting of $n$ variables. We can measure the size of the information contained in $F_t$ by considering the sizes of the information (information contents) contained in each of these variables. The problem is, when summing over these sets, we over-estimate by counting intersections more than once. Equation (13) excludes these “over-included” intersections. In the process, we exclude too much and thus need to add back a final term.\footnote{One can readily check, with a Venn diagram, that the above formula is correct.} The formula above is a special case of the general formula, attributed to Abraham De Moivre, an 18th century French mathematician. The general formula for evaluating the size of a given set by considering the sizes of the different, non-mutually exclusive subsets of which it is comprised is
given in equation (14). I express the formula in the context of the present discussion and therefore information content \((IC)\) is the “size of the information”.

\[
IC(F_t) = \sum_{i=1}^{n} IC(v_{it}) - \sum_{i,j:1\leq i<j\leq n} IC(v_{it}) \cap IC(v_{jt}) + \sum_{i,j,k:1\leq i<j<k\leq n} IC(v_{it}) \cap IC(v_{jt}) \cap IC(v_{kt}) - \cdots + (-1)^{n-1} IC(v_{1t}) \cap \cdots \cap IC(v_{nt})
\]

(14)

I admit that equation (14) is cumbersome to follow. It is the classic example of condensing something rather complicated into one precise formula! Nevertheless, one can check their understanding at this point by plugging \(n = 3\) into equation (14); you should recover equation (13). Notice the alternating signs. This helps to ensure that anything we over-include in summing over the information contents of all the variables, we make sure to exclude. Applying equation (14), when \(n = 4\), will precisely give the figure 8 intuitive result.

What is the point of all of the math you ask? The idea is that aggregating is a delicate task which requires caution. We need to ensure that we only “sweep” out information content once, as in Figure 8. Also, equation (14) highlights the importance of defining the intersections. I turn next to this problem.

I will define the intersection of information contents as follows:

\[
IC(v_{it}) \cap \cdots \cap IC(v_{jt}) = \frac{1}{\binom{n}{2}} IC(v_{k^*t}) \sum_{i=I^*}^{n-1} \sum_{j=i+1}^{n} |\rho_{ij}^t| 
\]

(15)

where \(v_{k^*t} = \) variable with smallest information content, \(i \leq k^* \leq j\)

\(I^* = \) starting variable index

\(n = \) ending variable index

\(\rho_{ij}^t = \) Pearson correlation between \(i\) and \(j\) through time \(t \geq 2\)

\(\xi = \#\) of variables intersected

Equation (15) is a general expression for the intersection of information contents between any combination of variables beginning with variable \(i\) through variable \(j\). Equation (15) calculates the information content of variable \(i\) that is already subsumed in the other variables up through variable \(j\). Everything needed in equation (14) to calculate \(IC(F_t)\), given the financial statement variables, is provided by equation (15). Precisely, given variables \(v_{it}, \ldots, v_{jt}\), equation (15) is the average of all possible pair-wise Pearson correlations multiplied by the smallest information content. This is consistent with the notion that the information content of the intersection can be no greater than the smallest information content. For example, suppose you are given \(IC(v_{1t}) = 0.2, IC(v_{2t}) = 0.3, IC(v_{3t}) = 0.1\) and \(\rho_{12} = 0.6, \rho_{13} = 0.9\) and \(\rho_{23} = 0.95\). The information content common to all three variables is \(0.1 * \frac{1}{\binom{3}{2}} * (0.6 + 0.9 + 0.95) \approx 0.08167\). Although this is likely an estimate, it is intuitive with the notion that the information content common to all three variables is an increasing
function of the correlation between the variables. The absolute value sign prevents the meaningless case of negative joint information. Although any two variables could be negatively correlated, knowing this is just as informative as knowing that they are positively correlated with each other! In other words, I don’t lose anything by disregarding the sign of the correlation between the two variables.

One will recall from the Apple Inc. example that \( IC(v_{jt}) \) can be signed (superscripted) to indicate whether uncertainty increased or decreased. When aggregating the information contents, however, I disregard whether \( v_{jt} \) increases or decreases the individuals’ uncertainty. I assume that a realization of \( v_{jt} \) which increases our uncertainty regarding future realizations of \( v_j \) is equally as informative as a realization of \( v_{jt} \) which decreases our uncertainty regarding future realizations of \( v_j \). Therefore, theoretically, \( 0 \leq IC(F_t) \leq n \).\(^{50}\) Calculating \( IC(F_t) = 4 \), for example, where \( F_t \) consists of ten variables, implies that the information content of \( F_t \) is 40% of the maximum amount it could have theoretically been.\(^{51,52}\)

A different way to express the information content of the financial statements, that may be more meaningful, is to divide \( IC(F_t) \) by the sum of the individual variable information contents. If there is no information common to any of the variables then \( IC(F_t) \) would equal this sum and we could say that the financial statements contained maximum information relative to the information contents of each of the variables therein. Thus a more meaningful measure of financial statement information content is given below:

\[
IC^*(F_t) = \frac{IC(F_t)}{\sum_{i=1}^{n} IC(v_{it})}
\]

(16)

The above discussion implies that \( IC(F_t) \) is maximized, theoretically, when each of the financial statement variables is independent of the others. If this is the case, the numerator and denominator of equation (16) are equal and \( IC^*(F_t) = 1 \). The equivalent of this pictorially, is where the circles in Figure 8 do not intersect each other. The converse of this occurs when each of the \( n \) variables is perfectly correlated with each of the others. If this is the case, \( IC(F_t) = IC(v_{k+t}) \) and \( IC^*(F_t) = IC(v_{k+t}) / \sum_{i=1}^{n} IC(v_{it}) \). Pictorially, this implies that the circles in Figure ?? are perfectly superimposed over each other such that the information in the smallest information content variable is the most the user can obtain from the statements.

Next, I provide an empirical example to help illustrate some of the potential applications of \( IC(v_{jt}) \). In Appendix B I continue with the Apple Inc. example to illustrate \( IC^*(F_t) \) and \( IC^{**}(F_t) \).\(^{53}\)

\(^{50}\)Recall that \( 0 \leq IC(v_{jt}) \leq 1 \) for all \( j \).

\(^{51}\)60% shy of “perfect” information. Perfect in the sense that all of the information contained in each variable \( j \) is unique to \( j \) and is equal to 1.

\(^{52}\)I call this percentage \( IC^{**}(F_t) \) in Appendix B.

\(^{53}\)In Appendix B I deal with the problem of signing \( IC(F_t) \). See Equations (17) and (??).
Appendix B: The Information Content of Apple’s Financial Statements

To illustrate the calculation of $IC(F_t)$ for a given company, consider Apple Inc. over the time period (1980-1990). Without loss of generality, suppose Apple’s financial statements, for each of these years, consisted of only four variables: earnings, total revenue, total assets and total liabilities. Thus, we consider the following set:

$$F_t = \{\text{earnings}_t, \text{total revenue}_t, \text{assets}_t, \text{liabilities}_t\} = \{v_{1t}, v_{2t}, v_{3t}, v_{4t}\}$$

First, calculate the correlations $\rho_{ij}$ between each pair of variables through each time period $t \geq 2$. Thus, for each time period, a vector of six correlations is produced. For example, through $t = 5$, the correlation between earnings and total revenue is $\rho_{12}^5 = 0.7726$. These correlations are listed in Table 3. Next, calculate the information contents of each of the four variables over the (1980-1990) time frame. These are displayed in Table 4. Now apply equations (14) and (15) to calculate the information content of the financial statements. For example, to illustrate $IC(F_8)$ you get the following:

54Variables $NI$, $REVT$, $TA$ and $TL$ respectively from Compustat.
\[ IC(F_8) = IC(v_{18}) + IC(v_{28}) + IC(v_{38}) + IC(v_{48}) \]
\[ - IC(v_{18}) \cap IC(v_{28}) - IC(v_{18}) \cap IC(v_{38}) - IC(v_{18}) \cap IC(v_{48}) \]
\[ - IC(v_{28}) \cap IC(v_{38}) - IC(v_{28}) \cap IC(v_{48}) - IC(v_{28}) \cap IC(v_{38}) \cap IC(v_{48}) \]
\[ + IC(v_{18}) \cap IC(v_{28}) \cap IC(v_{38}) + IC(v_{18}) \cap IC(v_{28}) \cap IC(v_{48}) \]
\[ + IC(v_{18}) \cap IC(v_{38}) \cap IC(v_{48}) + IC(v_{28}) \cap IC(v_{38}) \cap IC(v_{48}) \]
\[ - IC(v_{18}) \cap IC(v_{28}) \cap IC(v_{38}) \cap IC(v_{48}) \]
\[ = 0.0762 + 0.4142 + 0.0432 + 0.0432 \]
\[ - (0.0762 \times 0.8447) - (0.0432 \times 0.9024) - (0.0432 \times 0.8911) \]
\[ - (0.0432 \times 0.9896) - (0.0432 \times 0.9880) - (0.0432 \times 0.9931) \]
\[ - \frac{0.0432}{3} \times (0.8447 + 0.9024 + 0.8911) \]
\[ - \frac{0.0432}{3} \times (0.8447 + 0.8911 + 0.9880) \]
\[ - \frac{0.0432}{3} \times (0.9024 + 0.8911 + 0.9931) \]
\[ - \frac{0.0432}{6} \times (0.8447 + 0.9024 + 0.8911 + 0.9896 + 0.9880 + 0.9931) \]
\[ \approx 0.4278 \]

Now use equation (16) to express \( IC(F_8) \) in percentage terms and derive \( IC^*(F_8) \):

\[ IC^*(F_8) = \frac{0.4278}{0.0762 + 0.4142 + 0.0432 + 0.0432} \]
\[ \approx 0.7417 \text{ or 74\%} \]

Thus the information content of Apples’ 1987 financial statements was 74\% of the maximum it could have been, given the information contained in the variables of which the statements were comprised. That is, given the information contents of each of the variables, if the variables were all independent, the maximum information content would have been \( \sum_{i=1}^{4} IC(v_{it}) = 0.5768 \).

However, given the additional constraint that the variables not only be independent but also have maximal information themselves,\(^{55}\) the information content of the 1987 financial statements was \( IC^{**}(F_t) = IC(F_t) / n = 0.4278 / 4 \approx 0.1070 \). Thus the 1987 statements changed our uncertainty regarding future financial statements 10.7\% of the maximum amount it possibly could have. Note in Table 4 that this was an uncertainty-increasing change.

\(^{55}\)i.e. \( IC(v_{it}) = 1 \) for all \( i \).
Notice that I have not signed $IC(F_t)$. To accomplish this, consider the proportion of variables that increased uncertainty to those that decreased uncertainty. Call this function $p(t)$:

$$p(t) = \frac{\phi(t)}{\psi(t)}$$

where

$\phi(t) = \# $ of uncertainty-increasing variables at time $t$

$\psi(t) = \# $ of uncertainty-decreasing variables at time $t$

Thus the information content of the financial statements at time $t$ can be signed as follows:

$$IC(F_t) = \begin{cases} 
IC(F_t)^+ & \text{if } p(t) > 1 \\
IC(F_t)^- & \text{if } p(t) < 1 
\end{cases}$$

(17)

Now if $p(t) = 1$ then half of the variables increased our uncertainty regarding future realizations of those variables and half of the variables decreased our uncertainty. In this case, we consider the magnitude of the information contents of the uncertainty-increasing and uncertainty-decreasing variables. Define a new function as follows:

$$z(t) = \omega(t) - \zeta(t)$$

where

$\omega(t) = \text{mean information content of uncertainty-inc. variables at time } t$

$\zeta(t) = \text{mean information content of uncertainty-dec. variables at time } t$

The information content of the financial statements in this case is defined as follows:

$$IC(F_t) = \begin{cases} 
IC(F_t)^+ & \text{if } z(t) > 0 \\
IC(F_t)^- & \text{if } z(t) \leq 0 
\end{cases} \quad \text{if } p(t) = 1$$

(18)

Table 5 reports $IC(F_t), IC^*(F_t)$ and $IC^{**}(F_t)$ for Apple from 1980-1990.
Figures

**Figure 1**
Accounting as a Communication System

This figure displays how accounting fits into the classical framework of a communication system.

**Figure 2**
Information Conceptualized

This figure illustrates how information relates to initial and ex post state probabilities and time.
Figure 3
Information and the Change in Uncertainty

This figure illustrates how information increases with the change in uncertainty.

![Figure 3 Diagram](image1)

Figure 4
Uncertainty of Compound State-Probability Distribution

This figure is a probability tree of a compound state-probability distribution.

![Figure 4 Diagram](image2)
This figure illustrates how uncertainty increases relative to the maximum amount it could have increased by.

\[ U(P_I) \quad U(P_a) \quad \max U(P) \]

This figure illustrates how uncertainty decreases relative to the maximum amount it could have decreased by.

\[ U(P_a) \quad U(P_I) \quad \max U(P) \]
This figure depicts the earnings state mapping function, $f(j_t)$ from equation (6).

$$f: \ j_t \rightarrow S = \{L, \ Lm, \ Hm, \ H\}$$

<table>
<thead>
<tr>
<th>L</th>
<th>Lm</th>
<th>Hm</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_j - \sqrt{2} \sigma_j$</td>
<td>$\mu_j$</td>
<td>$\mu_j + \sqrt{2} \sigma_j$</td>
<td></td>
</tr>
</tbody>
</table>

This figure illustrates the non-overlapping information content of the financial statements reporting on four variables only.
Figure 9
Uncertainty Ratio for Apple and Microsoft Over Time

This figure plots the uncertainty ratio in equation (10) for Apple and Microsoft from their inception as a company through 2012.

Figure 10
Annual Return for Apple and Microsoft Over Time

This figure plots the annual share price return for Apple and Microsoft from their inception as a company through 2012.
### Tables

#### Table 1
Information Content of Apple Inc. Earnings (1980-2012)

This table reports the information content of each of Apple’s thirty-three earnings releases from its inception as a company through 2012 using the information content measure from equation (9).

<table>
<thead>
<tr>
<th>Year</th>
<th>$j_t$ (millions)</th>
<th>$P_{jt}$</th>
<th>$P_{At}$</th>
<th>IC($j_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>$11.70$</td>
<td>{1/4,1/4,1/4,1/4}</td>
<td>{1/4,1/4,1/4,1/4}</td>
<td>$0$</td>
</tr>
<tr>
<td>1981</td>
<td>$39.42$</td>
<td>{1/4,1/4,1/4,1/4}</td>
<td>{0,1/2,1/2,0}</td>
<td>$0.5^-$</td>
</tr>
<tr>
<td>1982</td>
<td>$61.306$</td>
<td>{0,1/2,1/2,0}</td>
<td>{0,1/3,2/3,0}</td>
<td>$0.0817^-$</td>
</tr>
<tr>
<td>1983</td>
<td>$76.714$</td>
<td>{0,1/3,2/3,0}</td>
<td>{0,1/2,1/2,0}</td>
<td>$0.0755^+$</td>
</tr>
<tr>
<td>1984</td>
<td>$64.055$</td>
<td>{0,1/2,1/2,0}</td>
<td>{1/5,1/5,3/5,0}</td>
<td>$0.3710^+$</td>
</tr>
<tr>
<td>1985</td>
<td>$61.223$</td>
<td>{1/5,1/5,3/5,0}</td>
<td>{1/6,1/6,2/3,0}</td>
<td>$0.0870^-$</td>
</tr>
<tr>
<td>1986</td>
<td>$153.963$</td>
<td>{1/6,1/6,2/3,0}</td>
<td>{0,5/7,1/7,1/7}</td>
<td>$0.0821^-$</td>
</tr>
<tr>
<td>1987</td>
<td>$217.496$</td>
<td>{0,5/7,1/7,1/7}</td>
<td>{0,3/4,1/8,1/8}</td>
<td>$0.0762^-$</td>
</tr>
<tr>
<td>1988</td>
<td>$400.258$</td>
<td>{0,3/4,1/8,1/8}</td>
<td>{0,2/3,2/9,1/9}</td>
<td>$0.1738^+$</td>
</tr>
<tr>
<td>1989</td>
<td>$454.033$</td>
<td>{0,2/3,2/9,1/9}</td>
<td>{0,7/10,1/10,1/5}</td>
<td>$0.0552^-$</td>
</tr>
<tr>
<td>1990</td>
<td>$474.895$</td>
<td>{0,7/10,1/10,1/5}</td>
<td>{0,7/11,2/11,2/11}</td>
<td>$0.1809^+$</td>
</tr>
<tr>
<td>1991</td>
<td>$309.841$</td>
<td>{0,7/11,2/11,2/11}</td>
<td>{0,7/12,1/4,1/6}</td>
<td>$0.1088^+$</td>
</tr>
<tr>
<td>1992</td>
<td>$530.373$</td>
<td>{0,7/12,1/4,1/6}</td>
<td>{0,8/13,4/13,1/13}</td>
<td>$0.1051^-$</td>
</tr>
<tr>
<td>1993</td>
<td>$86.589$</td>
<td>{0,8/13,4/13,1/13}</td>
<td>{0,4/7,2/7,1/7}</td>
<td>$0.1838^+$</td>
</tr>
<tr>
<td>1994</td>
<td>$310.178$</td>
<td>{0,4/7,2/7,1/7}</td>
<td>{0,8/15,1/3,2/15}</td>
<td>$0.0335^+$</td>
</tr>
<tr>
<td>1995</td>
<td>$424$</td>
<td>{0,8/15,1/3,2/15}</td>
<td>{0,9/16,3/8,1/16}</td>
<td>$0.1086^-$</td>
</tr>
<tr>
<td>1996</td>
<td>$(816)$</td>
<td>{0,9/16,3/8,1/16}</td>
<td>{1/17,8/17,8/17,0}</td>
<td>$0.0218^+$</td>
</tr>
<tr>
<td>1997</td>
<td>$(1045)$</td>
<td>{1/17,8/17,8/17,0}</td>
<td>{1/9,7/18,1/2,0}</td>
<td>$0.1605^+$</td>
</tr>
<tr>
<td>1998</td>
<td>$309$</td>
<td>{1/9,7/18,1/2,0}</td>
<td>{2/19,7/19,10/19,0}</td>
<td>$0.0160^-$</td>
</tr>
<tr>
<td>1999</td>
<td>$601$</td>
<td>{2/19,7/19,10/19,0}</td>
<td>{1/10,7/20,11/20,0}</td>
<td>$0.0172^-$</td>
</tr>
<tr>
<td>2000</td>
<td>$786$</td>
<td>{1/10,7/20,11/20,0}</td>
<td>{2/21,8/21,10/21,1/21}</td>
<td>$0.3553^+$</td>
</tr>
<tr>
<td>2001</td>
<td>$(25)$</td>
<td>{2/21,8/21,10/21,1/21}</td>
<td>{1/11,9/22,5/11,1/22}</td>
<td>$0.0067^-$</td>
</tr>
<tr>
<td>2002</td>
<td>$65$</td>
<td>{1/11,9/22,5/11,1/22}</td>
<td>{2/23,10/23,10/23,1/23}</td>
<td>$0.0088^-$</td>
</tr>
<tr>
<td>2003</td>
<td>$69$</td>
<td>{2/23,10/23,10/23,1/23}</td>
<td>{1/12,5/12,11/24,1/24}</td>
<td>$0.0104^-$</td>
</tr>
<tr>
<td>2004</td>
<td>$276$</td>
<td>{1/12,5/12,11/24,1/24}</td>
<td>{2/25,11/25,11/25,1/25}</td>
<td>$0.0081^-$</td>
</tr>
<tr>
<td>2005</td>
<td>$1335$</td>
<td>{2/25,11/25,11/25,1/25}</td>
<td>{1/13,11/26,6/13,1/26}</td>
<td>$0.0094^-$</td>
</tr>
<tr>
<td>2006</td>
<td>$1989$</td>
<td>{1/13,11/26,6/13,1/26}</td>
<td>{2/27,4/9,11/27,2/27}</td>
<td>$0.1995^+$</td>
</tr>
<tr>
<td>2007</td>
<td>$3496$</td>
<td>{2/27,4/9,11/27,2/27}</td>
<td>{1/14,4/7,2/7,1/14}</td>
<td>$0.0514^-$</td>
</tr>
<tr>
<td>2008</td>
<td>$4834$</td>
<td>{1/14,4/7,2/7,1/14}</td>
<td>{0,23/29,4/29,2/29}</td>
<td>$0.3918^-$</td>
</tr>
<tr>
<td>2009</td>
<td>$8235$</td>
<td>{0,23/29,4/29,2/29}</td>
<td>{0,5/6,1/5,1/10}</td>
<td>$0.1228^-$</td>
</tr>
<tr>
<td>2010</td>
<td>$14,013$</td>
<td>{0,5/6,1/5,1/10}</td>
<td>{0,25/31,4/31,2/31}</td>
<td>$0.0629^+$</td>
</tr>
<tr>
<td>2011</td>
<td>$25,922$</td>
<td>{0,25/31,4/31,2/31}</td>
<td>{0,27/32,3/32,1/16}</td>
<td>$0.1236^-$</td>
</tr>
<tr>
<td>2012</td>
<td>$41,733$</td>
<td>{0,27/32,3/32,1/16}</td>
<td>{0,9/11,4/33,2/33}</td>
<td>$0.0605^+$</td>
</tr>
</tbody>
</table>
Table 2
Annual Returns as a Function of the Uncertainty Ratio

This table reports the companies used in the equation (11) test relating annual returns to the annual uncertainty ratio, $q_{it}$, given in equation (10). The companies were chosen from the S&P 100 arranged in alphabetical order by company name. If a particular company did not have the required data to calculate the variables of interest (over its respective life from inception through 2012) the next company in the alphabetically arranged S&P 100 was chosen.

<table>
<thead>
<tr>
<th>COMPANY</th>
<th># of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>15</td>
</tr>
<tr>
<td>Amgen</td>
<td>28</td>
</tr>
<tr>
<td>Apple</td>
<td>31</td>
</tr>
<tr>
<td>Dell</td>
<td>24</td>
</tr>
<tr>
<td>Fed Ex</td>
<td>33</td>
</tr>
<tr>
<td>Home Depot</td>
<td>31</td>
</tr>
<tr>
<td>Microsoft</td>
<td>26</td>
</tr>
<tr>
<td>Nike</td>
<td>31</td>
</tr>
<tr>
<td>Starbucks</td>
<td>20</td>
</tr>
<tr>
<td>Walmart</td>
<td>39</td>
</tr>
</tbody>
</table>

Total Firm-Year Obs. 278

Table 3
Correlations Among Apple’s Financial Statement Variables

This table reports the correlations of Apple’s financial statement variables (earnings, total revenue, total assets and total liabilities) in the Appendix B example.

| Year | $t$ | $|\rho_{12}^t|$ | $|\rho_{13}^t|$ | $|\rho_{14}^t|$ | $|\rho_{23}^t|$ | $|\rho_{24}^t|$ | $|\rho_{34}^t|$ |
|------|-----|-------------|-------------|-------------|-------------|-------------|-------------|
| 1981 | 2   | 1.0000      | 1.0000      | 1.0000      | 1.0000      | 1.0000      | 1.0000      |
| 1982 | 3   | 0.9944      | 0.9949      | 0.9975      | 0.9786      | 0.9845      | 0.9995      |
| 1983 | 4   | 0.9632      | 0.9834      | 0.9358      | 0.9891      | 0.9922      | 0.9822      |
| 1984 | 5   | 0.7726      | 0.8250      | 0.6654      | 0.9933      | 0.9867      | 0.9702      |
| 1985 | 6   | 0.6888      | 0.7332      | 0.5973      | 0.9942      | 0.9932      | 0.9823      |
| 1986 | 7   | 0.7222      | 0.8335      | 0.7804      | 0.9825      | 0.9827      | 0.9895      |
| 1987 | 8   | 0.8447      | 0.9024      | 0.8911      | 0.9896      | 0.9880      | 0.9931      |
| 1988 | 9   | 0.9296      | 0.9376      | 0.9612      | 0.9941      | 0.9917      | 0.9885      |
| 1989 | 10  | 0.9623      | 0.9665      | 0.9786      | 0.9970      | 0.9943      | 0.9928      |
| 1990 | 11  | 0.9737      | 0.9762      | 0.9810      | 0.9978      | 0.9935      | 0.9940      |
Table 4
Information Content of Apple’s Financial Statement Variables

This table reports the information content of Apple’s financial statement variables (earnings, total revenue, total assets and total liabilities) in the Appendix B example using equation (15).

<table>
<thead>
<tr>
<th>Year</th>
<th>t</th>
<th>( IC(v_{1t}) )</th>
<th>( IC(v_{2t}) )</th>
<th>( IC(v_{3t}) )</th>
<th>( IC(v_{4t}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>1</td>
<td>0.0000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1981</td>
<td>2</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>1982</td>
<td>3</td>
<td>0.0817</td>
<td>0.0817</td>
<td>0.0817</td>
<td>0.0817</td>
</tr>
<tr>
<td>1983</td>
<td>4</td>
<td>0.0755</td>
<td>0.0755</td>
<td>0.0755</td>
<td>0.0755</td>
</tr>
<tr>
<td>1984</td>
<td>5</td>
<td>0.3709</td>
<td>0.3709</td>
<td>0.0290</td>
<td>0.3709</td>
</tr>
<tr>
<td>1985</td>
<td>6</td>
<td>0.0870</td>
<td>0.1402</td>
<td>0.0282</td>
<td>0.0870</td>
</tr>
<tr>
<td>1986</td>
<td>7</td>
<td>0.0821</td>
<td>0.3247</td>
<td>0.3787</td>
<td>0.1699</td>
</tr>
<tr>
<td>1987</td>
<td>8</td>
<td>0.0762</td>
<td>0.4142</td>
<td>0.0432</td>
<td>0.0432</td>
</tr>
<tr>
<td>1988</td>
<td>9</td>
<td>0.1737</td>
<td>0.0384</td>
<td>0.0384</td>
<td>0.0384</td>
</tr>
<tr>
<td>1989</td>
<td>10</td>
<td>0.0552</td>
<td>0.1441</td>
<td>0.0415</td>
<td>0.0297</td>
</tr>
<tr>
<td>1990</td>
<td>11</td>
<td>0.1808</td>
<td>0.1808</td>
<td>0.0196</td>
<td>0.0950</td>
</tr>
</tbody>
</table>

Table 5
Information Content of Apple’s Financial Statements (1980-1990)

This table reports the information content of Apple’s financial statements using each the measures from equations (15) and (16) respectively.

<table>
<thead>
<tr>
<th>Year</th>
<th>t</th>
<th>( IC(F_t) )</th>
<th>( IC^*(F_t) )</th>
<th>( IC^{**}(F_t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1981</td>
<td>2</td>
<td>0.5000</td>
<td>0.2500</td>
<td>0.1250</td>
</tr>
<tr>
<td>1982</td>
<td>3</td>
<td>0.0837</td>
<td>0.2563</td>
<td>0.0209</td>
</tr>
<tr>
<td>1983</td>
<td>4</td>
<td>0.0813</td>
<td>0.2692</td>
<td>0.0203</td>
</tr>
<tr>
<td>1984</td>
<td>5</td>
<td>0.5135</td>
<td>0.4497</td>
<td>0.1283</td>
</tr>
<tr>
<td>1985</td>
<td>6</td>
<td>0.1837</td>
<td>0.5367</td>
<td>0.0459</td>
</tr>
<tr>
<td>1986</td>
<td>7</td>
<td>0.4132</td>
<td>0.4324</td>
<td>0.1033</td>
</tr>
<tr>
<td>1987</td>
<td>8</td>
<td>0.4278</td>
<td>0.7416</td>
<td>0.1069</td>
</tr>
<tr>
<td>1988</td>
<td>9</td>
<td>0.1775</td>
<td>0.6144</td>
<td>0.0443</td>
</tr>
<tr>
<td>1989</td>
<td>10</td>
<td>0.1469</td>
<td>0.5425</td>
<td>0.0367</td>
</tr>
<tr>
<td>1990</td>
<td>11</td>
<td>0.1865</td>
<td>0.3916</td>
<td>0.0466</td>
</tr>
</tbody>
</table>