INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

The Usefulness of Accounting and Non-Financial Information in Explaining Revenues and Valuations for Internet Firms

by

Anthony R. Kozberg

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN BUSINESS ADMINISTRATION
DEPARTMENT OF ACCOUNTING
THE LEONARD N. STERN SCHOOL OF BUSINESS
NEW YORK UNIVERSITY
SEPTEMBER, 2001

Committee:
Stephen G. Ryan
Joshua Livnat
James Ohlson

UMI Number: 3028671

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

I, Anthony Kozberg,

hereby swear and affirm that no part of the dissertation which I have submitted for publication has been heretofore published, and/or copyrighted in the United States of America, except in the case of passages quoted from other published sources; that I am the sole author and proprietor of said dissertation; that the dissertation contains no matter which, if published, will be libelous or otherwise injurious, or infringe in any way , the copyright of any other party; and that I will defend, indemnify and hold harmless New York University against all suits and proceedings which may be brought, and against all claims which may be made against New York University by reason of the publication of said dissertation.

Anthony Kozberg

Sworn and subscribed to before me this 21st

Day of Aug. 2001

DEBRA A. CUNNINGHAM
Notary Public, State of New York
No. 01CU5004212
Qualified In New York County
Commission Expires Nov. 16, 2002

Table of Contents

Chapter 1: The Value Drivers of Internet Stocks: A Business Models Approach

1.1 Introduction

Over the past five years, the technology-laden NASDAQ index has experienced unprecedented price volatility, largely attributable to internet stocks. Because internet firms have often lacked positive net income and had market values that greatly exceed revenues, valuation methods for these firms have necessarily been ad hoc. Moreover, relatively little progress has been made into the difficult problem of how to incorporate the various amorphous business models they employ into their analysis. These models have often been defined by the source of the firm's current or potential revenues, such as advertising, sponsorship, sales, subscription services and licensing. Alternatively, definitions that loosely reflect the target markets for these firms, such as B to B, B to C, and C to C, have been used.[1] Such approaches, however, fail to reflect the dynamic and varied nature of internet firms.

This dissertation contributes to the literature on firm valuations using non-financial measures, focusing on the internet, in three ways. First, it highlights the importance of distinguishing the business models employed by internet firms in determining their value-drivers. Second, it develops a conceptual framework which is used to create a more comprehensive set of non-financial value drivers and employs a number of these in empirical tests (e.g., the percentage of the internet audience reached and the number of pageviews and advertisements shown to those audience members). Finally, it constructs a larger sample of firms over a longer time horizon enabling a more detailed examination of changes in the pricing of financial and non-financial

[1] B to B refers to companies who focus on clients whose business is selling to other companies. B to C firms target end consumers and C to C firms focus on individuals who want to interact with others having similar interests.

variables during boom (through February, 2000) and bust (since) periods in the market for internet stocks.

The first contribution is an examination of the importance of identifying the business models employed by internet firms. Despite the prevailing view that it represents a "new economy," the internet is, at its core, a technology serving as a point of convergence for a number of different traditional and non-traditional industries such as media, telecommunications, hardware, software, retailing, and consulting among others. Failure to appreciate the characteristic differences across these industries introduces noise and/or potential biases into empirical analysis. To address this problem, I classify my sample firms into seven groups based upon the principal business model used: portals, content-community, e-tailers, financial services, enablers, ISP/Infrastructure and non-sensitive firms.[2]

These business models provide a richer definition of firm types and reflect distinct operating characteristics such as the type of products sold (e.g., information, software or a tangible good), the types of customers targeted and the relative level of importance of internet activity to their websites and those of their customers. While firms can and do change particular aspects of their operations over time, these models should avoid mistaking small differences across firms in their target market (i.e., a change from selling to consumers to selling to businesses) with larger operational changes in the underlying product or services offered. In

[2] Portals are designed to be gateways to the Internet. Most feature news, information organized by category, and search capabilities. Content-community firms are organized around specific content (sports, politics, stocks, etc.) and personal or professional interests. E-tailers sell products online, to consumers, business, or both. Financial Services firms include online stockbrokers, loan processors, credit card providers, banks, and venture capital companies. Enablers provide software that enables other firms or individuals to conduct business or entertainment activities. ISPs/Infrastructure (I) firms provide Internet access to computers, corporate clients (VPNs), wireless devices, etc. This group also includes firms which try to improve the performance of the Internet (e.g., cable access providers, caching server vendors, and router and switch makers). Non-Sensitives firms are, ex ante, not expected to have any dependence to the amount of activity their websites generate. These companies include those that develop security or performance software and consultants/designers. A more detailed description of these firms is provided in section 3.

contrast, prior papers have either analyzed internet firms as if they were one large homogenous sample (Hand, 2000(a,b), Rajgopal, et. al., 2000 and Demers and Lev, 2001) or have restricted their studies to firms that are *ex ante* believed to have the greatest sensitivity to the non-financial measures studied (Trueman, Wong and Zhang, 2001(a) and Demers and Lev, 2001).

Empirical results indicate that there are noticeable differences in both the mean and median levels of financial statement and non-accounting data and in the percentage of firms with reported non-financial information across the various business models. As an example of the differences in the levels of financial and non-financial variables, portal and content-community firms have from half to an order of magnitude size difference for most of the variables employed in this study. Reported internet activity levels vary from 5% for firms in the "non-sensitive" business model classification to 89% for portals.

Empirical results show a number of differences in the information content of financial and non-financial variables across the business models. (1) Most web-usage variables are positively and significantly associated with firm valuations for portals and e-tailers. Results across other business models show sensitivity to only a few (content-community) or none of the variables examined (financial services firms) over the full time period studied. This is despite a relatively high occurrence of reported web activity for these four business models (greater than 40% of each sub-sample). For sub-samples with lower occurrences of reported activity (enablers, ISP/Infrastructure and non-sensitives), there is little evidence to suggest that these measures have any value relevance. (2) The significance of the accounting variables differs in predictable ways across the seven business models. For example, while research and development (R&D) expenses have been shown to have limited explanatory power in prior valuation studies (e.g., Demers and Lev, 2001 and Trueman et. al., 2001(a)), it provides a

noticeable improvement in R^2s for particular business models. Specifically, R&D is positively and significantly related to valuations for ISP/Infrastructure, portal, and content-community firms. Additionally, portals and ISP/Infrastructure firms appear to be valued more like traditional firms with earnings positively priced.

The second contribution of this paper is an examination of the path from expenditures on SG&A and R&D through non-financial measures to revenue generation. From this, it is possible to develop a more complete set of non-financial variables. In addition to measures of total audience, pageviews, visits, and time spent online that have been examined in prior research, this paper examines the informativeness of the number of advertisements shown on a firm's web property and the number of times those advertisements have been "clicked-through" by its visitors. For the overall sample, these previously unexamined measures are significant in explaining firm valuations both when regressed individually (with the accounting data) and incrementally significant when combined with other internet activity data. For specific business models, advertisements per person show a positive and significant coefficient for portals, e-tailers and, to a lesser extent, content-community business models. This result reflects the relative importance that advertising plays in the revenue streams for these model types and provides evidence of the usefulness of isolating model-specific variables in the valuation of internet firms.

Within the last year, firms involved in the internet have seen most, or all, of their stock gains from the late 1990's evaporate. A number of questions have since arisen over the continuing relevance of non-financial information in this later period and whether or not investors have come to appreciate the importance of accounting fundamentals for these stocks. The third major contribution of this paper is the development of a more extensive database

(through the first quarter of 2001) from which it examines the question of how the pricing of internet stocks has changed from the boom (through February 2000) to bust (starting March 2000) periods in the markets for these stocks. The positive pricing of earnings before taxes for the entire internet sample appears to be driven by observations in the bust period. The coefficients on earnings for ISP/Infrastructure and portal firms are robust to the different time periods. Results indicate that the negative pricing of earnings observed by Hand (2000a,b) and others would appear to be isolated to online retailing firms and those which develop enabling technology for other companies to conduct business on the internet in the pre-crash period. Disaggregating investments into SG&A and R&D from earnings provides further evidence that accounting fundamentals have become increasingly relevant in the later time period with earnings (SG&A) positive and significant for 5 (6) of the 7 models. Despite the increasing relevance in accounting data in the later time period, the previously value-relevant non-financial measures generally continue to be significantly priced in the post-crash period as well. Previously, only Demers and Lev (2001) has provided any tests of changes in information content over time, covering the period just after the initial market crash.

The remainder of this paper is divided into seven sections. Section 2 discusses the existing internet valuation literature. Section 3 describes the business models used in this paper. Section 4 details the data collection process. The empirical results for the full and business model partitioned samples are presented in Sections 5 and 6, respectively. Section 7 summarizes the findings of this study and provides suggestions for future testing.

1.2 Literature Review

An example of early work on the usage of non-financial report data for valuation of dynamic industries includes Amir and Lev's (1996) examination of the wireless telephone industry. Amir and Lev showed that information about current market penetration and the number of potential wireless subscribers, in conjunction with earnings and book value, is value relevant. Since then, a number of other papers have emerged examining the value relevance of non-financial information such as patents (Deng, 1999), trademarks (Seethamraju, 2000), brand valuations (Barth et. al., 1999), and customer satisfaction (Ittner and Larker, 1998) in various industries. Similar to Amir and Lev, these papers focus on measures that attempt to explain existing or future economic opportunities of the firms in question, borrowing from economic concepts such as reputation effects and barriers to entry.

With its (initially) low cost of entry (a web address and hosting of simple homepage can cost as little as $20 a year) and the potential for scale economies, millions of websites have appeared over the last 5 years and several hundred publicly traded firms have come into existence. These firms have often reported negative or slightly positive earnings for their entire life, making it difficult to value them using accounting data alone. The academic literature has just begun to examine the value-relevance of either type of information in internet stock valuations. I briefly discuss five prominent papers below. Salient features and results of these papers are summarized in Table 1.

Hand (2000a) finds a positive relationship between log market values and log accounting data for internet firms with positive core net income (CNI). For negative CNI firms, however, the coefficient is negative and largely attributable to the market's positive pricing of marketing and R&D expenses. Hand (2000b) further incorporates audience measurement data, including

demographic data, and three supply and demand variables for the firms' stock. Results suggest that forecasted earnings and book value contain more explanatory power than either internet traffic or supply and demand variables. Valuation is marginally related to unique visitors, although not to either pageviews or total hours spent online at the firm's web property. Results for supply and demand variables are consistent with prior expectations that higher market value firms are more likely to be shorted, have a smaller share float and greater levels of institutional ownership. For firms without reported web traffic, the coefficients on forecasted earnings and supply and demand appear to be greater. Demographic data does not appear to be priced.

Trueman, Wong and Zhang (TWZ, 2001a) find unique users and pageviews are positively, and net income is negatively, associated with market values. When they partition the sample into two business models, e-tailing and portal and content (P&C), results indicate a negative association between net income and market value for the former and a significantly positive association for the latter. Pageviews are slightly more relevant than visitors (based on R^2) for P&C firms but far more relevant for e-tailers. Using earnings components, the positive coefficient on gross margin holds for both while sales and marketing expense are significantly *negative* for P&C firms. TWZ (2001a) attributes these results to P&C firms being more like offline firms (e.g., more periodic expenses) than do e-tailers. While true in some respects, this explanation overlooks the fact that e-tailers must still deal with issues such as product procurement and fulfillment while P&C firms can be almost completely information-based.

Rajgopal, Kotha and Venkatachalam (RKV, 2000) is one of the first papers to discuss the possibility of a network effect ("critical mass") for internet firms which, if achieved, can later be converted into revenue. Depending on the business model, however, increased activity at a site will not always result in a better experience for users as it could lead to greater lags and difficulty

in locating desired content. RKV (2000)'s results are mixed, although not materially different from other papers. RKV (2000) also expand the literature's methodology to account for endogeneity by simultaneously estimating for audience (reach) and market value. Reach continues to be positive and significant under this specification. A quarterly, returns regression indicates that changes in the reach variable are positively significant, while earnings and changes in earnings are not. A final contribution of RKV (2000) is an examination of the acquisition prices for 42 (public and private) firms, in which they find that internet activity is positively related to acquisition price.

Demers and Lev (2001) attempt to explain the price reactions of internet stocks before and after their first downturn in the spring of 2000. Using *Nielsen//NetRatings* data and factor analysis, they identify three factors that are referred to as reach, stickiness, and customer loyalty. Using price-to-sales in place of the more common market-to-book ratio, results are generally consistent with the other literature with reach and stickiness being positively priced for internet stocks. Demers and Lev (2001) also provide some evidence of changes in the way internet stocks are priced before and after the market downturn, although both reach and stickiness continue to be priced. This approach, however, may not accurately reflect the potential impact of excess stickiness in certain business models.

In this paper, I examine the value-relevance of both financial and industry-specific, non-financial information for internet firms. In the next section, I detail how differences across internet business models often lead to very different predictions regarding the sensitivity of firm valuations to non-financial measures employed in these studies.

1.3 Internet Business Models

Due to the dynamic nature of the industry, the number of different business models for internet firms potentially exceeds the number of public firms available. Devising a methodology for grouping these firms based upon such models is difficult but critical. Aggregation of such a broad collection of business models into a single sample (e.g., Hand, 2000(a,b)) will increase the heterogeneity of the sample and lower the potential explanatory power of any tests. Alternatively, a dichotomous portal & content (P&C) vs. e-tailing classification and/or a study which focuses only on firms with reported web activity (e.g., TWZ, 2001(a,b)) overlooks the contributions that can be made from a more expansive study of internet firms that are not *ex ante* known to be reliant on web traffic for some portion of their revenues.

In order to gain a better understanding of some of the business models involved with the internet a simple framework is provided in Figure 1 showing the theoretical paths that web-activity-dependent firms follow from start up to revenue generation. Firms begin by making large expenditures on R&D to develop a site's quality, improving their ability to retain viewers (proxied for by visits and time spent per person) and attract new ones via reputation effects. In addition, firms engage in major advertising campaigns and other promotions (SG&A) oriented towards attracting larger audiences.[3] As audience increases so does the number of pages viewed, increasing the advertising and promotion based revenue opportunities for the firm. Increased audience could also lead to additional opportunities resulting from network economies of scale and scope. In essence, Figure 1 shows potential internet equivalents to the market penetration measure used in Amir and Lev (1996) for the wireless industry and may also proxy for future

[3] Noe and Parker (2000) show analytically that two internet firms, competing in a two-period, winner-take-all model, will advertise aggressively and make large investments in site quality in order to capture market share.

growth opportunities of the firm. Similar frameworks and other techniques could also be developed to determine potential value-relevant measures of market potential and/or penetration for other types of internet firms (as well as "offline" firms).

This paper examines a more complete universe of internet firms, characteristic of Hand (2000a,b), while focusing on the differences among these firms in terms of their business models. I begin with the Wall Street Research Network's (WSRN.com) twelve business model classifications cited in Demers and Lev (2001). Due to similarities in predictions across some of the categories and in order to increase power, I ultimately aggregate these classifications into seven groups: portal, content-community, e-tailing, financial services, enablers, ISP/Infrastructure, and non-sensitive firms. Prior research has made few predictions based upon the different characteristics of these business models.[4] Based upon the process shown for activity-dependent firms in Figure 1, I make predictions about each of these business models below which are summarized in Table 2.[5]

> Portals (10) – These sites (e.g., Yahoo!) provide a starting point for web browsing and information searches. Increases in audience directly translate into increased advertising revenues. Activity, once at the websites, could matter more, as greater levels of pageviews should lead to improvements in search technologies as well as the ability to better target advertising to users based upon those searches. As advertising represents a significant proportion of firm revenues, the number of advertisements shown and/or clicked-through are also expected to be value-relevant.[6]

Under this model, any variables that are (linearly) related to pageviews should be explained, although not necessarily in a linear fashion.

[4] TWZ (2001a) suggest that P&C firms are more likely to show sensitivity to internet data (when compared to e-tailers), due to a greater reliance on advertising revenues.

[5] The numbers in parentheses represents the WSRN type variable in stock list database. The descriptions of these firms are the author's.

[6] Click-through rates refer to the percentage of banner advertisements that are clicked upon, leading a visitor to the advertised site.

Content-Community (3) – These sites draw in visitors through the provision of information (e.g., CNET) and/or the ability to interact with others (e.g., TalkCity). Successful firms are able to encourage users to stay longer and return more frequently. These firms should show the clearest relationship between pageviews and stickiness (how long an individual remains) and firm valuations, since those metrics directly translate into increased revenues for the firm. Measures of unique visitors may not have as clear a relationship with firm valuation, however, since the network effects that lead to increased profitability from each additional user could be counteracted by difficulties in targeting desired demographic groups, slower page delivery times, and/or increasing difficultly in navigating the site.[7,8] Similar to portals, ads shown and/or click-throughs are expected to be value-relevant for content-community firms. To date, studies have aggregated portals and content-community firms into one category. This study examines the descriptive characteristics and information content of the data employed for each model independently.

E-tailing (5) – These firms (e.g., Amazon.com) earn revenues in much the same way as the more traditional "bricks and mortar" (B&M) stores do, through sales. Their sites are characterized by high upfront expenditures in technology (R&D effectively replaces the construction of physical storefronts), SG&A, and advertising (when accounted for separately). Getting browsers to these sites is essential, but inducing them to make purchases is the key driver of revenues. Therefore, the best non-financial measures for these firms are how many visitors respond to their advertisements/promotions (click-through rate) and how many visitors complete a purchase once at their site (conversion rate). These rates can also be used to measure how effective a firm has been in translating its operational investments (e.g. advertising and R&D) into revenues.

Financial Services (6) – While some of the firms in this group are holding companies (e.g., CMGI), the majority earn revenues by encouraging people to subscribe to their site (open accounts) and subsequently selling them services, stocks, mutual funds, and other financial products (e.g., Ameritrade). Pageviews and, to a lesser extent, stickiness should translate into higher revenues. A measure of repeat users, especially those using fee-based services, and of transactions conducted (trades, loans processed, etc...) should improve valuations. For those financial services firms whose major product is the information content on its sites, I expect them to act similarly to content-community firms. Holding companies should show sensitivity to the same variables as the underlying business models of their holdings.

---

[7] To the extent firms require membership and/or collect information on their users, they should be able to continue providing more focused services and advertising.

[8] While not a member of this particular group, concerns about the size and scope of its site, as its target markets grew, prompted Amazon.com to reduce the number of "tabs" on its websites in order to improve the sites' usability.

The following three groups comprise the enabler sub-sample of firms:[9]

Advertisers (1) – These firms (e.g., Mypoints/Cybergold), encourage users to visit their sites and, once there, attempt to build brand awareness of the product and/or traffic to the sites of their clients. This group also includes firms (e.g., Doubleclick.net and 24/7 Media) which sell advertising space and/or manage and distribute ("serve") advertisements for other websites. The former type of advertiser should see its valuation increase with the traffic to, pageviews on, and "stickiness" of its website. Additionally, revenues should improve as click-through rates increase. Higher rates result from better design and targeting of advertisements/promotions. Finally, firms will benefit from improvements in the technologies used to deliver these advertisements, such as incorporating Macromedia's Flash™ or Sun's Java™ in order to make the ads more dynamic, or by increasing the delivery speed of the pages via firms such as Akamai.

E-commerce Enablers (4) – These firms provide technology for businesses and their consumers to conduct transactions via the internet. This group also includes business-to-business enablers and software makers who earn revenues from sales or other transactions processed rather than for the number of visitors to its websites. A measure of these transactions would be advantageous for firm valuations and would probably not correlate well with the audience measures used in recent studies. Firms that enable by means of their own websites are effectively portals for a particular good or service and could show a relationship similar to those firms, although many of those firms also sell their technology to others (e.g., email hosting firms such as mail.com and yesmail provide their own web-based email services to consumers as well as handling the outsourcing of that service for other firms). Pure software firms are not expected to appear in the Nielsen-NetRatings database or to show any relationship between internet activity and firm valuation if they do.

Internet Services (8) – These firms are similar to the e-commerce enablers except for a focus on serving the portal and content-community (P&C) firms. Many firms in this category (e.g., HotJobs.com) would appear as P&C firms in other studies.

Together the following two classifications make up the ISP/Infrastructure sample:

Internet Service Providers (7) - These firms generate revenue from their installed base of users in much the same way telecommunications companies do; by providing services and/or equipment for an up-front fee and/or a monthly charge. Internet service providers (ISP) often have an advantage in the provision of content and portal services to their installed customers and may therefore show some sensitivity to internet usage measures in a manner similar to those model types.

[9] Firms which enable by means of owned and operated websites (e.g., eBay enables people to auction and bid on goods by way of its ebay.com site) are classified into the previous model types when possible.

Speed and Bandwidth (12) – These firms are essentially ISP enablers, a few of which are arguably ISPs themselves. Internet data is not expected to be present for most of these firms and, when available, any relationships are expected to be weak.

The expected non-sensitive classifications are:

Consultants / Designers (2) – The success of these firms relies on how well their clients' websites appear and perform. They should not show any sensitivity to their own web usage statistics and are unlikely to have enough activity to appear in the Nielsen-NetRatings database (described later). These firms will remain sensitive to the well being of the industry as a whole, however, and may provide a good basis for drawing comparisons across different types of business models.

Performance Software (9) and Security (11) – Similar to the design and consulting firms, one would expect their revenue creation to come from sales of their software and/or services. There should be little association between their valuations and the activity on their sites. Some firms involved with performance software (e.g., Internet Pictures Corp.) are essentially enablers to P&C sites and may show a mild reaction to pageviews, as this would translate into more potential users of their product.

1.4 Data Collection

1.4.1 Sample Selection

An initial list of firms was chosen based upon the Internet World 50 lists for 1998 and 1999 of the top 50 public internet firms, ranked by revenues, used in RKV (2000). This list was subsequently merged with the InternetStockList utilized by the remaining literature (from Internet.com, the same source as the former list).[10] The original lists were downloaded on March 20, 2000.[11] While reviewing the sample firms, it became apparent that a considerable number of these were involved in mergers and acquisitions activity during the time period examined. Since such transactions could lead to the omission of acquired firms from the sample (survival bias)

[10] The principle means by which to define an internet firm is whether the majority of its revenues come from the internet. See Hand (2000a) for a more detailed explanation of the InternetStockList.
[11] The list was later expanded, slightly, using the June 1st, 2000 list, which added approximately 30 firms.

and the misreporting of growth and change variables for merged firms (e.g., changes in variables will treat both internal and acquisition-related growth the same), public firms were added back into the database based upon their reported M&A activity in press releases dating back to the firm's IPO or the beginning of 1999, whichever was more recent. When press releases were not available or did not date back far enough, financial statements were examined for such activity. In total, 332 firms were found using this procedure. To the author's knowledge, this is the most comprehensive list of publicly traded internet firms to be examined to date.

1.4.2 Financial Statement and Stock Price Data

Accounting data for these firms comes from Compustat via the 2000 quarterly tape, which includes quarters ending in 1999 through March 2001. Price data are collected from the CRSP tapes for observations from February 1999 through May 2001. The top rows of Table 4 provide descriptive financial statistics for the full sample of internet firms. The average (median) market value of these companies is $4.17 billion ($356 million), while average (median) revenues are only $73.9 million ($13.3 million). Mean (median) net income is -$23.3 million (-$6.7 million) and the market-to-book ratio is 8.76 (3.90).

1.4.3 Non-financial Data

Data for the initial analysis are taken from the Nielsen//NetRatings (NNR) "Internet Audience" database which carries detailed information on the web browsing habits of approximately 57,000 non-business panel members as of June 2000.[12] At any time, NNR's website contains data for the most recent 8 weeks and 13 months worth of reports. NNR's data is maintained at the site, domain, and property levels. A site refers to a unique web address (e.g., finance.yahoo.com). A domain includes all the sites that contain the same root name at the end of their address (e.g., the sites games.yahoo.com and finance.yahoo.com are members of the yahoo.com domain). Properties generally contain all the domains owned or controlled by a particular firm (e.g., Yahoo! would include yahoo.com, geocities.com, and broadcast.com among others).

Data for the valuation regressions comes from the February 1999 – May 2001 records of *home* users and are aggregated at the *property* level. This database is similar to the one used in Demers and Lev (2001). Data includes:

> Unique Audience (UNQAUD) – Defined as the number of different individuals visiting a website within the month. In practice, this measure can only detect the number of unique web browsers not unique visitors. For instance, a single networked computer could be used by several people but would have only one IP address and typically two browsers, Internet Explorer and Netscape. As a result, the reported measure of unique audience is likely to understate the total number of visitors.
>
> Reach (REACH) – This figure represents the percentage of internet users that visit a particular web property within a month. It is the internet's equivalent of Nielsen's television share or ratings points, depending on whether the deflator used includes active internet users or all internet users. Similar to Demers and

[12] NNR also provides data on usage at work and recently added a combined home and work database. Their smaller panel sizes and time series, however, prevent meaningful tests based on this data from being performed. To the extent that some firms in this sample may show relatively different sensitivity to work versus household users, this introduces a sampling bias in the data.

Lev (2001), the latter is used as the measure of audience market share, since it more accurately reflects a firms' ability to reach its entire potential market.[13]

Pageviews (PAGEVIEW) – In the NNR database, pageviews refers to the total number of pages seen by all users in the sample, regardless of the means by which they are viewed (see cache below). While sometimes referred to as "hits," pageviews are a more accurate measure of how many times a particular web property has been seen. The methodology commonly used to account for total "hits" leads to inflated measures of a firm's web activity. Pageviews itself is not a consistent measure across different data sources, even in terms of definition. NNR claims that a pageview is only counted when that the page is allowed to load fully. Due to technological constraints, internally generated measures of pageviews by a firm generally register a view at the time a page is requested, even if its not successfully delivered or the users quits the request. This would lead to inflated results relative to a third party measurement such as NNR's.

Pageviews per person (VIEWSPP) – Refers to the total number of pages viewed by the average audience member. In the NNR database, this value has been rounded to the nearest integer. In order to preserve information content, VIEWSPP is recalculated as PAGEVIEW / UNQAUD.

Visits per person (VISITSPP) – Indicates the number of different times an average audience member visits a particular property within a month. Visits are a common measure of how "loyal" a viewer is to a site. NNR does not begin reporting this statistic until August 1999.

Time spent per person (TIMEPP) – Indicates the total amount of time an audience member spends at a property over the month. This variable is commonly referred to as the "stickiness" of a web property is, although pageviews per person has also been used in this regard. The measure is likely to overstate the time spent due to the effects of idling, although NNR controls for browsers that have been inactive for longer than 30 minutes.

Cache (CACHE) – In percentage terms, the amount of time that a page is viewed from a user's own hard drive. Caching is used to store data on recently/commonly visited web pages locally, in order to speed the time it takes a page to load.

A second set of NNR databases contains information regarding internet advertising, including the number of times a banner advertisement has been seen/served and the percentage of times it has been clicked-through. It is organized by both those firms making and those

[13] The reach variable is UNQAUD divided by a number that is constant across firms in any given month hence the

delivering the advertisement for each individual button or banner ad and aggregated across all ads delivered on a single domain. Converting these ads delivered (served) to the property level requires a complicated aggregation process. The domain level data is merged based upon a July 2000 list of the various web properties included in the NNR database and their respective domains.[14] TOTADS represents the total number of *delivered* ad impressions each month across all reported domains for a given property. CLICKS represents the number of advertisements that are clicked-through and is calculated as the product of the reported click-through rate and the ad impressions of each domain and then aggregated to the property level. If the rate is not given it is assumed to be 0.[15] To avoid over-counting, acquired domains were eliminated from the sample for the months prior to the date of acquisition. NNR also maintains several other databases including demographic data for all the properties used above.[16,17] Hand (2000b) examines similar data from Media Metrix and finds no incremental information content beyond that contained in other non-financial measures in the audience database.

Descriptive statistics for these variables are provided in Table 4 for the "web sample" of firms (those firms with data reported in the NNR database). The average firm reaches 2.75% of the estimated population of internet users in the U.S. while the median firm enjoys an audience only about one-third as large. These data suggest that there are a small number of firms which

---

correlations between these two measures are almost perfect.

[14] This process of aggregation may overstate (understate) reported values for the time period prior to (after) July 2000 for firms engaged in M&A activity. The overall impact of this measurement error is likely to be small.

[15] This action appears reasonable given the database's ability to identify firms that served as few as .13 million advertisements.

[16] NNR specially compiled advertising data at the property level for this researcher for the month of July 2000. NNR also collects data on conversion rates and other e-commerce statistics, however, that data was not available to researchers.

[17] The business models discussed in Section 3 suggest other potentially useful sources of data. It may be possible, subject to reporting bias, to assemble some of this data from financial reports. Preliminary evidence from content-community firms suggests that a majority of firms in that business model report at least some of these measure in their press releases and or financial statements. Some financial services firms also appear to provide data related to levels and changes in customer accounts, assets held and transactions processed similar to more traditional, "off-line" firms.

dominate the internet in terms of their market share of eyeballs. The average (median) user makes 2.20 (1.68) trips to a given property each month spending a total of 1.63 (0.18) hours. This order of magnitude difference between the means and medians for time spent online provides further evidence that a small number of firms dominate the attention of internet users.

These web sample firms show an average (median) of 166.32 (17.23) million pages carrying 187.35 (19.69) million ads. Interestingly, despite the large number of advertisements shown, only .22 (.02) million of these ads were clicked upon. As a result, firms able to deliver a high volume of click-throughs could command a premium in the marketplace. On the other hand, if advertising dollars on the net are focused upon enhancing brand value (similar to more traditional media), click-throughs will have a negligible impact on firm valuations.

1.5 Full Sample Results

This section briefly details the correlation and regression results for the full sample of internet firms. It serves as a basis for comparison to the prior literature and the business model partitioned regressions described in Section 6.

1.5.1 Correlations and initial testing

Table 5 presents the correlations among a number of accounting and internet activity variables.[18] Market value is significantly correlated with net income (.08), book value (.09), and the asset-deflated internet variables (.32-.40). Unique audience deflated (per person) internet variables also are also positive and significantly correlated with market value (.22-.26), with the

[18] All accounting variables are deflated by total assets. With the exception of reach, internet usage variables are deflated by both assets and unique audience.

exception of click-throughs per person.[19] Also, with the exception of click-throughs, all variables appear to be positive and significantly correlated with net income (ads shown per person is only marginally significant). The internet usage data shows mixed sign and significance on their correlations with book value, however, none of these correlations exceeds .13 in magnitude. Examination of the audience-deflated internet variables indicates that the correlations among them are noticeably lower than their asset deflated counterparts, suggesting the former choice of deflator would experience fewer multicollinearity problems during regression testing.[20,21]

Valuation regressions throughout this paper follow the basic format of:

$$\frac{MV_t}{DEF} = \alpha\frac{1}{DEF} + \beta\frac{BV_t}{DEF} + \gamma\frac{EBT_t}{DEF} + \delta WEBS_t + \varepsilon \qquad (1)$$

$$\frac{MV_t}{DEF} = \alpha\frac{1}{DEF} + \beta\frac{BV_t}{DEF} + \gamma\frac{EBT2_t}{DEF} + \varphi\frac{SGA_t}{DEF} + \omega\frac{RND_t}{DEF} + \delta WEBS_t + \varepsilon \qquad (1')$$

where t is the month in which firms disclose their quarterly earnings, as reported by Compustat. MV is the market value of the firm at the end of month t. BV and EBT are measured as the total book value and earnings before taxes for the current quarter. SGA and RND are the current periods expenditures on SG&A and R&D, respectively, and EBT2 = EBT + SGA + RND.[22]

---

[19] Due to the lack of reported data prior to August 1999, correlations for visits per person are run on a smaller set (564 observations). In order to preserve sample size, valuation regressions will be conducted with these missing values set to their sub-sample's mean, where applicable. Similarly, advertising data is not available prior to May 1999. Regressions using this dataset are restricted to this later time frame in order to avoid biasing results from zeroing observations which do not appear to warrant it (i.e., zeroing the data appears to introduce a negative bias as larger firms are more likely to appear in the sample during these earlier months).

[20] As a result of this relationship and the easier interpretability of the per person data in the regressions, most of the empirical results reported will use per person variables for internet activity.

[21] Results of deflating accounting variables by total book value are generally consistent with those under total assets deflation with the exception of a negative and significant relationship between income and market value.

[22] R&D and SG&A are expressed as positive numbers.

WEBS reflects a single or set of web usage variables and DEF is either total assets or book value.[23]

In Table 6, regressions on the full sample of internet firms indicate that the coefficient on book value is positive and significant throughout the set of regressions, consistent with prior studies and the predictions in Table 1. Regressing on earnings and book value alone (equation 1) shows that the coefficient on earnings before taxes is *positive* and significant. This result contrasts with TWZ (2001a) and RKV both of which find a *negative* coefficient on earnings, however, only the former is significant. The coefficient on earnings of 3.547 appears low considering that internet firms are generally characterized as "growth stocks." Previous research studies have suggested that expenditures on SG&A and R&D may be viewed as investments rather than a period expense by the market (e.g., Demers and Lev 2001). Failure to control for the potential investment effects of SG&A and/or R&D could negatively bias the coefficient on net income. In order to determine whether the expensing of SG&A and R&D affects the coefficient on net income, I replace equation 1 with 1'. Results indicate that the coefficients on EBT2, SG&A and R&D are all *positive* and significant with the coefficient on EBT2 more than twice that of EBT in equation 1. These findings are consistent with TWZ (2001a) who find a positive and significant coefficient on gross profits for their full sample, however, the coefficients on R&D and marketing expenses (a component of SG&A) are positive but generally insignificant. Differences in the results of those studies and mine are likely the due to the longer time-series employed and the larger scope of the full internet sample in this paper.

When each internet usage variable is individually added to the regression, all except click-throughs per person are positive and significantly associated with market value (ads shown

[23] Reported results, including the constant, use total assets as the deflator. Book value was not used due to an apparent negative bias from the close relationship between book value and net income for these short-lived firms.

only marginally so). Click-throughs, on the other hand are negative and significant. The result is surprising since click-throughs represent a measure of the quality of a browser's attention to the advertising being delivered. Considering the scarcity of click-throughs across most firms (about one in a thousand advertisements are clicked upon) in the full sample and the heterogeneity introduced by the pooling of the various business models, however, it is difficult to draw any conclusions from this result. Regressing on the first four non-financial (non-advertising) variables together, reach and visits per person are both found to be significant and positive, time spent per person is positive but not significant, while pageviews per person is negative but not significant. Together, these results provide solid evidence that firms attracting larger audiences or more repeat visitors to their websites have higher valuations.

In the subsequent regression, the newly generated advertising data appear to have incremental explanatory power. Controlling for other aspects of user activity such as pageviews and time spent online, advertisements shown per person is positive and significant, indicating that firms able to deliver more advertising are given higher valuations.[24] Clicks-throughs per person continue to be negative and (marginally) significant, possibly the result of the high correlation between ads and click-throughs (0.7). Inclusion of these two advertising measures does not alter the coefficients on the other variables. Overall, the findings on the non-financial variables in this section are consistent with prior studies with the exception of Hand's (2000b) finding of no significance for hours and pageviews.

Earlier results were generally consistent for alternative specifications of the deflator and earnings measure.

[24] Due to the different, smaller sample employed for regressions with advertising data it is not possible to directly compare the R^2s across the advertising and non-advertising data included regressions. Restricting both tests to the smaller sample, advertising remains incrementally relevant.

1.5.2 Factors analysis and date-partitioned results

Since February 2000, firms involved in the internet have seen most, if not all, of their stock gains from the late 90's evaporate (the ISDEX index associated with these internet.com stocks has fallen 75% from then through May 2001). As a result, a number of questions have arisen over the continuing relevance of non-financial information in the latter period and whether or not investors have come to appreciate the importance of accounting fundamentals for these stocks. With the development of a longer time-series in this study it should now be possible to examine the question of how the pricing of internet stocks has changed from the boom (through February 2000) to bust (since) periods in the markets for these stocks. Previously, only Demers and Lev (2001) has provided any tests of changes in information content over time although their post-boom sample size is severely limited (fewer than 100 observations).

Table 8 re-examines the value-relevance of accounting and non-financial data for these pre and post-crash time periods.[25,26] Results for the pre-crash period indicate that earnings are negative but not significantly priced. Disaggregating SG&A and R&D, the coefficient on earnings changes sign but remains insignificant. SG&A and R&D are both positive, although only the former is (marginally) significant. Regressions including the non-financial measures indicate that reach and pageviews per person are positive and significant while advertisements per person loses significance. In the post-crash period both EBT and EBT2 are positive and

[25] The term post-crash is used for exposition purposes and is not meant to indicate that the fall in internet stock prices is isolated to March 2000. In practice, the "post-crash" period includes the general downward market conditions these firms have faced since February 2000.

[26] It should be pointed out that the post-crash period's sample is limited as the result of a number of mergers & acquisitions and other means by which a firm is no longer trading (7% of firms in the initial sample have de-listed or gone bankrupt by the beginning of March 2001). In addition, another 30% of firms were trading at below $2 per share at that time and many have since delisted. This will introduce some survival bias into the sample and could also indicate the "going-concern" assumption implicit in the valuation model is violated. In future research I intend to examine new methods of defining and testing the implications of distress in high-technology and other "new economy" firms.

significant as are SG&A and R&D. Even as accounting fundamentals have become increasingly value-relevant, non-financial measures continue to be priced. From Table 8, it can be seen that the more commonly cited reach and pageview measures are positive and significant in both time periods and that advertisements seen only becomes significant in the more recent period. These results suggest that the non-financial measures contain information on the future growth opportunities of these firms beyond that contained in earnings.

In order to address potential problems with multicollinearity across the non-financial measures, I also employ a factors analysis model. The approach is based on the model used in Demers and Lev (2001). Differences in their model and mine result from the inclusion of the two new advertising measures and the removal of the unique audience (which is virtually identical to reach) and cache (which does not contribute noticeably to any factor in initial testing) variables. Results from Table 8, Panel A, indicate that the first factor weighs most heavily on the various "per-person" variables (most notably pageviews and time spent per person). The second, "volume," factor reflects the overall reach and pageviews (to a lesser extent visits per person as well) for the firms. These two factors roughly correspond with the "stickiness" and "reach" factors in Demers and Lev (2001). The third factor consists primarily of the two new "advertising" variables. Consistent with previously reported results, the "volume" factor is positive and significant, however, the other factors are not significantly priced. Overall, results from the factor analysis are consistent with prior results in this paper and others on the informativeness of non-financial measures in general for internet firms. In the next section, I examine the explanatory power of this dataset on the individual business models.

1.6 Business Model Results

1.6.1 Portals

In the next two sub-sections, I examine the descriptive statistics and value relevance of accounting and internet usage data for firms with portal and content-community business models, respectively. While generally aggregated in both the academic literature (e.g., TWZ, 2001(a,b)) and the popular press, content-community firms are noticeably different from portals.

Table 9, Panel A, presents the descriptive statistics for portal-based business models. In general, portals are larger than the average internet firm in Table 4 with mean (median) market values of $16.6 ($1.67) billion. Likewise, the mean and median market-to-book ratios (13.79 and 5.16, respectively) for portals are about 50% larger than for the overall sample, suggesting a larger share of portal firms' value are in intangible assets. Mean earnings before taxes for portals are *positive* ($32.0 million), however, the median firm in this sample continues to lose money (-$19.0 million). The reliance of this business model on websites and the larger size of these firms lead portals to have the highest presence of reported web activity of any business model for both non-advertising (reach, pageviews, time spent, and visits) and advertising (ads shown and click-throughs) data (about 89% and 76%, respectively). With the exception of visits per person, mean values for the internet variables range from 3-5 times larger than for the full sample of internet firms.

Panel B presents the correlation matrix for these firms. Despite the relatively small number of total observations (n=61-82) in this sub-sample, the correlations between market value and net income and all the per-person internet variables are positive and significant (click-throughs per person and market valuation is only marginally significant). None of the variables

show any significant correlation with book value. There is a high degree of collinearity (.55 or higher) among the internet usage variables, suggesting potential problems with multi-collinearity given the smaller sample size.

Panel C summarizes the results of regressions of market value on accounting and internet usage variables. In the first regression, using book value and income alone, both variables are positive and significantly associated with firm values. Similar to the full sample, earnings are next decomposed in order to control for any investment effects from SG&A and R&D. All three of the new variables are positive and significantly priced and book value retains its' significance (R^2 increases by about 5.5%). These results are consistent with the findings of TWZ (2001a) on an aggregated sample of portal and content-community firms except that they find a negative and significant coefficient on marketing expenses and a positive but insignificant coefficient on R&D. The difference in results for this paper, however, are not surprising given the need of these firms to increase their websites' activity through large marketing expenditures and technological improvements and given the longer time-series examined.

Adding reach into the regression, the coefficient on book value loses significance. Reach itself is positive and significantly associated with firm valuations with R^2 improving 13%. Replacing reach with either pageviews, time spent online or visits per person produces positive and significant results, with R^2 increases of 11.6%, 16.7%, and 5.4% respectively. TWZ (2001a) shows similar results for reach and pageviews, with the latter showing a slightly greater increase in R^2s. With regards to the advertising data, regressing on advertisements shown and click-throughs (separately) produces positive and significant coefficients for each. Throughout these regressions, the coefficients on the accounting variables remain positive and significant.

Combining the non-advertising internet usage data, time spent per person is positive and significant, pageviews are negative and significant, and reach and visits per person are positive but not significantly priced. These results support the contention that increased activity within a portal's websites is more value relevant than attracting new audience members as increases in time spent online by the average browser will likely result in improved profiling and targeting of promotions to those visitors. Further inclusion of the advertising measures to the other non-financials leads to no significant results on either variable, although book value loses significance and reach becomes positive and significant. Due to the lower number of observations in this sub-sample and the high level of multi-collinearity across the internet usage variables, drawing any clear interpretations from this model would be difficult. In order to address problems with multi-collinearity in this smaller sample, I re-ran the factor analysis for this business model (based on the factors calculated for the full sample). In results not shown, the first activity-based, "per-person" factor is positive and significantly priced as is the "volume" factor most closely associated with reach, pageviews and visits per person (t-statistics of 1.96 and 4.21 respectively). The "advertising" factor is also positive, although not quite significant (t-statistic of 1.64).

1.6.2 Content-Community

Table 10, Panel A, presents the summary statistics for firms employing the content-community business model. The mean (median) market value for these firms is $786 ($215) *million*, an order of magnitude lower than for portals. Mean (median) revenues and earnings before taxes are $18.40 ($7.82) and $–34.11 (-$9.23) million respectively. Mean (median) market-to-book is 5.76 (2.91), about 40% (60%) that of portals. Similar to portals, content-

community firms have a large percentage (66-81% depending on the variable) of observations reported in the NNR database. Unlike portals, however, the magnitude of these variables is a half to full order lower for content-community firms. The mean (median) firm in this sub-sample reaches only 1.9% (1.2%) of the estimated universe of internet users in the U.S., versus the 12.0% (6.8%) reach for portals, and has only 45.1 (25.7) million pageviews per month. Given the distinct differences discussed in Section 3 and shown here, the implicit assumption of homogeneity between these groups made in prior studies would appear to be invalid.[27]

In Panel B, the correlations for the content-community business model tend to be far less significant than for portals despite the larger sample size of the former. Only reach is positive and significantly correlated with market value (.23). Reach is also significantly positively correlated with EBT (.12) at the 10% level. Time spent, pageviews and visits are all negatively correlated with book value (-.15 to -.17), otherwise neither accounting variable is significantly correlated with any of other variables. Compared to portals, content-community firms show less correlation among the internet usage variables. Reach shows a mild positive correlation with visits (.32) and time spent per person (.15, significant at the 10% level). Pageviews per person are highly, positively correlated with time spent online (.85) and more moderately with visits (.45) and ads shown (.44). Similarly, time spent is correlated with both visits and ads shown per person (.57 and .40 respectively). Finally, click-throughs are not significantly correlated with any other measure of usage than ads shown (.52).

Panel C summarizes the results of the linear regressions of market value on accounting and internet usage variables. Regressing on book value and EBT alone, both variables are positive, although only book value is significant. This result contrasts with the finding of a

[27] Tests on the differences in means between the portal and content-community firms indicates that all the variables shown in Table 10a are significantly different from those in Table 9a.

positive and significant coefficient on EBT for portals. Decomposing earnings, EBT2, R&D and SG&A are all positive and significant. Adding reach into the regression, the coefficient on book value becomes marginally significant. As predicted, reach itself is positive and significant with R^2 increasing by 3%, less than one-quarter the increase for portals. Contrary to predictions, replacing reach with either pageviews, time spent online or visits per person, there is no significant relationship between internet usage and market values. Again, these results directly contrast with the findings for portal firms in which these variables are significant. Not surprisingly, using these variables collectively, reach is positive and significant, book value loses significance and the other non-financial variables are not significantly different from zero. Unlike portals, regressing on ads shown and click-throughs (separately) does not produce any significance. Combining the two advertising variables with the other non-financials, however, reach and advertisements per person are positive and click-throughs are negative and significantly associated with market valuations.

The lack of significance for many of the non-financials, despite the relatively larger sample size for content-community firms (when compared to portals), is somewhat surprising. Given the smaller, more focused nature of content-community firms, this sub-sample may be more volatile and heterogeneous than the one for portals and therefore have greater noise in their valuations. Similarly, these smaller firms may not have as large analyst or investor followings as do portals, which may reduce the effectiveness with which non-financial information is processed into firm valuations. Also, these firms may have yet to achieve a "critical mass" after which they would be more capable of translating internet activity into increased revenues (e.g., through the better targeting of advertisements from increased pageviews and time spent online).

Overall, the results in this section do not support the existing practice of aggregating portal and content-community firms. The significance of some of the internet activity measures in previous papers would appear to be driven by the former. The lack of robust findings in those studies may be the result of noise generated by merging these two heterogeneous samples.

1.6.3 Online Retailers

Table 11, panel A, presents the summary statistics for online retailing firms (e-tailers). The mean (median) market value for these firms is $1.28 ($.186) billion. Mean (median) revenues and earnings before taxes are $41.09 ($17.10) and -$26.4 (-$11.37) million respectively. Mean (median) market-to-book is 5.79 (2.41), about equal to content-community and below that of portal firms. Slightly lower than the prior business models, 67% of e-tailers have reported activity in the NNR audience database, however, only 43% of firms have sufficient advertising activity to appear in that sample. These results are not surprising, considering the relative importance the e-tailing model places on generating sales over advertising revenues. The mean (median) firm in this sub-sample reaches only 1.43% (.64%) of the internet population, lower than the full sample and the previously examined sub-samples. E-tailers also show slightly below average activity, consistent with the interpretation that other variables are more important for evaluating these firms and that excessive pageviews may be an undesirable trait.

Panel B presents the correlation matrix for these firms. The correlations for both market value and net income and all but one of the internet variables (click-throughs per person) are positive and significant. Market value and net income are not correlated with each other. Net income, reach, pageviews and time spent per person are significantly correlated with book value

(.12, -.13, .15 and .16 respectively). Similar to portals, there is a high degree of collinearity (.67 or higher) among internet usage variables (except for click-throughs per person), although potential problems with multi-collinearity should be less of a concern given the larger sample size (n=261).

In the first regression on Panel C, the coefficient on net income is negative but not significantly different from zero for online retailers. Similar to the prior sub-samples, decomposing earnings produces a positive and significant coefficient on earnings and SG&A, however, R&D is not significantly priced for e-tailers. Reach, pageviews, time spent, visits, and advertisements shown are all positive and significant while click-throughs are not significantly different from zero when regressed independently. These results are generally consistent with those of TWZ (2001a) which find reach and pageviews are positive and significantly priced, although reach (visitors) loses significance when earnings components are used in their paper. Inclusion of these variables, however, eliminates the significance of the coefficient on earnings and produces only mixed significance on SG&A. This suggests that, unlike the previous models, non-financial information may be more of a substitute than a complement for accounting information. In comparison, TWZ (2001a) does not find any significant relationship for marketing expense and a weak positive relationship between firm R&D and valuations in one regression.

Regressing on all four non-financial (non-advertising), variables, time spent per person is positive and significant, pageviews are negative and significant, while reach and visits per person are not significantly different from zero. Similar to the findings for portals, these results indicate that firms attracting browsers who spend more time at their websites (who are more likely to make a purchase during their visit) tend to have higher valuations. Additionally, the negative

coefficient on pageviews suggests that sites that require consumers to navigate more pages to find what they are interested in (who are less likely to make a purchase) have lower firm valuations. Unlike portals or content-community firms, the further inclusion of the two advertising measures with the prior measures does not produce a positive coefficient on advertisements shown per person. In addition, click-throughs are *negative* and significantly related to market values. One possible explanation for this result is that firms generating larger volumes of click-throughs may be distracting their customers and/or may be penalized by the market for concentrating too greatly on advertising relative to commerce revenues. Replacing the non-financial measures with the previously described factors, results are generally consistent with expectations. Both the "per-person" and "volume" activity measures are positive and significant, while "advertising" is negative and significant (4.34, 2.37, and –1.92 respectively).

1.6.4 Other Business Models

This section summarizes the results for financial services, enabling, ISP/Infrastructure and non-sensitive firms. As discussed below, these business models have a limited amount of observations with reported internet usage and, in some cases, similar results. Hence, to conserve space, results are summarized more briefly here than for prior business models.

The largest number of observations of any business model belongs to the enabling firms (n=603), however, only 50 of those observations possess enough internet activity data to appear in the NNR database. Similarly, expected non-sensitive firms represent the second largest sub-sample and have an extremely low presence of activity data (19 out of 380 observations). A third sub-sample, ISP/Infrastructure firms, represents another 293 firms of which about one-quarter (77 observations have audience data, only 33 possess data on ads shown and clicked-

through) appear in one of the non-financial databases. In a study of internet firms, exclusion of these three groups (e.g., TWZ 2001a,b) eliminates two-thirds of the available sample. Inclusion of these firms while failing to account for differences in the sensitivities of their business models to the data (e.g., Hand 2000a,b), could lead to erroneous conclusions regarding the usefulness of accounting and non-financial information for valuation purposes.

The final group of firms examined in this section are those involved in financial services. Consistent with the interpretation that attracting users to their services is an important first step, audience appears to be an important element in the operations of about 40% of these firms (41 out of 96). Unlike other traffic dependent business models, however, advertising is not expected to be a major component of revenues and only 19 observations containing sufficient activity to be reported in the NNR's database.

Table 12, Panel A, provides some descriptive characteristics for the various business models. On average, ISP/Infrastructure firms tend to have much larger market values, financial firms are of average size, and enablers and non-sensitive firms are smaller than average when compared to the overall sample of firms. Revenues are, on average, much higher for ISP/Infrastructure and financial services firms than for the other two models. Mean net income is, however, negative for all 4 groups and is of about the same size. Enablers and non-sensitive firms both have negative mean EBT of -$7.8 and -$3.4 respectively. With the exception of financial services firms, which are similar to content-community in terms of the magnitudes of the internet usage variables (slightly higher on average), these business models have negligible activity on their websites (when reported).

Even though mean net income is negative for ISP/infrastructure firms, market values are positive and significantly correlated with net income (.19).[28] These results suggest that markets may be less willing to tolerate negative earnings from these firms. For the 77 observations with reported activity there is no relationship between the non-financial variables and either market value or earnings with the exception of pageviews and visits per person which are negative and significantly correlated with both. It is likely that much of the activity to these sites is oriented towards customer service and other information needs of their current or potential customers and may therefore be treated as expenses by the market.

Among the other three models earnings and market value are only significant for enablers (.10). For enablers and non-sensitive firms, the correlations with net income are not significantly correlated for any other variable. For financial firms, net income is positive and significantly correlated with pageviews (.27), time spent (.30), and visits per person (.31). For both enablers and financial services reach is positive and significantly (.24 and .48 respectively) correlated with market values. Otherwise, none of the non-financial variables are significantly correlated with market values for any of these three models. Intra-web the results are generally consistent with the un-partitioned sample.

Table 12, Panel B, summarizes the results of regressing firm valuations on accounting data alone. ISP/Infrastructure firms have a positive and significant coefficient on both value and EBT. Disaggregating earnings, all four accounting variables are positive and significantly priced with a R^2 of 11%. These results are not surprising given the generally better developed and understood telecommunications firms involved in this model. The significant coefficients on both SG&A and R&D are consistent with these expenditures representing investments in their customer base and in developing new technologies, respectively. Both non-sensitive and

[28] Due to the low occurrence of significant results in this section, correlations for these firms are not shown in tables.

financial firms show no significant coefficient on earnings in either its full or decomposed specification. Further research into value-relevant measures seems necessary in order to develop better pricing models for these firm types (e.g., number of customers and their satisfaction level for the former and accounts or transactions processed for the latter). With respect to the informativeness of the available, non-financial variables, financial services, ISP/Infrastructure and non-sensitive firms do not appear to have any significant coefficients. While section 3 does suggest some proxies for the success of these models, overall, these results suggest the need to examine their business models in greater detail in order to determine better statistical testing for them in the future.

Enablers provide the most interesting results among these sub-samples. Regressing on book value and EBT, the former is found to be positive and significant, while the latter is negative but insignificant. After decomposing earnings, EBT2 is positive and (marginally) significant and R&D and SG&A are positive with the latter significantly different from zero. Enabling firms also show a positive and significant coefficient on visits per person when regressed collectively with the other non-financial variables suggesting increasing visits may serve as a proxy of customer interest. Given the importance of developing standards for enabling technologies, to best achieve economies of scale, these firms must make large up-front investments in R&D. Due in part to rapid industry change and the high level of competition among firms to become these standards, however, the returns to these investments in R&D are highly noisy. The negative coefficient on aggregated earnings does suggest that the marketplace recognizes the needs of these firms to sacrifice current profits in order to succeed long-term. The willingness of the market to reward such activities may be short-lived, however, and in the next sub-section I examine whether or not this and other relationships have changed over time.

1.6.5 Time-Partitioned Results

Similar to the full sample examined in Section 5, results for the individual business models are probably sensitive to changes in the market's perceptions of these firms over time. In this sub-section, I repeat the time-partitioned tests conducted on the full sample for each individual model. Due to the sheer volume of regressions involved, I will only be able to highlight some of the main results.

For the full sample of internet firms most of the evidence for positive pricing of earnings appears to be isolated to the post-crash period. In this section, both portals and ISP/Infrastructure firms showed evidence of positive pricing. Partitioning these models into pre and post-crash periods indicates that EBT is positive and significantly priced for each business model for *both* time periods, although only marginally so for portals in the pre-crash period.[29] The coefficients on earnings for the other models are generally insignificant and of mixed sign over the entire sample period. After time-partitioning these models, the coefficients on earnings for content-community and non-sensitive firms are insignificant in the pre-crash period but positive and significant post-crash. Earnings are negative but not significantly priced for online retailers in the pre-crash period and are positive and significant after the crash. Earnings for financial services firms are not significantly priced in either period. Interestingly, earnings are *negative* and significant for enablers during the pre-crash period and not significantly different from zero post-crash. The prevalence of this sample over the entire universe of firms (about one-third of the pre-crash sample in this paper consists of enablers) likely explains the negative

[29] It should be noted that the samples sizes are small for portals (38 and 44 for the pre and post-crash periods, respectively) which reduces the chances of finding significant results.

pricing discovered in Hand 2000(a,b). In addition, the results for portals, content-community and e-tailers in the pre-crash period are generally consistent with TWZ (2001a).

SG&A increases in significance over time and is positive and significant for each model except for online retailing (versus being significant for only ISP/Infrastructure and portal firms in the pre-crash period). The increase in significance for SG&A is likely the result of greater spending controls imposed as funding has dried up in the post-crash period and firms concentrate their spending on more profitable and/or less risky promotion. The results on R&D are mixed with the coefficient losing its' significance with online retailing, while becoming positive and significant for content-community firms. The significance of the non-financial measures are generally unchanged with the exception of a loss in significance in advertisements for content-community firms and positive and significant coefficients on reach and advertisements for financial services firms (although the value of the latter is quite low, 0.004) from the pre-crash to the post-crash period.

1.7 Conclusions and Suggestions for Further Research

In this paper, I describe the various business models that exist for internet firms and discuss the importance of distinguishing among them. This includes generating predictions on the relevance of both accounting and other forms of data for firm valuations. Consistent with these developments, I have shown that seven categories of firms involved with the internet, based upon their business models, differ with respect to both descriptive characteristics and the usefulness of accounting and non-financial measures in explaining their stock valuations. In particular, portal and content-community firms, grouped together in prior research, are shown to vary noticeably in the magnitude and sensitivity to the data employed. Among those groups with a high level of reported web activity on Nielsen//NetRatings (40% or higher of the sub-sample), portals and online retailers show a positive and significant coefficient on nearly all the usage variables. Content-community firms are sensitive to reach as expected but do not show any significant relationship with pageviews. In contrast to predictions, financial services firms have no relationship to the data over the full sample period. Firms with low reported activity show no association between these measures and firm valuations.

Previously unexamined data on the advertising capacity of these firms (ads shown and click-throughs) are found to be incrementally useful in explaining valuations in the full sample of internet firms. For individual models, advertising data is positive and significantly related with firm valuations for portal, content-community, and e-tailing firms. For portals and content-community firms, these data are also found to be incrementally relevant after the inclusion of other internet usage data. Future research could be conducted with respect to business model specific variables that are suggested as being value relevant in Section 3 (e.g., registered users for content-community firms). In addition, while some of the measures in this study may proxy

for the valuation of internet brands and trademarks (reach in particular), additional data relating to firm visibility on the internet (e.g., prevalence in search engines) could also developed as a possible explanation for these intangibles.

Given the dramatic rise and fall of internet stocks over the last few years (the ISDEX index associated with these internet.com stocks has fallen 75% from the end of February 2000 to May 2001), the case can be made that, even if these non-financial measures were value-relevant in 1999 through early 2000, they may not be currently. Through the development of a more extensive database (through the first quarter of 2001), I examine the question of how the pricing of internet stocks has changed from the boom (through February 2000) to bust (starting March 2000) periods in the markets for these stocks. The positive pricing of earnings before taxes for the entire internet sample appears to be driven by observations in the bust period. The coefficients on earnings for ISP/Infrastructure and portal firms are robust to the different time periods. Results indicate that the negative pricing of earnings observed by Hand (2000a,b) and others would appear to be isolated to firms in the pre-crash period which develop enabling technology for other companies to conduct business on the internet and from e-tailers. Most importantly, despite the increasing relevance in earnings in the later time period, the value-relevant non-financial measures generally continue to be significantly priced.

The internet remains a rapidly developing technology that serves as a point of convergence for many traditional and newer industries. While firm turnover is high, a general understanding of how these firms behave and how information regarding them is used by the marketplace is nonetheless beneficial to both academia and the investment community. For all the questions answered so far, many more remain. In particular, there has been little or no research examining the impact of the internet in general (and these variables in particular) on

either "bricks and mortar" (offline) or "bricks and clicks" (both online and offline) firms. Among the questions that could be asked are: Is web activity useful in the pricing of these firms? Has the internet reduced or enhanced the revenue opportunities for these firms? Also, to date, academic research has also focused on the internet as a source of revenues. How has the internet changed the cost structure of existing firms and are the ultimate gains of the new economy to be found from new markets (increased revenues) or increased efficiency (reduced expense)? I intend to examine some of these questions in future research.

Table 1

Summary and Comparison of Prior Research

	Hand (2000a)	Hand (2000b)	TWZ (2001a)	RKV (2000)	Demers and Lev (2001)	Kozberg (2001)
Firm Lists	InternetStockList (ISL)	ISL	ISL	InternetWorld 50 and Business.com	ISL	ISL and others
Accounting Data	Marketguide.com	Marketguide.com	Press Releases	10-Qs	10-Qs	Compustat
Firms	167	212	56	86	84	316
Observations	729	212	179	149	236	1977
Internet Data	none	PC Data	Media Metrix	PC Data – public	Nielsen//NetRatings	Nielsen//NetRatings
Business Models	All in one	All in one	P&C and S	All in one	P, C, F, S, and "services" in one	P, C, F, S, E, I and N
Forecasts	none	Earnings from IBES	None	none	none	none
Basic Model	Log-linear	Log-linear	Linear	Linear	Linear, factor analysis	Linear, factor and path analysis
Dependent	Market Value	Market Value	Market Value	Market Value	Market Value	Market Value
Deflator(s)	None	None	Book value	Assets and Book Value	Sales	Assets (shown) and Book Value
Book Value	+	+	+	+	NA	Depends on model
Income Variable	Core Net Income	Forecasted Net Income	Net Income	Earnings Before Extraordinary	Earnings are decomposed	EBT and decomposed earnings
Income	NA	NA	-	- (insignificant)	NA	Depends on model
Positive Income	+	+	NA	NA	NA	NA
Negative Income	-	-	NA	NA	NA	NA
Zeroed webs	NA	Yes	No	Yes	Yes	Yes
Reach / Audience	NA	+	+	+	+	Depends on model
Other Internet Results	NA	Pageviews: 0 Hours spent: 0 Demographics: 0	Pageviews: +	NA	Stickiness: + Loyalty: 0	Depends on model
R&D	NA	NA	+, mixed significance	NA	Positive, mixed significance	Depends on model
SG&A or Marketing	NA	NA	Marketing: - for P&C	NA	Positive, mixed significance	Depends on model
COGS	NA	NA	NA	NA	Negative, mixed significance	NA
Other Variables	NA	Stock supply and demand data	Gross Margin: +	Sales: +	Cashburn: -	Test new advertising measures and make predictions for other possible variables.

The firm types used above are (abbreviations in parenthesis): Portals (P), Content-community (C), Financial Services (F), E-tailers (S), Enablers (E), ISP/Infrastructure (I) and Non-Sensitives (N). A "yes" for zeroed webs indicates that internet activity levels for firms with no reported audience data are set equal to zero (no indicates a missing result and removal of the firm-observation from the sample).

Table 2

Predictions for Business Models

Business Model	Net Income	Book Value	Unique Audience	Pageviews/ Stickiness	Ads Shown and/or clicked-through	Conversion Rates
Content – Community	+	+	Mild/Strong	Strong	Strong	Mild
Portals	+	+	Strong	Strong	Mild	None
Financial	+	+	Mild	Mild	Weak	Mild
Retail	+	+	Mild	Strong	Weak	Strong
Enablers	+	+	None	Weak	Mild	Mild
ISP/Infrastructure	+	+	Weak	Weak	Weak	None
Non-Sensitives	+	+	None	None	None	None

Predictions on net income are after controlling for all other possible financial and non-financial information. None indicates that no prediction is being made. The impact of conversion rates on values is expected for those sites which sell some product or service directly to a consumer. Due to the lack of available information, conversion rates are not examined in this paper.

Table 3

Sample Breakdown

Firms in initial sample:	332	
Firms (observations) with complete accounting and price data:	316 (1977)	a.k.a. "Full Sample"
Firms (observations) with reported *monthly* web audience data:	150 (2824)	
Firms (observations) with accounting, price and reported web audience data:	128 (649)	a.k.a. "Web Sample"

Table 4

Descriptive Statistics

Variable	N	Mean	Median	Std Dev	Min.	Max.
Market value	1977	4174.49	356.27	31347.42	10.10	902311.9
Book value	1977	522.86	93.50	2002.65	0.42	27641.0
Revenues	1977	73.89	13.25	411.66	0.00	6830.0
Net income	1977	-23.29	-6.72	155.05	-2435.2	1363.0
Market-to-book	1977	8.76	3.90	14.37	0.04	139.83
Unique audience	649	3.59	1.12	7.11	0.10	56.14
Reach	649	2.75	0.85	5.49	0.07	38.78
Pageviews	649	166.32	17.23	697.65	0.31	7163.3
Hours per person	649	1.63	0.18	6.47	0.00	78.32
Visits per person	564	2.20	1.72	1.42	1.00	8.70
Ad impressions	429	187.35	19.69	620.98	0.14	6450.4
Clicked ads	429	0.22	0.02	0.68	0.00	5.52

All results except for Market-to-book, reach and the two per person measure are reported in million.

Table 5

Full Sample Correlations

Pearson correlations for accounting and internet usage variables deflated by total assets, with the exception of reach and per person variables.

Variable	BV	INC	UNQAUD	VIEWS	HOURS	TOTADS	CLICKS	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP
MV	**.09**	**.08**	**.32**	**.34**	**.36**	**.40**	**.33**	**.38**	**.22**	**.26**	**.24**	**.22**	-.01
BV	1	*.06*	.06	*.08*	*.07*	.07	.03	.05	*-.08*	**-.13**	**-.13**	*.09*	.05
INC		1	**.18**	**.14**	**.14**	**.18**	**.16**	**.19**	**.13**	**.13**	**.16**	*.10*	.02
UNQAUD			1	**.82**	**.86**	**.80**	**.86**	**.984**	**.41**	**.43**	**.62**	**.25**	.05
VIEWS				1	**.98**	**.96**	**.83**	**.77**	**.66**	**.64**	**.69**	**.43**	.07
HOURS					1	**.96**	**.86**	**.82**	**.61**	**.62**	**.67**	**.39**	.07
TOTADS						1	**.80**	**.75**	**.69**	**.68**	**.77**	**.52**	*.08*
CLICKS							1	**.83**	**.38**	**.42**	**.62**	**.28**	**.19**
REACH								1	**.38**	**.43**	**.61**	**.23**	.05
VIEWSPP									1	**.93**	**.73**	**.61**	.02
TIMEPP										1	**.80**	**.57**	.03
VISITSPP											1	**.63**	**.16**
ADSPP												1	**.61**
CLICKSPP													1

Variable definitions are given in Appendix A. Correlations shown in bold (italics) are significant at least at the 5% (10%) level.

Table 6

Full Sample Regressions

Refer to appendix A for definitions of variables. All accounting variables are deflated by total assets. With the exception of reach, all internet usage variables are deflated by the web property's unique audience. White adjusted T-statistics are in parenthesis. Reported R^2s are adjusted.

CONST	BV	EBT	EBT2	RND	SGA	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP	R^2	Obs.
15.867	**7.781**	**3.547**										0.2769	1977
(5.63)	**(21.07)**	**(2.76)**											
1.242	**6.324**		**8.337**	**35.896**	**17.995**							0.2970	1977
(0.16)	**(16.44)**		**(4.73)**	**(3.91)**	**(4.92)**								
3.194	**5.675**		**7.966**	**36.794**	**17.619**	**0.451**						0.3154	1977
(0.46)	**(14.55)**		**(4.58)**	**(3.93)**	**(4.85)**	**(4.57)**							
2.461	**5.968**		**8.241**	**37.586**	**17.482**		**0.039**					0.3019	1977
(0.33)	**(15.23)**		**(4.66)**	**(3.97)**	**(4.76)**		**(3.04)**						
2.808	**5.870**		**8.253**	**38.406**	**17.284**			**4.922**				0.3034	1977
(0.39)	**(14.83)**		**(4.65)**	**(3.99)**	**(4.68)**			**(3.19)**					
1.884	**3.656**		**7.889**	**35.055**	**15.788**				**1.104**			0.3082	1977
(0.25)	**(6.50)**		**(4.59)**	**(3.84)**	**(4.36)**				**(5.20)**				
3.108	**5.884**		**7.965**	**37.544**	**17.042**					*0.024*		0.2863	1824
(0.41)	**(14.48)**		**(4.50)**	**(3.70)**	**(4.57)**					*(1.75)*			
2.535	**6.103**		**7.988**	**36.672**	**17.092**						*-2.256*	0.2842	1824
(0.33)	**(15.19)**		**(4.53)**	**(3.67)**	**(4.61)**						*(-1.70)*		
3.017	**4.221**		**7.780**	**35.872**	**16.424**	**0.382**	-0.025	1.553	**0.676**			0.3175	1977
(0.43)	**(8.65)**		**(4.54)**	**(3.87)**	**(4.54)**	**(3.68)**	(-0.68)	(0.38)	**(3.29)**				
3.555	**3.970**		**7.475**	**35.741**	**15.693**	**0.301**	-0.003	-2.668	**0.777**	*0.028*	**-11.537**	0.3005	1824
(0.49)	**(8.00)**		**(4.37)**	**(3.60)**	**(4.28)**	**(2.79)**	(-0.09)	(-0.63)	**(3.63)**	*(1.76)*	**(-2.60)**		

Results in bold (italics) are significant at least the 5% (10%) level.

Table 7

Pre and Post-Crash Regressions

White adjusted T-statistics are in parenthesis. Reported R^2s are adjusted.

Panel A: Pre-crash

CONST	BV	EBT	EBT2	RND	SGA	REACH	VIEWSPP	ADSPP	R^2	Obs.
8.982	**10.681**	-2.140							0.3630	922
(1.61)	**(16.98)**	(-0.94)								
-4.115	**8.910**		8.706	**54.852**	*19.981*				0.3839	922
(-0.35)	**(13.26)**		(1.53)	**(3.53)**	*(1.83)*					
-1.867	**8.068**		7.309	**56.220**	*17.744*	**0.618**			0.4089	922
(-0.17)	**(10.52)**		(1.31)	**(3.66)**	*(1.65)*	**(4.40)**				
-3.040	**8.412**		8.576	**58.254**	*19.320*		**0.061**		0.3896	922
(-0.27)	**(12.31)**		(1.51)	**(3.73)**	*(1.77)*		**(2.39)**			
-1.456	**8.467**		8.348	**66.853**	18.984			0.205	0.3759	769
(-0.13)	**(11.43)**		(1.33)	**(3.55)**	(1.57)			(1.04)		

Panel B: Post-crash

CONST	BV	EBT	EBT2	RND	SGA	REACH	VIEWSPP	ADSPP	R^2	Obs.
9.672	**5.159**	**4.644**							0.2123	1055
(2.21)	**(13.36)**	**(4.80)**								
-7.277	**4.434**		**5.871**	**28.180**	**9.516**				0.2244	1055
(-1.00)	**(12.00)**		**(4.39)**	**(3.23)**	**(3.62)**					
-4.963	**4.050**		**5.759**	**28.513**	**9.469**	**0.258**			0.2349	1055
(-0.70)	**(10.52)**		**(4.31)**	**(3.21)**	**(3.58)**	**(2.36)**				
-4.319	**4.103**		**5.765**	**29.346**	**8.745**		**0.034**		0.2331	1055
(-0.60)	**(10.85)**		**(4.29)**	**(3.21)**	**(3.23)**		**(2.27)**			
-5.278	**4.194**		**5.833**	**29.135**	**9.139**			*0.301*	0.2295	1055
(-0.73)	**(11.20)**		**(4.32)**	**(3.23)**	**(3.40)**			*(1.87)*		

Table 8

Factors Analysis

Panel A: Factors loadings

Variable	Factor 1	Factor 2	Factor 3
REACH	0.21937	**0.79332**	0.11816
PAGEVIEWS	0.36145	**0.80639**	0.11784
VIEWSPP	**0.89013**	0.37255	0.15335
TIMEPP	**0.86815**	0.38488	0.14395
VISITSPP	*0.41782*	*0.53464*	0.11386
ADSPP	*0.46652*	0.26626	**0.72636**
CLICKSPP	0.00910	0.04935	**0.78407**

Coefficients greater than .8 (4) are highlighted in bold (italics).

Panel B: Regression results

	CONST	BV	EBT	EBT2	RND	SGA	FACTOR 1	FACTOR 2	FACTOR 3	R^2	Obs.
Full Sample	**17.426**	**7.399**	**2.894**				0.053	**1.645**	-0.180	0.2805	1824
	(3.49)	**(22.28)**	**(2.43)**				(0.23)	**(6.49)**	(-0.71)		
	3.474	**5.915**		**7.582**	**36.514**	**17.153**	0.094	**16.697**	-0.103	0.3015	1824
	(0.48)	**(15.20)**		**(4.41)**	**(3.63)**	**(4.69)**	(0.41)	**(3.71)**	(-0.72)		
Pre-crash	*12.394*	**10.430**	-3.492				-0.039	**2.975**	0.005	0.3733	769
	(1.83)	**(17.09)**	(-1.37)				(-0.09)	**(5.45)**	(0.01)		
	-0.398	8.518		7.236	**65.869**	**17.830**	0.176	**2.975**	0.084	0.3973	769
	(-0.04)	(11.86)		(1.17)	**(3.51)**	**(1.50)**	(0.89)	**(4.98)**	(0.68)		
Post-crash	11.868	**5.062**	**4.288**				0.219	**1.122**	-0.352	0.2296	1055
	(1.44)	**(15.26)**	**(3.94)**				(0.93)	**(4.90)**	(-0.93)		
	-5.897	**4.284**		**5.596**	**27.859**	**9.713**	0.224	**1.145**	-0.290	0.2427	1055
	(-0.62)	**(10.79)**		**(4.86)**	**(4.23)**	**(3.40)**	(0.96)	**(5.03)**	(-0.77)		

Table 9

Portal Results

Panel A: Descriptive Statistics

Variable	N	Mean	Median	Std. Dev.	Min.	Max.
Market value	82	16634.81	1671.34	34802.63	60.52	149031
Book value	82	1660.30	478.35	4208.08	21.43	22995.0
Revenues	82	364.6	29.32	1216.25	2.37	6830
Net income	82	32.03	-19.04	205.64	-245.0	1178.0
Market-to-book	82	13.79	5.16	17.27	0.07	65.87
Unique audience	73	15.47	10.19	14.81	0.22	56.14
Reach	73	12.02	6.81	11.45	0.17	38.78
Pageviews	73	873.89	179.71	1734.58	2.15	7163.3
Hours per person	73	8.75	1.54	16.48	0.02	78.32
Visits per person	61	3.33	2.51	2.11	1.50	8.7
Ad impressions	62	723.71	182.56	1274.83	20.88	6450.49
Clicked ads	62	0.93	0.28	1.46	0.00	5.52

Reported statistics are in millions except for market-to-book, reach and the per person measures.

Panel B: Correlations

Variable	BV	INC	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP
MV	-.02	.18	**.46**	**.46**	**.55**	**.48**	**.47**	*.24*
BV	1	.03	-.17	.00	-.17	-.18	.02	-.08
INC		1	**.66**	**.39**	**.41**	**.52**	**.50**	**.40**
REACH			1	**.73**	**.79**	**.93**	**.74**	**.57**
VIEWSPP				1	**.95**	**.90**	**.91**	**.61**
TIMEPP					1	**.92**	**.90**	**.69**
VISITSPP						1	**.86**	**.69**
ADSPP							1	**.55**
CLICKSPP								1

Panel C: (Linear) Regressions

Note: All accounting variables and the constant are deflated by total assets. T-statistics in parenthesis are White adjusted when homogeneity is rejected at the 10% level. Reported R^2s are adjusted.

CONST	BV	EBT	EBT2	RND	SGA	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP	R^2	N
1290.920	**11.866**	**64.054**										0.4458	82
(5.01)	**(4.80)**	**(2.82)**											
1159.297	**8.301**		**94.575**	**307.686**	**105.244**							0.5007	82
(2.58)	**(4.57)**		**(3.13)**	**(3.37)**	**(3.13)**								
1526.792	-0.101		**41.584**	**252.068**	10.978	**0.633**						0.6326	82
(3.91)	(-0.04)		**(2.04)**	**(3.19)**	(0.32)	**(5.35)**							
1558.988	2.168		**63.772**	**233.528**	*45.728*		**0.195**					0.6170	82
(6.14)	(1.14)		**(3.19)**	**(2.72)**	*(1.89)*		**(3.40)**						
1522.536	1.153		**53.643**	**223.831**	28.475			**23.337**				0.6686	82
(5.51)	(0.62)		**(3.02)**	**(2.79)**	(1.26)			**(3.70)**					
1316.847	-0.350		**69.362**	**286.103**	53.613				**2.293**			0.5546	82
(3.09)	(-0.10		**(3.29)**	**(3.31)**	(1.50)				**(3.21)**				
1285.197	-0.368		*36.883*	**173.967**	21.228					**0.288**		0.6115	76
(3.37)	(-0.15)		*(1.96)*	**(2.32)**	(0.71)					**(5.42)**			
1086.317	4.013		**57.120**	**212.383**	54.008						**100.726**	0.4948	76
(2.51)	(1.51)		**(2.71)**	**(2.49)**	(1.60)						**(2.54)**		
1476.832	3.503		**45.661**	**224.198**	19.723	0.302	-0.173	**38.468**	-1.468			0.6875	82
(3.98)	(1.12)		**(2.42)**	**(3.03)**	(0.63)	(1.40)	(-1.36)	**(2.89)**	(-1.31)				
1463.602	1.807		31.935	**180.453**	7.963	*0.454*	-0.035	13.782	-1.394	0.149	-59.193	0.6356	76
(3.81)	(0.58)		(1.66)	**(2.42)**	(0.25)	*(1.97)*	(-0.23)	(0.81)	(-1.21)	(1.46)	(-1.30)		

Table 10

Content-Community Results

Panel A: Descriptive Statistics

Variable	N	Mean	Median	Std. Dev.	Min.	Max.
Market value	262	786.04	215.41	1441.38	10.10	11715.34
Book value	262	321.05	77.03	1076.88	4.18	9463.6
Revenues	262	18.40	7.82	34.83	0.0	280.51
Net income	262	-34.11	-9.23	148.96	-1348.9	573.44
Market-book	262	5.76	2.91	9.11	0.19	75.86
Unique audience	213	2.51	1.58	2.53	0.11	12.19
Reach	213	1.88	1.20	1.91	0.07	12.33
Pageviews	213	45.07	25.70	58.27	0.77	331.54
Hours per person	213	0.52	0.26	0.65	0.01	3.56
Visits per person	189	1.95	1.81	0.60	1	4.45
Ad impressions	172	53.72	22.87	73.45	0.52	393.27
Clicked ads	172	0.08	0.02	0.14	0	0.80

Panel B: Correlations

Variable	BV	INC	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP
MV	.01	.06	**.23**	-.03	-.01	-.07	.08	-.08
BV	1	.13	-.06	*-.15*	*-.15*	*-.17*	-.09	-.09
INC		1	*.12*	-.03	-.05	-.02	-.02	.09
REACH			1	.11	*.15*	**.32**	.02	-.01
VIEWSPP				1	**.85**	**.45**	**.44**	.00
TIMEPP					1	**.57**	**.40**	-.04
VISITSPP						1	.17	.04
ADSPP							1	**.52**
CLICKSPP								1

Panel C: (Linear) Regressions

Note: All accounting variables and the constant are deflated by total assets. T-statistics in parenthesis are White adjusted when homogeneity is rejected at the 10% level. Reported R^2s are adjusted.

CONST	BV	EBT	EBT2	RND	SGA	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP	R^2	N
40.625	**4.798**	2.215										0.2708	262
(1.81)	**(6.59)**	(0.91)											
13.871	**3.696**		**6.363**	**41.232**	**15.580**							0.3020	262
(0.78)	**(5.02)**		**(2.59)**	**(2.31)**	**(2.55)**								
39.310	*1.583*		**6.784**	**44.756**	**16.521**	**0.807**						0.3311	262
(1.93)	*(1.78)*		**(2.47)**	**(2.47)**	**(2.51)**	**(2.25)**							
9.939	**4.079**		**6.338**	**38.401**	**16.140**		-0.022					0.3003	262
(0.49)	**(3.63)**		**(2.56)**	**(2.13)**	**(2.56)**		(-0.59)						
11.794	**3.923**		**6.304**	**39.697**	**15.768**			-1.113				0.2996	262
(0.60)	**(3.68)**		**(2.58)**	**(2.24)**	**(2.58)**			(-0.36)					
14.827	**2.713**		**6.523**	**40.800**	**15.498**				0.432			0.3006	262
(0.83)	**(2.24)**		**(2.67)**	**(2.30)**	**(2.57)**				(1.30)				
17.418	**2.342**		**4.493**	**38.974**	**11.616**					0.023		0.3709	245
(0.58)	**(2.96)**		**(3.04)**	**(2.56)**	**(2.82)**					(0.76)			
7.542	**2.857**		**4.333**	**35.807**	**12.088**						-2.859	0.3645	245
(0.26)	**(4.79)**		**(2.93)**	**(2.36)**	**(2.84)**						(-1.22)		
31.188	2.073		**6.765**	**36.497**	**18.117**	**0.908**	-0.049	-1.632	0.218			0.3306	262
(1.25)	(1.32)		**(2.43)**	**(2.15)**	**(3.00)**	**(3.66)**	(-0.64)	(-0.22)	(0.27)				
32.118	1.384		**4.305**	**36.071**	**11.638**	**0.426**	0.010	-4.157	0.220	**0.043**	*-13.349*	0.3805	245
(1.31)	(1.22)		**(2.15)**	**(2.92)**	**(2.59)**	**(2.13)**	(0.18)	(-0.76)	(0.42)	**(2.18)**	*(-1.94)*		

Table 11

Online Retailing (E-tailer) Results

Panel A: Descriptive Statistics

Variable	N	Mean	Median	Std. Dev.	Min.	Max.
Market value	261	1284.86	185.91	4104.84	10.23	38628.1
Book value	261	161.87	77.20	216.37	3.32	1279.10
Revenues	261	41.09	17.10	70.69	0.00	676.04
Net income	261	-26.40	-11.37	33.70	-323.21	38.30
Market-book	261	5.79	2.41	10.17	0.07	87.53
Unique audience	176	1.88	0.76	2.69	0.10	15.18
Reach	176	1.43	0.64	1.96	0.07	9.32
Pageviews	176	141.46	12.46	540.31	0.36	3445.84
Hours	176	1.14	0.13	3.97	0.00	26.16
Visits per person	153	1.78	1.41	1.39	1.08	8.49
Ad impressions	112	166.08	4.39	600.15	0.27	3684.46
Clicked ads	112	0.06	0.01	0.23	0.00	1.83

Panel B: Correlations

Variable	BV	INC	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP
MV	*.11*	.07	**.28**	**.34**	**.38**	**.52**	**.66**	.01
BV	1	*.12*	*-.13*	*.15*	*.16*	.12	.16	.05
INC		1	*.18*	*.19*	*.19*	*.19*	**.32**	.06
REACH			1	**.71**	**.68**	**.73**	**.67**	.09
VIEWSPP				**1**	**.95**	**.97**	**.95**	*.21*
TIMEPP					1	**.98**	**.87**	*.19*
VISITSPP						1	**.95**	*.22*
ADSPP							1	*.22*
CLICKSPP								1

Panel C: (Linear) Regressions

Note: All accounting variables and the constant are deflated by total assets. T-statistics in parenthesis are White adjusted when homogeneity is rejected at the 10% level. Reported R^2s are adjusted.

CONST	BV	EBT	EBT2	RND	SGA	REACH	VIEWSPP	TIMEPP	VISITSPP	ADSPP	CLICKSPP	R^2	N
2.273	**4.512**	-3.12										0.2632	261
(0.11)	**(6.22)**	(-1.44)											
-9.374	**3.759**		**6.388**	18.872	**17.861**							0.2804	261
(-0.61)	**(4.61)**		**(2.84)**	(1.37)	**(4.11)**								
17.379	**1.978**		4.401	9.936	*13.991*	**1.239**						0.3575	261
(0.86)	**(2.51)**		(1.16)	(0.64)	*(1.87)*	**(5.63)**							
11.597	**2.005**		2.935	10.523	11.804		**0.056**					0.3629	261
(0.59)	**(2.57)**		(0.77)	(0.68)	(1.58)		**(5.84)**						
15.580	**1.614**		2.178	10.569	10.079			**7.656**				0.3793	261
(0.80)	**(2.06)**		(0.58)	(0.69)	(1.36)			**(6.47)**					
-5.384	-0.334		4.165	15.388	*13.033*				**1.861**			0.3539	240
(-0.27)	(-0.33)		(1.09)	(0.99)	*(1.73)*				**(5.52)**				
0.552	**2.550**		3.311	12.503	12.993					**0.058**		0.3302	240
(0.02)	**(3.16)**		(0.82)	(0.74)	(1.62)					**(5.08)**			
-17.864	**3.883**		**6.656**	20.132	**18.629**						-13.127	0.2579	240
(-0.91)	**(4.11)**		**(2.67)**	(1.37)	**(3.72)**						(-0.96)		
14.243	-0.098		2.445	11.607	9.647	0.450	*-0.071*	**11.733**	0.968			0.3832	261
(0.70)	(-0.08)		(0.65)	(0.76)	(1.31)	(1.38)	*(-1.87)*	**(2.90)**	(1.55)				
5.016	0.073		3.663	14.397	12.381	0.458	-0.061	**11.422**	1.010	0.001	**-45.905**	0.3818	240
(0.21)	(0.05)		(0.92)	(0.87)	(1.59)	(1.33)	(-1.13)	**(2.56)**	(1.33)	(0.04)	**(-2.53)**		

Table 12

Other Business Models

Panel A: Descriptive Statistics

	Enablers	Financial	ISP/Infrastructure	Non-sensitives
Total observations	603	96	293	380
Observations with audience data	50	41	77	19
Observations with advertising data	24	19	33	7
Market value	1198.55	3437.55	15739.00	1793.10
Market-to-book	8.70	7.00	11.08	10.52
Revenues	25.12	195.10	195.79	24.72
Net Income	-23.48	-44.78	-20.42	-22.13
Unique audience	0.92	2.75	2.07	0.77
Reach	0.72	2.15	1.60	0.52
Pageviews	10.62	88.81	70.00	4.52
Hours per person	0.13	0.85	0.80	0.05
Visits per person	1.64	3.14	2.72	2.19
Ad impressions	10.31	168.37	127.04	3.62
Clicked ads	0.02	0.21	0.28	0.03

Reported statistics are industry means

Panel B: Regressions

Business Model	CONST	BV	EBT	EBT2	RND	SGA	R^2	N
Enablers (E)	**31.825**	**6.520**	1.229				0.3398	603
	(8.02)	**(13.51)**	(0.80)					
E	-1.049	**5.271**		*2.792*	7.967	**15.583**	0.3603	603
	(-0.08)	**(9.63)**		*(1.79)*	(0.91)	**(4.29)**		
Financial (F)	19.645	**6.435**	2.257				0.3055	96
	(0.61)	**(4.92)**	(0.51)					
F	24.312	**6.494**		2.122	-0.839	1.305	0.2904	96
	(0.47)	**(4.58)**		(0.47)	(-0.04)	(0.13)		
ISP/Infrastructure (I)	1.378	**12.397**	**20.511**				0.2696	293
	(0.06)	**(7.28)**	**(2.54)**					
I	-10.317	**6.280**		**29.787**	**227.855**	**54.475**	0.3775	293
	(-0.25)	**(4.23)**		**(4.41)**	**(6.86)**	**(3.88)**		
Non-sensitives (N)	**13.619**	**8.931**	4.055				0.2870	380
	(3.98)	**(10.61)**	(1.23)					
N	6.321	**7.536**		**12.436**	-3.633	**29.521**	0.2983	380
	(0.88)	**(7.69)**		**(2.94)**	(-0.17)	**(3.04)**		

Appendix A

Variable Definitions

Historical accounting data is from the quarterly, June 2000 Compustat tapes. More recent Compustat data was acquired through Factset.

EBT (data23) – Earnings before taxes (bottom-line net income was used previously and the results have not changed)
BV (data60) - Total book value was used in place of book value of common equity in order to reduce the number of negative observations and due to the predominately equity-like characteristics of preferred stock in startup, internet firms.
TOTASS (data44) – Total firm assets.
SGA (data1) - Sales, general and administrative. When a firm reports no cost of goods sold this variable is COGS instead and this variable is reported as 'C.'
SALES (data2)
RND (data4)
SHARES (data15) - Fully diluted shares

From CRSP & Factset:

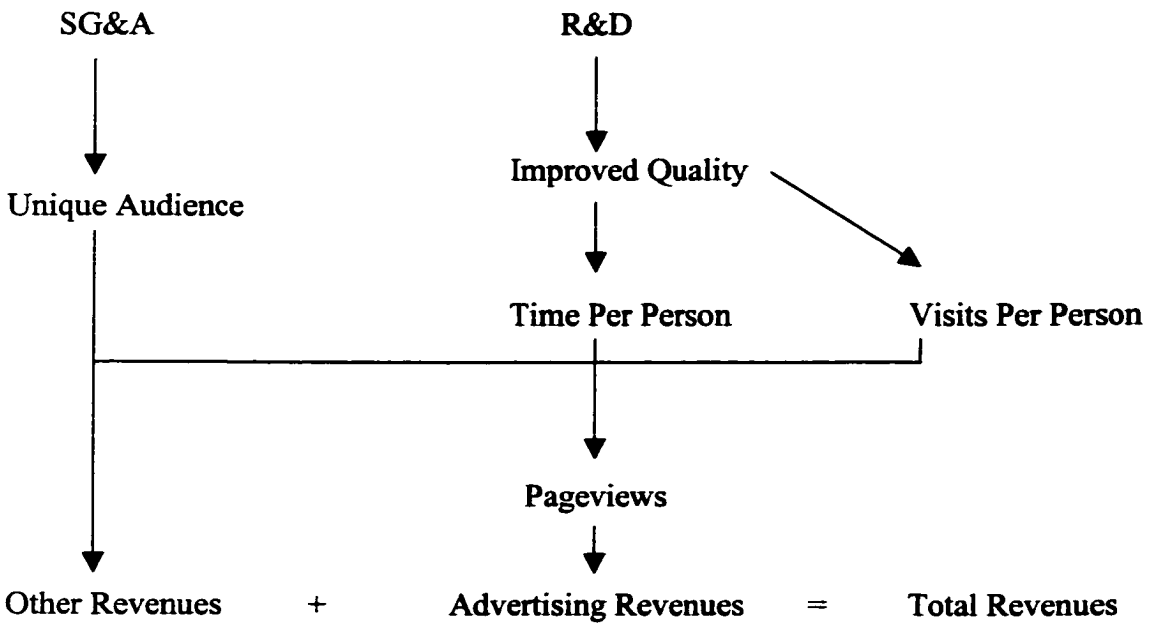
PRICE - End of month, stock price as reported
MV - Equals PRICE * SHARES (from Compustat)

From Nielsen//NetRatings (NNR):

UNQAUD - Unique audience as reported in the monthly audience measurement database.
VIEWS - Total pageviews as reported in the monthly audience measurement database.
REACH - Percentage of total estimated internet audience as reported in the monthly audience measurement database.
VIEWSPP - Average page views per person as reported in the monthly audience measurement database.
TIMEPP - Average time (in hours) spent per person as reported in the monthly audience measurement database.
PAGESPP - Redefined as VIEWS / UNQAUD since NNR rounds their reported variable.
HOURS – Total amount of time (in hours) spent on a particular property by all users, calculated as (UNQAUD * TIMEPP).
TOTADS - The number of ad impressions served by all the domains in a property, aggregated from domain level data reported by NNR.
ADSPP – TOTADS / UNQAUD
CLKRATE - The percentage of ad impressions clicked upon.
CLICKS - The total number of ads clicked upon, defined as TOTADS * CLKRATE for each domain and then aggregated to the property level.
CLICKSPP – CLICKS / UNQAUD

Advertising by sample firms on the internet is available as well but is not included in this study. Audience, views, and ad impressions are in millions. Rates are reported in percentages (10.3) rather than decimal form (.103).

Figure 1



References

Amir, E. and B. Lev, 1996, Value-relevance of Non-financial Information: The Wireless Communications Industry, Journal of Accounting and Economics 22, 3-30.

Barth, M.E, M.B. Clement, G. Foster and R. Kaznik, 1999, Brand Values and Capital Market Valuation, Review of Accounting Studies 3, 41-68.

Demers, E. and B. Lev, 2001, A Rude Awakening: Internet Shakeout in 2000, forthcoming in Review of Accounting Studies.

Deng, Zhen, 1999, Patents, Citations and Equity Valuation, Working Paper, Baruch College.

Hand, J., 2000a, Profits, Losses, and the Pricing of Internet Stocks, Working Paper, Kenan-Flagler Business School, UNC Chapel Hill.

Hand, J., 2000b, The Role of Economic Fundamentals, Web Traffic, and Supply and Demand in the Pricing of U.S. Internet Stocks, Working Paper, Kenan-Flagler Business School, UNC Chapel Hill.

Ittner, C. and D. Larcker, 1998, Are Non-financial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction, Journal of Accounting Research 36, 1-46.

Li, Charlene, August 1999, Internet Advertising Skyrockets, The Forrester Report.

Noe, T. and G. Parker, 2000, Winner Take All: Competition, Strategy, and the Structure of Returns in the Internet Economy, Working Paper, Tulane University.

Ohlson, J., 1995, Earnings, Book Values and Dividends in Security Valuation, Contemporary Accounting Research 11, 661-687.

Rajgopal, S., S. Kotha and M. Venkatachalam, 2000, The Relevance of Web Traffic for Internet Stock Prices, Working Paper, University of Washington and Stanford University Berkeley.

Seethamraju, C., 2000, The Value-Relevance of Trademarks, Working Paper, Washington University.

Trueman, B., M.H.F. Wong and X. Zhang, 2001a, The Eyeballs Have It: Searching for the Value in Internet Stocks, forthcoming in Journal of Accounting Research.

Trueman, B., M.H.F. Wong and X. Zhang, 2001b, Back to Basics: Forecasting the Revenues of Internet Firms, forthcoming in Review of Accounting Studies.

Chapter 2: The Revenue Drivers of Internet Stocks: A Path Analysis Approach

2.1 Introduction

Prior academic literature on the relevance of accounting and non-financial measures for internet firms has generally focused on explaining their stock valuations. In the absence of clear relationships between earnings and these valuations, analysts, corporate insiders and researchers have concentrated their attention on explaining other measures of internet firms' value. These include earnings components, such as revenues and gross margin, and non-financial measures such as unique audience and pageviews. With the exception of an examination of revenue forecast errors by Trueman, Wong and Zhang (2001b), however, there has been no research attempting to explain how these activity measures are generated or their effect on firm revenues, which is the focus of this chapter.

As discussed previously, a better understanding of the relationships among accounting and non-financial measures should improve the identification of value drivers and the means by which they are specified. Figure 1, from Chapter 1, provides a conceptual path diagram from initial management decisions on the levels of SG&A and R&D expenditures through to revenue realization for those firms which rely upon website activity. This chapter refines the path diagram and uses it to test whether firm expenditures on SG&A and R&D translate into increased web activity and whether said activity results in increased revenue opportunities for the firm.

In addition, Chapter 1 illustrates the hazards of testing a sample of heterogeneous firms involved in the internet (distinguished by their business models) as one collective sample. Heterogeneity is only one of several statistical issues that can arise regarding current methodologies for testing these firms, however. For instance, little attention has been paid by

the existing literature to the likely relationships among the accounting and internet activity variables used to explain firm valuations. Chapter 1 shows evidence of high multicollinearity among the internet activity measures for the full sample (Section 5) and certain business models (Section 6). One method employed to correct for multicollinearity in the Chapter 1 and in Demers and Lev (2001) is factor analysis, which replaces raw or deflated internet usage measures with a smaller set of orthogonal factors. This approach, however, allows the data to determine the factors and is inevitably followed by a researcher's ad hoc attempt to interpret the factors. In addition, the choice of factors is highly sensitive to the combination of variables chosen and the approach taken in calculating the factors.[1]

While high correlation and endogeneity are not the same thing, this relationship suggests that some or all of these variables could be endogenous, violating an assumption made in OLS estimation. Treating these variables as exogenous when they are in fact endogenous could result in a number of statistical problems including measurement error and bias. Ideally, these factors should be specified *ex ante*, while still providing the researcher with the ability to control for variable endogeneity.

The methodology employed in this chapter is based upon a path analysis estimation technique first used by Wright (1921). This approach allows a researcher to address issues of factor identification and endogeneity simultaneously. In addition, it permits separate testing of the direct and indirect (through intermediate variables) effects of the selected independent variables. Path analysis is based upon a diagram of the hypothesized relationships among the independent and dependent variables. In the analysis, variables examined are classified into two types, exogenous or endogenous, based upon whether or not they appear as dependent

[1] For instance, Demers and Lev (2001) choose the almost perfectly correlated reach and unique audience as factor components in their model. This choice influences their first factor to load predominately on these two variables.

variables in any of the system of equations. Among the variables employed in this study, expenditures on R&D and SG&A are treated as exogenous while website activity measures and revenues are endogenous. The path diagram is presented in Figure 2, an expanded version of Figure 1, which specifies empirically testable relationships among the data. In Figure 2, single arrows indicate the predicted direction of causation from the exogenous to the endogenous variables.

Empirical testing of this path diagram provides several interesting results regarding the use of activity data for the analysis of internet firms. Consistent with findings in Chapter 1, accounting data on firm expenditures in SG&A and R&D have explanatory power over both website activity measures and firm revenues. R&D, a proxy for investments made to develop website quality, reduces the amount of time a browser needs to spend online at a firm's website. SG&A, which should proxy for efforts to increase website activity levels, is positively and significantly related to time spent and number of visits per person for financial services and online retailing firms. It is also positively and significantly related to time spent per person for portal and content-community firms. Consistent with expectations, both SG&A and R&D are positively and significantly related to unique audience. Finally, SG&A is positively and R&D is negatively and significantly associated with firm revenues, with the latter relationship appearing to be driven by financial services and online retailing firms. These results indicate that at least some portion of firm expenditures are directed towards improving website quality and visitor activity.

Internet activity measures are systematically related to firm revenues as well. As unique audience and time spent per person increase, so do pageviews. Pageviews have the direct effect of increasing firm revenues and increasing the amount of advertising seen. This

direct effect on revenues is most likely the result of the ability of pageviews to proxy for other, non-advertising, firm revenue opportunities associated with greater site activity (e.g., mailing lists and user profiling for portal and content-community firms and transactions for financial services or online retailing firms). Finally, while initial results for advertising data do not show explanatory power over revenues, alternative tests provide evidence that click-throughs are positively and significantly associated with firm revenues.

This paper includes seven sections. Section 2 provides a brief review of the relevant literature. Section 3 details the data collection process and provides summary statistics for the variables. Section 4 describes the path analysis methodology employed. Sections 5 and 6 give the initial and expanded results from empirical testing, respectively. Section 7 summarizes the findings and provides suggestions for future testing.

2.2 Literature Review

A number of recent papers have attempted to value internet firms using a combination of accounting and non-financial measures. Hand (2000a,b), Trueman, Wong and Zhang (TWZ, 2001a), Rajgopal, Kotha and Venkatachalam (RKV, 2000) and Demers and Lev (2001) provide evidence that internet firms' earnings are generally not priced (or in some cases negatively priced). In the absence of positive and significant results for net earnings, several of these earlier papers attempt to use earnings components such as revenues to explain firm valuations. The evidence from those studies is generally mixed, with revenues, marketing expenses (a component of SG&A) and R&D all showing some signs of being positively and significantly valued. Results from the Chapter 1, which includes more recent data than prior studies, provides evidence that net earnings have become positively priced for internet firms in general

and for most business models over time. In addition, SG&A and R&D both show stronger evidence of being positively and significantly priced for the overall sample as well as most individual business models. Finally, non-financial measures such as reach, pageviews and advertisements are shown to be priced for internet firms in general. None of these papers, however, make any attempt at directly examining the determinants of activity and the ability of firms to convert that activity into revenues.

Trueman, Wong and Zhang (TWZ, 2001b) utilize current financial and non-financial data in the prediction of internet firm revenues, which it suggests are a key driver in the valuation of these firms.[2] It focuses on the types of firms for which one would *ex ante* expect web activity measures to have relevance: portal, content-community and online retailing. TWZ (2001b) examines how well different accounting and internet usage variables correlate with analysts' forecast errors (measured in percentages). It finds that analysts systematically underestimate revenue growth from 1999 to early 2000. Growth rates in historical revenues and internet usage seem to have power in explaining these errors for portal and content-community firms, while growth in internet usage is significant in explaining errors for online retailers. While TWZ (2001b) examines the relationship between revenue estimates and their realized values, it does not examine the usefulness of accounting or non-financial information in explaining either analysts forecasts or realized revenues directly. If the influences of the web activity measures are already accurately impounded into the revenue estimates made by analysts, then these measures should have little or no ability to explain errors.

Given the availability of internet activity data from several sources (Nielsen//NetRatings, Media Metrix and PC Data) on a monthly or even weekly basis, it is not

surprising that the explanatory ability of the tests conducted in TWZ (2001b) are somewhat low (R^2s of .15 or less). In addition, given the emphasis placed on the importance of revenue growth for internet firms, these firms may attempt to influence their reported numbers through such activities as the inclusion of "grossed-up" and/or barter revenues as discussed in Bowen, Davis and Rajgopal (2001). Over a long enough time horizon, such adjustments would naturally reverse and/or lead to a higher denominator used for the calculation of revenue growth (implying a negative correlation between past growth and the error). However, over the shorter time horizon examined in TWZ (2001b), it may be possible for management to continue to manipulate revenues in this fashion. These management actions could result in the systematic underestimating of revenues that TWZ (2001b) document.

With the exception of TWZ (2001b), no previous research has examined the ability of either financial or non-financial data to explain other fundamental economic data than internet firm valuations. This chapter extends upon the previous literature by examining the financial and non-financial determinants of firm revenue, while addressing the endogenous and multicollinear nature of these measures.

[2] Justification for their usage of audience measurement data comes from the suppositions that: (1) higher usage reflects greater demand for products and services; (2) increased traffic leads to greater advertising revenues; and (3) higher usage brings in more advertisers and, at least indirectly, higher advertising rates.

2.3 Data Collection

Table 1 provides a breakdown of the number of firms and observations in the samples studied in this chapter. Unlike Chapter 1, this chapter restricts its focus to firms with positive levels of internet activity. This is done in order to restrict the sample to firms that are dependent on web activity for revenues, for which the hypothesized path diagram is more likely to be a reasonable description. Accounting data for these firms comes from Compustat for quarters ending in 1999 through March 2001.

The top rows of Table 2 provide descriptive financial statistics for these internet firms. The average (median) market value of these companies is $3.21 billion ($464 million) and average (median) revenues are about the same at $80.0 million ($17.1 million). Mean (median) net income is -$66.9 million (-$14.9 million) and the market-to-book ratio is 8.48 (2.99).[3] These descriptive statistics are consistent with the larger sample examined in the prior chapter.

The internet activity data for this study are taken form Nielsen//NetRatings "Audience Measurement" and Bannertrack™ databases from February 1999 through May 2001. The data employed include[4]:

> Unique Audience (UNQAUD) – Defined as the number of different individuals visiting a website within the month. In practice, this measure can only detect the number of unique web browsers rather than unique visitors.
>
> Reach (REACH) – This figure represents the percentage of internet users that visit a particular web property within a month.

[3] Market values and net income are presented for descriptive purposes only and are not used in any tests in this chapter. Similarly, book value is not used, therefore the constraint that firms have a book value over 0 is not necessary (leading the market-to-book ratio to be negative for some observations and biasing the ratio lower relative to the full sample in the prior chapter).

[4] In tests conducted using advertising data, the time period examined begins in May 1999 rather than February 1999. For a more detailed explanation of the databases and a longer description of terms, I refer the reader to Chapter 1.

Pageviews (PAGEVIEW) – In the NNR database, pageviews refers to the total number of pages seen by all users in the sample, regardless of the means by which they are viewed.

Visits per person (VISITSPP) – Indicates the number of different times an average audience member visits a particular property within a month. NNR does not begin reporting this statistic until August 1999.

Time spent per person (TIMEPP) – Indicates the total amount of time an audience member spends at a property over the month.

Advertisements served (ADSEEN) - The total number of *delivered* ad impressions each month across all reported domains for a given property. NNR does not begin reporting this statistic until May 1999.

Click-throughs (CLICKS) - The number of advertisements shown that are clicked upon by the browser. NNR does not begin reporting this statistic until May 1999.

Descriptive audience statistics for these variables are provided in the lower rows of Table 2.[5] The average firm reaches about 2.36% of the estimated population of internet users in the U.S. while the median firm enjoys an audience only one-third as large. These data suggest that there are a small number of firms which dominate the internet in terms of their market share of unique browsers. The average (median) user makes 2.04 (1.75) trips to a given property each month spending a total of 0.19 (0.15) hours.[6] These firms show an average (median) of 69.9 (13.9) million pages carrying 85.7 (16.5) million ads but only .15 (.02) million of these ads were clicked upon. As a result, firms that are able to deliver a high volume of click-throughs could command a premium in the marketplace. On the other hand, if advertising

[5] The differences in the number of observations in this sample and those in the "web sample" in the prior chapter result from slight differences in the matching and truncation criterion employed in this study. Observations are matched based upon the final month of the firm quarter in question rather than the month a firm announces earnings. Observations more than 3 standard deviations from the mean are removed.

[6] The previous chapter showed an almost order of magnitude difference between the means and medians for time spent online as well as considerably larger means than medians for other activity measures as well. Due to the greater need to control for outliers using a path analysis framework this relationship has been considerably mitigated.

dollars on the net are more focused upon enhancing brand value (similar to more traditional media), click-throughs may have a negligible impact on firm revenues.

2.4 Methodology

This section presents an alternative approach for examining the interrelated nature of the accounting and non-financial variables used in the valuation of internet firms. Figure 1, from Chapter 1, specifies a hypothetical path for web-activity-dependent firms from start-up to revenue generation. This chapter refines Figure 1 to develop a more detailed, empirically testable, path diagram.

Conceptually, management initiates expenditures on R&D, intending to establish (or enhance) a website's quality. The potential effects of this spending may offset one another, however. Increased site quality should improve a firm's ability to retain viewers, which can be proxied for by the amount of time spent and the number of visits made per person to its websites. On the other hand, website R&D expenditures could be focused upon aspects of quality such as improved delivery times (lowering the average time spent online) rather than on adding further content (potentially increasing time online). Regardless of the means by which quality improves, however, the websites should generate larger audiences as the result of improved brand recognition from reputation effects.

In addition to spending on R&D, firms may choose to engage in major advertising campaigns and other promotions (SG&A) designed to attract new visitors to their websites. These increases in audience should improve the quantity of user generated content. It should also allow more opportunities for members to develop into communities with those possessing similar interests. As a result, increased SG&A could have the secondary effect of encouraging

existing members to use their websites more frequently. Overall, expenditures on SG&A should enhance the "network effects" from having more users online with whom to interact and share information.[7]

As audience increases so does the total number of pages viewed, increasing advertising revenue opportunities for the firms. In addition, pageviews should increase as individual audience members visit and/or spend more time at a website. Increased pageviews translates into more opportunities for firms to deliver advertisements or other forms of sponsored content to their viewers. Naturally, increases in the number of delivered advertisements leads to additional chances for browsers to click-through to the website of an advertiser. On the other hand, as time spent per person increases, browsers are more likely to have seen the same advertisements previously or already viewed those advertised sites reducing their likelihood of clicking-through.

Apart from their impact on the quantity of advertisements shown, increased audience and pageviews could also generate an improved ability to target content and promotions to their viewers which could further increase advertising revenues. Additionally, audience, pageviews, SG&A and R&D could all influence firm revenues directly, proxying for other revenue opportunities such as: (1) online or offline sales of goods and services; (2) the creation and use of mailing lists; (3) alliances; and/or (4) services rendered and content delivered for other sites.

Building upon the logic contained in Figure 1, the methodology used for estimation in this paper focuses on path analysis, a statistical technique based upon a linear equation system that was first developed by Sewall Wright (1921). While uncommon in the financial

[7] Noe and Parker (2000) show analytically that two internet firms, competing in a two-period, winner take all model, will advertise aggressively and make large investments in site quality in order to capture market share. Under this model, any variables that are (linearly) related to pageviews should be explained, although not necessarily in a linear fashion.

accounting literature,[8] it has been utilized frequently in the social sciences and ecological literatures. Path analysis' popularity in those literatures results from its explicit recognition of possible causal relationships among variables. In so doing, it enables the researcher to decompose the correlations between each pair of variables into the different effects that flow from the causal variable(s) to the dependent variable. These effects may be either direct (e.g., increased audience should lead directly to more individuals seeing a site's webpages) or channeled indirectly through other variables (increased audience directly leads to increased pageviews and indirectly causes more advertisements to be seen). Thus one may examine both the direct and various indirect effects of firm expenditures and activity generation measures and assess the impact of each.

This focus on intermediate pathways along which these effects travel makes the application of this technique particularly appealing for internet firms. As discussed previously, understanding the path from firm expenditures to revenue creation provides a clearer understanding of what may be driving the value of internet firms. The analysis begins with a path model that diagrams the expected relationships among the independent and dependent variables. It should be noted, however, that the pathways in these models represent hypotheses of researchers, and cannot be statistically tested for the direction of causality. Figure 2 provides a more developed version of Figure 1 expressed as a path diagram. In path analysis, the variables examined are broken into two types, exogenous or endogenous, based upon whether or not they appear as dependent variables in any of the system of equations. Among the variables employed in this study, expenditures on R&D and SG&A are treated as exogenous while site activity and revenues are endogenous.

[8] An example of the application of path analysis in the accounting literature is Amit and Livnat (1988), which examines the direct and indirect effects of diversification, operating risk and leverage on a firm's systematic risk.

In the main model tested there are four exogenous variables, SG&A and R&D deflated by *both* total firm assets and unique audience (per-person). In any particular equation tested, however, only one of the two deflated sets of variables is used. The decision as to which set to use is based primarily upon which deflator is employed for the dependent variable. The primary reasons for this specification are to maintain consistency across deflators within a single equation, to allow easier interpretability of the results and to avoid introducing competing effects into the data. In Figure 2, single arrows indicate the predicted direction of causation from the exogenous to the endogenous variables that is suggested from the earlier discussion in this section.

The coefficients generated in a path analysis are standardized regression coefficients (betas), showing the direct effect of an independent variable on its dependent variable in the path diagram. Thus, when the model has two or more causal variables, path coefficients are partial regression coefficients that measure the extent of the effect of a causal variable and its dependent in the path model controlling for other prior variables. The path analysis typically uses standardized data or a correlation matrix as an input. In terms of its practical application, the path analysis amounts to the following system of simultaneous equations, processed iteratively.[9]

$$\text{UNQAUD} = \beta_{11}\text{SGA} + \beta_{13}\text{RND} + \varepsilon_1 \quad (1a)$$

$$\text{TIMEPP} = \beta_{22}\text{SGAPP} + \beta_{24}\text{RNDPP} + \varepsilon_2 \quad (1b)$$

$$\text{VISITSPP} = \beta_{32}\text{SGAPP} + \beta_{34}\text{RNDPP} + \varepsilon_3 \quad (1c)$$

$$\text{PAGEVIEW} = \beta_{45}\text{TIMEPP} + \beta_{46}\text{VISITSPP} + \beta_{47}\text{UNQAUD} + \varepsilon_4 \quad (1d)$$

$$\text{ADSEEN} = \beta_{58}\text{PAGEVIEW} + \varepsilon_5 \quad (1e)$$

$$\text{CLICKS} = \beta_{65}\text{TIMEPP} + \beta_{69}\text{ADSEEN} + \varepsilon_6 \quad (1f)$$

$$\text{SALES} = \beta_{71}\text{SGA} + \beta_{73}\text{RND} + \beta_{75}\text{TIMEPP} + \beta_{76}\text{VISITSPP} + \beta_{77}\text{UNQAUD} + \beta_{78}\text{PAGEVIEW} + \beta_{79}\text{ADSEEN} + \beta_{710}\text{CLICKS} + \varepsilon_7 \quad (1g)$$

[9] The subscripts are written here in a manner consistent with other statistical tests. The standard convention for path analyses is for the first number to indicate the causal variable and the latter the dependent variable.

Variables ending in 'PP' are deflated by unique audience. All other measures are deflated by the total assets of the firm. Per-person measures are used for time spent online and visits as these are the variables reported on NNR and are more descriptive of the characteristics of a website's audience than total hours spent or visits would be. A summary of the predictions for the signs of these coefficients is given in Table 3.

As is the case with other statistical techniques, path analysis suffers from a number of limitations related to model specification. As mentioned previously, the most important among these is the fact that it cannot explicitly test for directionality in the relationships. The directions of the arrows in a path diagram represent the researcher's hypotheses regarding causality; however, the actual direction could be the reverse or the correlation could be spurious. In particular, if a variable specified as prior to another given variable is really consequent to it, it should be estimated to have no path effect. However, when it is included as a prior variable in the model, it could erroneously lead to changes in the coefficients for other variables in the model. Another important limitation is that techniques such as these often require substantially more data than single equation regressions in order to assess significance. The conventional wisdom in the literature is that the total number of observations should exceed the number of parameters tested by at least 10-20 times.

In addition, the coefficients in path analyses are sensitive to specification error when a significant causal variable is left out of the model. When this happens, the path coefficients will reflect their shared covariance with such unmeasured variables and will not be accurately interpretable in terms of their direct and indirect effects. Finally, the researcher's choice of variables and pathways represented will limit the model's ability to recreate the sample covariance and variance patterns that are observed in the data. Because of this, there may be

several models that fit the data equally well. Nonetheless, the path analysis approach remains useful in structuring relational data which is a good first step in understanding the intricate nature of the data involved.

2.5 Results

The description of the path analysis above focuses on the actions of web-activity-dependent firms. While the sample studied here includes a small number of observations for business models in which activity is not *ex ante* expected to be a substantial source of long-term revenues in their earlier descriptions (in the prior chapter), these firms are likely to prove exceptions to the rule. If firms are attempting to maximize revenue streams from multiple sources, primary or not, then website activity should translate into increased revenues for these companies as well.

Due to the use of partial regression coefficients in the path analysis, it would first be helpful to examine the overall correlations among the variables tested.[10] The correlations in Table 4 are sorted from left-to-right (top-to-bottom) based upon the particular variables' position in Figure 2. From Table 4, it can be seen that a number of pairs of variables are highly correlated, such as pageviews and advertisements shown (.74). This result would seem to support the need for a mechanism to control for possible endogeneity problems suggested by high multicollinearity in the data. From the organization of the data, it can be seen that these high correlations among the variables tends to fall as the number of hypothesized steps between them increases. These correlations are, therefore, consistent with the predicted effects in the

[10] In a perfectly specified model the sum of the effects from the direct and indirect pathways between any two variables would equal the correlation for those two variables.

last section (e.g., pageviews influences advertisements shown which in turn has some, albeit smaller, effect on click-throughs as a result of this intermediate step).

With respect to the accounting data, SG&A and R&D are mildly positively correlated (.24 and .41 when deflated by total assets and unique audience respectively). Interestingly, the two measures for SG&A are slightly *negatively* correlated (-.07), suggesting that each measure may provide different insights during testing. The two R&D measures have a small positive correlation (0.26). SG&A deflated by total assets is significantly related to all the other variables (negatively for the per person measures). Deflating by unique audience, however, the correlations are largely negative and significant except with the other per-person measures. The R&D measures show a similar relationship, although generally not as strong as for SG&A.

Table 5, Panel A, displays the results of the full path analysis described in Figure 2.[11] Regressing time spent online per person on SG&A and R&D, the former variable is not significantly different from zero and the effect of R&D is *negative* and significant (t-statistic of -2.33). The latter result is consistent with the interpretation that firm expenditures on R&D have been more focused on improving page delivery times (reducing time spent) than on the expansion of content and/or services (which would increase time spent). With respect to visits per person, neither SG&A nor R&D is significantly different from zero.

The results for SG&A are particularly surprising when one considers that it is common practice for firms to use advertising to increase the use of its products and services by existing customers (which would increase time spent and/or visits per person). However, increases in spending on SG&A should also increase the number of new browsers. If new users are, on average, less active than existing users, then failure to account for this indirect path would negatively bias the coefficient. To test for this possibility, the path analysis is re-estimated

including unique audience as an explanatory variable for both time spent and visits per person.[12] The resulting coefficients are negative but not significant and do not change the sign or significance for the other coefficients.

As suggested previously, regardless of the means by which SG&A and R&D improve website quality, unique audience is expected to increase in both of these measures. Results from Panel A are consistent with this expectation, as both measures are positively and significantly associated with unique audience. In addition, pageviews are found to be positively and significantly related to both time spent per person and the unique audience variable as predicted. Surprisingly, the coefficient for pageviews on visits per person is negative and significant. This result suggests that, once controlling for time spent per person, sites attracting more repeat activity over the course of a month may do so at the expense of depth of activity once browsers are at the site (i.e., through the use of bookmarks and/or greater experience with a site, users are better able to find desired content in a reduced number of pageviews).

Consistent with predictions, the direct effect of pageviews on advertisements shown is positive and significant. In turn, these advertisements are significantly positively related to click-throughs. Additionally, the direct effect on click-throughs of time spent per person is negative and significant, indicating that there are likely to be diminishing returns to increased time spent online as browsers become less sensitive to repeated advertisements. Finally, revenues are positively and significantly associated with SG&A and pageviews and negative

[11] Results are calculated using the PROC CALIS procedure in SAS using the RAM statement.
[12] The impact of the indirect effects on time and visits spent per person depends on the comparative magnitudes of the direct and indirect effects and on the ratio of new to existing browsers.

and (marginally) significant with unique audience. Contrary to expectations, advertisements shown and click-throughs are not significantly different from zero.

As suggested previously, all three of these measures could proxy for additional revenue opportunities. After modeling the (indirect) effect of SG&A on time spent, visits, and unique audience, it is likely that the remaining (direct) effect contains information regarding non-audience related revenues. Pageviews, on the other hand, should proxy for the ability of the firms to leverage their existing site activity through such actions as new ventures, alliances and more efficient targeting of content and promotions to audience members. The negative direct effect of unique audience probably controls for some un-modelled effects of the data or possibly serves as an indication of increased costs or decreasing benefits from attracting new browsers.[13]

The lack of significance on either advertisements or click-throughs may be the result of the smaller sample size and competing effects for these measures. The latter possibility is similar to problems experienced for visits and time spent per person. Advertising revenues include two major elements, the number of advertisements shown (or click-throughs) and the amount received per advertisement. If these two elements are negatively correlated, then the omission of the latter variable in the path diagram would result in a model mis-specification in which the coefficient on advertisements shown (click-throughs) would be negatively biased. Furthermore, if advertisements shown or click-throughs are negatively correlated with the rates charged, it is likely the result of individual users being shown more advertisements on each page (or altogether), thereby reducing the average value for each. This condition would also

[13] Newer browsers are likely to be among the slower adopters of the internet and technology in general and may not be as valuable an audience.

negatively bias the coefficients on unique audience and pageviews, which may explain the negative coefficient found on the former.

One possible method for detecting this hypothesized relationship would be to include interactive variables into the path analysis. The framework of the path analysis and the means by which it is calculated, however, makes the inclusion of such terms difficult. An alternative approach is to estimate the set of equations using per person deflation for all measures. If click-through rates are negatively correlated with the amount of advertising shown to a browser then the per person measures may be able to control for this.[14] In results not shown, the coefficient for advertisements shown per person is negative but not significantly related to revenues per person, similar to the asset deflated results above. On the other hand, click-throughs per person are positively and significantly associated with revenues per person (t-statistic of 3.73). Overall, these results are consistent with the interpretation that higher click-throughs lead to increased firm revenues, although the evidence of a negative effect from excessive advertising is inconclusive. With respect to the other variables, SG&A and pageviews retain their significance (the latter only marginally so) and R&D and reach are no longer significantly different from zero.

A second possible test is to regress the potential competing effects against revenues in a simple OLS framework. Assuming all revenues are generated from advertising, the ratio of revenues to total assets can be decomposed as follows:

$$\text{Sales/Assets} = (\text{Pageviews/Assets}) * (\text{Ads shown/Pageviews}) * (\text{Click-throughs/Ads shown}) * (\text{Sales/Click-throughs}) \quad (2)$$

[14] Since unique audience deflated by itself would result in a constant across all firms, this variable is replaced by the total audience deflated, "reach," measure to which it is nearly identical.

Taking the natural logarithm of each side and replacing the variables with suggestive notation produces the following result:

$$\text{Log(SALES)} = \text{Log(PAGEVIEW)} + \text{Log(EXPOSURE)} + \text{Log (CLKRATE)} + \text{Log(CPM)} \quad (3)$$

where SALES and PAGEVIEW are the asset-deflated values used in the previous tests. EXPOSURE reflects the ratio of advertisements shown to the number of pages viewed. CLKRATE corresponds to the conventional "click-through rate" definition used for internet firms (the percentage of advertisements that are clicked upon by the viewer). The final term, CPM, refers to the acronym usually quoted in the advertising industry for the cost per thousand viewers seeing an advertisement.[15] This final measure reflects overall conditions for the advertising market and is generally beyond the control of individual firms, after controlling for possible effects from the first three variables on CPM. Since any such relevant information would be contained in those variables and since CPM measures are only infrequently reported by firms, the final term is removed from the model leaving the following testable equation:[16]

$$\text{Log(SALES)} = \beta_1 * \text{Log(PAGEVIEW)} + \beta_2 \cdot \text{Log(EXPOSURE)} + \beta_3 * \text{Log (CLKRATE)} + \varepsilon \quad (4)$$

Similar to the per person path analysis, results (not shown) from this OLS equation indicate that pageviews and the click-through rate are positively and significantly related with sales (t-statistics of 12.40 and 24.08 respectively). The exposure measure is negative but not significantly different from zero (t-statistic of -0.43).

[15] In actual fact, the variable herein refers to a combination of the traditional CPM measure and the value placed on click-throughs on these advertisements.

[16] An initial examination of earnings announcements and quarterly statements indicates some firms report membership numbers and/or their cpms. The data, however, would be subject to a self-selection bias and the number of available observations appeared insufficient for testing purposes.

As mentioned above, in order to achieve interpretable results for the regression coefficients in a path analysis, it is customary to have at least 10-20 times as many observations as parameters. The ratio of about 20 for Panel A comes close to violating this condition. Therefore, it is uncertain whether the lack of significance for some of the coefficients above results from the reduction in observations imposed by requiring reported advertising data to be available. As a result of the insignificant findings on advertisements and click-throughs and as a check of robustness, Panel B shows a less restricted set of regressions conducted after removing equations 1e and 1f and reducing 1g to the following (resulting in an increase in the number of observations to 583 and a reduction of the number of parameters from 18 to 13):

$$\text{SALES} = \beta_{71}\text{SGA} + \beta_{73}\text{RND} + \beta_{75}\text{TIMEPP} + \beta_{76}\text{VISITSPP} + \beta_{77}\text{UNQAUD} + \beta_{78}\text{PAGEVIEW} + \varepsilon_7 \quad (1g')$$

The results on this larger sample (with a ratio closer to 40) are consistent with those reported in Panel A for the equations with time spent and unique audience. R&D now appears to be negatively and significantly associated with visits as well, providing further evidence that increased spending in R&D has been focused on streamlining the amount of activity necessary from audience members. For pageviews as the dependent variable, visits per person remains negative but loses significance and the other variables remain positive and significant. The direct effect of R&D for revenues becomes marginally significant, whereas, there is a loss of significance for the effect of R&D for revenues. The direct effect of SG&A on revenues remains positive and significant, while the effect of unique audience continues to be negative and significant. In light of these results, one possible explanation for the possible (direct) controlling effect for unique audience would be that firms with a smaller, more focused audience (most likely to be included in this larger sample but not in the one with advertising

levels restricted to being non-zero) are better able to leverage their audience through more targeted promotions, e-commerce initiatives, and the provision of premium "member" content and services.

To this point, the statistical tests have all been conducted with current SG&A and R&D explaining realized activity levels and firm revenues for the quarter. Evidence from the prior chapter of this dissertation and from other internet valuation papers have suggested that SG&A and R&D may be treated as investments by the investment community. If this interpretation is accurate, one should be able to predict future financial or non-financial data based upon these two variables. To examine this question, SG&A and R&D are replaced with their one-quarter lag values in equations 1a-c and 1g. In results not shown, the lag versions of R&D and SG&A are shown to be of the same sign and significance as the contemporaneous variables, consistent with the viewpoint that these variables do represent investments in future firm activity levels and revenues. Results are not materially different for any of the other variables in the other equations with the exception of an increase in significance for advertisements seen and loss in significance for unique audience in equation 1g.

In summary, results from this path analysis suggest that both SG&A and R&D have explanatory power over the website activity variables, consistent with the earlier contention in this dissertation that these expenditures represent investments in website quality. Evidence from the path analysis also indicates that both accounting and non-financial measures, in particular SG&A and pageviews, are significantly associated with firm revenues.

2.6 Expanded Testing

One limitation of any static, "levels", study is that the coefficient on any variable reflects the average effect of the data in question. As the internet develops and the technologies change, the relationships among these variables are likely to change with the scope of the firm (e.g., through network effects, increased efficiency or changes in browser demographics or habits) and over time, respectively. In addition, as described in Section 5, it is possible that some of the variables tested have competing effects which may confuse the results and cannot be easily modeled out, even within a path analysis framework. To examine the marginal effect of these variables, the complete set of regressions (1a-1g) are estimated using a changes specification, where the changes are defined as the difference between the reported quarterly accounting data and its one-quarter lag value.[17]

From Table 6, Panel A, it can be seen that, under this specification, neither SG&A nor R&D is significantly associated with either time spent online or visits per person. In addition, R&D is positive but no longer significantly related to unique audience, although the coefficient for SG&A and unique audience remains positive and significant. The lack of a coefficient for changes in R&D spending suggests that additional firm spending on R&D is most likely not associated with efforts to improve website activity. Overall the results for the changes specification are not as strong as those in the prior section. The results are, nonetheless, consistent with the interpretation that more primitive activity measures are relevant not only in the prediction of the other activity data but for the prediction of revenues as well (by way of pageviews). Additionally, while the evidence from R&D is mixed, SG&A shows strong

[17] Changes in the non-financial measures are similarly calculated as the difference between the reported activity in the last month of the firm quarter less the 3-month lag reported value.

evidence of being positively and significantly associated with firm revenues both directly and through its' influence on unique audience (which in turn increases pageviews).

To examine whether the relationships among the variables tested has changed over time, the set of equations for the reduced diagram (1a-d and 1g') are estimated for both the pre and post-crash period. In results not shown, R&D per person remains positive and significant in both time periods for both time spent and visits per person. SG&A, not significantly different from zero in Table 5, is now positively and significantly associated with visits per person in the pre-crash period and negative but not significant in the latter period. This result suggests that earlier firm expenditures on SG&A had been focused, at least in part, on increasing the user activity levels on their websites. With respect to unique audience, SG&A is positive and significant in both periods and R&D is not significantly different from zero in either period (most likely a victim of reduced sample sizes). Unique audience and time spent continue to be positively and significantly associated with pageviews for each time period and visits per person is negative and (marginally) significant in the later period. Similarly, the coefficients for both pageviews and SG&A with revenues remain robust to the time period selected. Unique audience remains negatively associated with revenues, although the significance is lost in the post-crash sample. On the other hand, the negative coefficient observed for R&D and sales appears to be isolated to the post-crash sample.

In addition to highlighting differences in the pricing of accounting and non-financial information over time, the prior chapter of this dissertation also stresses the importance of identifying and isolating different business models in order to reduce sample heterogeneity. In order to examine whether the type of business model employed by a firm influences the results in this chapter, two sub-samples of firms are separately tested: (1) portal and content-

community (P&C); and (2) (less advertising but still activity dependent) financial services and online retailing business models. Results for P&C firms are reported in Table 7. In the full sample of firms, SG&A per person is not significantly related to time spent per person. For P&C firms, however, this measure is positive and significant, consistent with these firms having a greater reliance on advertising and other promotional revenues which are generated directly from website activity levels of its users. In addition, unique audience (negative and marginally significant in the full sample) is positive but not significant in equation 1g. Other results are generally consistent with the full sample, with exception of a loss of significance on visits per person in equation 1d and R&D in 1g.

Table 8 shows the reduced diagram results for financial services and online retailing firms.[18] For equations 1a and 1b, SG&A is positive and significant and R&D is negative and significant for time spent and visits per person, respectively. The results suggest that these firms engage in promotional activities designed to increase site activity while trying to use technology to decrease the amount of time it takes users to conduct the transactions necessary for the firm's success (e.g., e-commerce sales or security trades). Consistent with this interpretation, SG&A is significantly associated with unique audience, whereas R&D is not significantly related to efforts to increase audience. Similar to the full sample, both time spent and unique audience are positively and significantly related to pageviews.

Unlike for P&C firms, financial services and online retailing firms have a negative and significant coefficient on visits per person for pageviews and would appear to be driving the similar results for the full sample. This suggests that the efficiency gains mentioned as a possible explanation are more prominent for these types of firms, perhaps from the benefits of

[18] Results for the full path diagram are not given as only 110 observations are available with advertising data which would result in an observation-parameter ratio of about 6.

having financial and/or credit information previously stored by these firms (e.g., one-click checkouts). Finally, results for the regression of revenues on these other measures also seem to indicate that financial services and online retailing firms may be responsible for the negative coefficients for R&D and unique audience and that SG&A and pageviews are positively and significantly related for this sub-sample as well.

2.7 Conclusions and Suggestions for Further Research

In the absence of definitive results regarding the pricing of net earnings in the earlier internet valuation literature, a number of papers have focused on revenues and other components used to calculate net income in order to explain firm valuations. To date, however, little empirical research has been conducted on how revenues are created by these firms. This chapter examines firm revenue creation, while addressing the potentially endogenous and multicollinear nature of the internet activity measures. This is accomplished through the development and testing of a path diagram (Figure 2), which specifies the route firms take from expenditures on SG&A and R&D through activity generation to revenue creation. This methodology allows for simultaneously addressing issues of factor identification and endogeneity. The focus on intermediate pathways permits separate testing of direct and indirect (through intermediate variables) effects. Its application is particularly appealing for internet firms, where understanding these relationships should provide a clearer understanding of what is driving the valuations of these firms.

The path analysis methodology presented in this paper could be easily adapted to other areas of accounting research. In particular, it could be used to improve measurement of other variables by decomposing components or effects of accounting and non-financial data. For

instance, evidence from this and other papers suggests that expenditures on SG&A and R&D might be regarded as investments and should therefore be capitalized. Path analysis could help address issues like these for all types of firms. It would allow for better amortization schedules by eliminating more transitory elements of these variables from those which should be capitalized. Similarly, path analysis could be used to develop better (possibly more recursive) accruals models by isolating the effects accounting variables have on each other. This could lead to better measures of non-discretionary versus discretionary accruals. Finally, this framework could be used to isolate and test the effects of more or less permanent components of earnings, while simultaneously giving researchers the ability to control for decisions made by managers on when to recognize such items as write-offs.

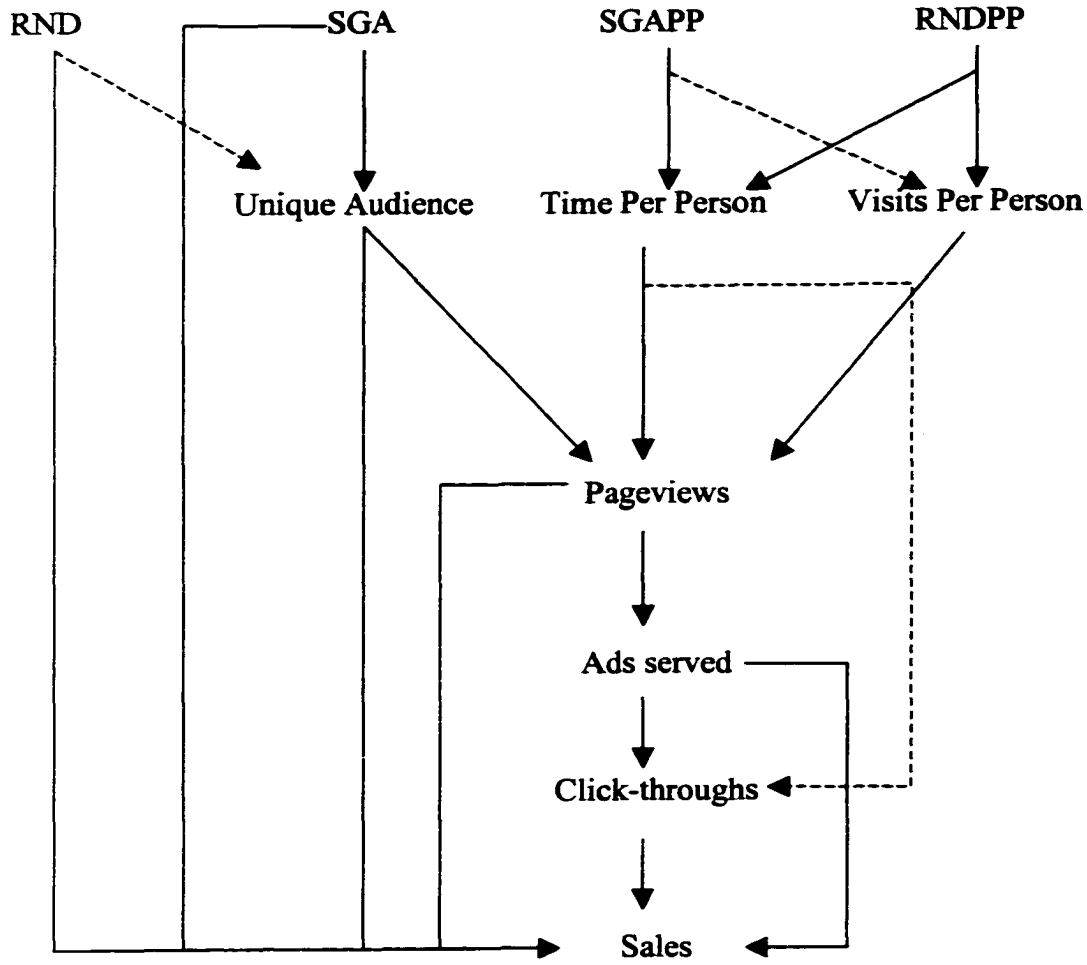
Empirical testing of the path diagram for internet firms provides evidence that firm expenditures on SG&A and R&D have explanatory power over both the generation of website activity and firm revenues. R&D per person reduces the amount of time a browser needs to spend online at a firm's website. SG&A, on the other hand, is positively and significantly related to time spent and number of visits per person for financial services and online retailing firms. It is also positively and significantly related to time spent per person for portal and content-community firms. Both SG&A and R&D, deflated by total firm assets, are positively and significantly related to unique audience. Finally, SG&A is positively and R&D is negatively and significantly associated with firm revenues, with the latter relationship appearing to be driven by financial services and online retailing firms.

The internet activity generated is systematically related to firm revenues as well. As unique audience and time spent per person increase so do pageviews. Pageviews have the direct effects of increasing firm revenues as well as increasing the amount of advertising seen.

This direct effect on revenues is most likely the result of the ability of pageviews to proxy for other, non-advertising, firm revenue opportunities associated with greater site activity (e.g., mailing lists and user profiling for portal and content-community firms and transactions for financial services or online retailing firms). Finally, while initial results for advertising data do not show explanatory power over revenues, alternative tests provide evidence that click-throughs are positively and significantly associated.

Figure 2

Path Analysis Diagram



Solid and dashed arrows both indicate the predicted direction of causality between any two variables.

Table 1

Sample Breakdown

Firms in initial sample:	332
Firms (observations) with complete accounting data:	317 (2049)
Firms (observations) also with data reported in the NNR audience database:	129 (583)
Firms (observations) with advertising data as well:	86 (373)

Table 2

Descriptive Statistics

Variable	N	Mean	Median	Std Dev	Min.	Max.
Market value	583	3215.90	464.38	12651.94	0.40	17140.2
Market-book	582	8.48	2.99	41.36	-45.64	900.01
Net Income	583	-66.86	-14.90	330.49	-5426.3	1178.0
Sales	583	80.01	17.10	406.06	0.00	6830.0
SG&A	583	38.06	21.38	54.63	0.00	425.00
R&D	583	4.86	1.50	12.06	0.00	159.72
Unique audience	583	3.03	0.96	5.14	0.10	44.56
Reach	583	2.36	0.78	4.24	0.07	37.38
Pageviews	583	69.89	13.87	177.91	0.27	1698.13
Time spent per person	583	0.19	0.15	0.13	0.02	0.86
Visits per person	516	2.04	1.75	0.99	1.03	6.24
Ad impressions	377	85.71	16.52	191.37	0.14	1821.05
Click-throughs	377	0.15	0.02	0.47	0.00	7.12

Table 3

Predictions for Direct Effects

The following table summarizes the predictions made in Section 4 for the direct effects of each accounting or internet-activity measure shown in Figure 2. Explanatory variables are given in the columns with the rows belonging to the relevant dependent variables. Variables ending in 'PP' are deflated by unique audience. All other variables are deflated by total assets. See Appendix 1 for further explanations of each term. A + (-) indicates an expected positive (negative) coefficient. A '0' indicates a variable that is being tested for which no prediction was made, while a '?' indicates a variable for which multiple, conflicting predictions are made.

	SGA	SGAPP	RND	RNDPP	TIMEPP	VISITSPP	UNQAUD	PAGEVIEW	ADSEEN	CLICKS
TIMEPP		+		?						
VISITSPP		+		?						
UNQAUD	+		+							
VIEWS					+	+	+			
ADSEEN								+		
CLICKS					-				+	
SALES	+		0				+	+	+	+

Table 4

Web Sample Correlations

Pearson correlations for accounting and internet usage variables deflated by total assets, with the exception of reach and per person variables.

Variable	SGA	SGAPP	RND	RNDPP	UNQAUD	TIMEPP	VISITSPP	PAGEVIEW	ADSEEN	CLICKS	SALES
SGA	1	*-.07*	**.24**	**-.08**	**.43**	**-.14**	**-.19**	**.28**	**.20**	**.23**	**.50**
SGAPP		1	-.03	**.41**	**-.29**	-.03	-.02	**-.26**	**-.21**	**-.21**	**-.24**
RND			1	**.26**	**.13**	**-.13**	**-.12**	.05	**.21**	**.14**	<.01
RNDPP				1	**-.21**	**-.13**	**-.15**	**-.18**	**-.13**	**-.13**	**.41**
UNQAUD					1	-.03	.05	**.76**	**.65**	**.50**	**.20**
TIMEPP						1	**.63**	**.28**	**.25**	.03	-.02
VISITSPP							1	**.19**	**.31**	**.25**	-.07
PAGEVIEW								1	**.74**	**.45**	**.19**
ADSEEN									1	**.54**	**.09**
CLICKS										1	**.12**
SALES											1

Variable definitions are given in Appendix A. Correlations shown in bold (italics) are significant at least at the 5% (10%) level.

Table 5

Path Analysis of Spending, Site Activity, and Revenues

Panel A: Full diagram path analysis results (n=377)

Dependent	SGA	SGAPP	RND	RNDPP	TIMEPP	VISITSPP	UNQAUD	PAGEVIEW	ADSEEN	CLICKS
TIMEPP		-0.010 (-0.16)		**-0.146** **(-2.33)**						
VISITSPP		-0.095 (-1.51)		-0.082 (-1.31)						
UNQAUD	**0.421** **(8.08)**		**0.167** **(3.20)**							
PAGEVIEW					**0.449** **(8.81)**	**-0.118** **(-2.31)**	**0.771** **(16.54)**			
ADSEEN								**0.743** **(20.12)**		
CLICKS					**-0.111** **(-2.12)**				**0.566** **(15.36)**	
SALES	**0.590** **(10.46)**		*-0.091* *(-1.73)*				*-0.104* *(-1.65)*	**0.159** **(2.62)**	-0.073 (-1.23)	0.019 (0.36)

Panel B: Reduced diagram results (n=583)

Dependent	SGA	SGAPP	RND	RNDPP	TIMEPP	VISITSPP	UNQAUD	PAGEVIEW
TIMEPP		0.027 (0.59)		**-0.142** **(-3.13)**				
VISITSPP		0.044 (0.96)		**-0.167** **(-3.66)**				
UNQAUD	**0.420** **(9.83)**		**0.031** **(0.72)**					
PAGEVIEW					**0.344** **(8.38)**	-0.066 (-1.61)	**0.773** **(20.28)**	
SALES	**0.544** **(11.81)**		**-0.126** **(-2.96)**				**-0.110** **(-2.15)**	**0.127** **(3.26)**

Table 6

Path Analysis for Changes in Accounting and Non-Financial Measures

Panel A: Full diagram path analysis results (n=302)

Dependent	SGACH	SGAPPCH	RNDCH	RNDPPCH	TIMEPPCH	VISITSPPCH	UNQAUDCH	PAGEVIEWCH	ADSEENCH	CLICKSCH
TIMEPPCH		-0.005 (-0.09)		0.055 (0.95)						
VISITSPPCH		-0.016 (-0.27)		0.045 (0.78)						
UNQAUDCH	**0.175** **(2.96)**		0.081 (1.37)							
PAGEVIEWCH					**0.483** **(8.39)**	-0.042 (-0.73)	**0.510** **(9.03)**			
ADSEENCH								**0.559** **(11.90)**		
CLICKSCH					-0.050 (-0.84)				**0.289** **(5.93)**	
SALESCH	**0.478** **(7.97)**		0.031 (0.53)				<0.001 (0.02)	**0.127** **(2.09)**	-0.019 (-0.31)	-0.059 (-1.02)

Panel B: Reduced diagram results (n=486)

Dependent	SGACH	SGAPPCH	RNDCH	RNDPPCH	TIMEPPCH	VISITSPPCH	UNQAUDCH	PAGEVIEWCH
TIMEPPCH		-0.035 (-0.78)		0.043 (0.95)				
VISITSPPCH		0.004 (0.09)		0.037 (0.80)				
UNQAUDCH	**0.125** **(2.67)**		0.060 (1.27)					
PAGEVIEWCH					**0.385** **(8.50)**	<0.001 (0.02)	**0.546** **(12.16)**	
SALESCH	**0.448** **(9.49)**		0.046 (0.98)				0.051 (0.85)	**0.111** **(2.62)**

Table 7

Path Analysis for Portal and Content-Community Business Models

Panel A: Full diagram path analysis results (n=216)

Dependent	SGA	SGAPP	RND	RNDPP	TIMEPP	VISITSPP	UNQAUD	PAGEVIEW	ADSEEN	CLICKS
TIMEPP		**0.182** **(2.53)**		**-0.167** **(-2.33)**						
VISITSPP		0.071 (0.99)		**-0.151** **(-2.09)**						
UNQAUD	**0.413** **(6.01)**		**0.198** **(2.88)**							
PAGEVIEW					**0.515** **(7.70)**	-0.020 (-0.30)	**0.740** **(12.03)**			
ADSEEN								**0.737** **(15.08)**		
CLICKS					**-0.205** **(-2.95)**				**0.486** **(9.84)**	
SALES	**0.529** **(7.12)**		-0.006 (-0.09)				0.120 (1.48)	**0.178** **(2.25)**	-0.109 (-1.44)	0.086 (1.29)

Panel B: Reduced diagram results (n=262)

Dependent	SGA	SGAPP	RND	RNDPP	TIMEPP	VISITSPP	UNQAUD	PAGEVIEW
TIMEPP		**0.173** **(2.31)**		**-0.204** **(-2.72)**				
VISITSPP		0.059 (0.79)		**-0.203** **(-2.73)**				
UNQAUD	**0.389** **(6.23)**		0.046 (0.74)					
PAGEVIEW					**0.475** **(7.80)**	-0.020 (-0.32)	**0.729** **(12.67)**	
SALES	**0.480** **(7.17)**		0.030 (0.48)				0.111 (1.50)	**0.182** **(3.26)**

Table 8

Other Activity-Dependent Business Models

Reduced diagram results (n=189) for financial services and online retailing firms. Results are not shown for the full set of equations due the low number of observations (n=101) relative to the number of parameters (18) leading to a ratio of about 6.

Dependent	SGA	SGAPP	RND	RNDPP	TIMEPP	VISITSPP	UNQAUD	PAGEVIEW
TIMEPP		**0.626** **(7.79)**		**-0.397** **(-4.95)**				
VISITSPP		**0.573** **(7.14)**		**-0.347** **(-4.32)**				
UNQAUD	**0.628** **(8.38)**		-0.110 (-1.47)					
PAGEVIEW					**0.249** **(3.85)**	**-0.149** **(-2.25)**	**0.884** **(14.20)**	
SALES	**0.582** **(6.62)**		**-0.216** **(-2.86)**				**-0.430** **(-4.52)**	**0.430** **(6.16)**

Appendix A

Variable Definitions

Historical accounting data is from the quarterly, June 2000 Compustat tapes.

SALES (data2)
TOTASS (data44) – Total firm assets.
SGA (data1) - Sales, general and administrative. When a firm reports no cost of goods sold this variable is COGS instead and this variable is reported as 'C.'
RND (data4)

From Nielsen//NetRatings (NNR):

UNQAUD - Unique audience as reported in the monthly audience measurement database.
VIEWS - Total pageviews as reported in the monthly audience measurement database.
REACH - Percentage of total estimated internet audience as reported in the monthly audience measurement database.
VIEWSPP - Average page views per person as reported in the monthly audience measurement database.
TIMEPP - Average time (in hours) spent per person as reported in the monthly audience measurement database.
PAGESPP - Redefined as VIEWS / UNQAUD since NNR rounds their reported variable.
ADSEEN - The number of ad impressions served by all the domains in a property, aggregated from domain level data reported by NNR.
ADSPP – TOTADS / UNQAUD
CLKRATE - The percentage of ad impressions clicked upon.
CLICKS - The total number of ads clicked upon, defined as TOTADS * CLKRATE for each domain and then aggregated to the property level.
CLICKSPP – CLICKS / UNQAUD

Advertising by sample firms on the internet is available as well but is not included in this study. Audience, views, and ad impressions are in millions. Rates are reported in percentages (10.3) rather than decimal form (.103).

Changes in the variables above have the suffix CH attached.

References

Amit, R. and J. Livnat, Winter,1988, Diversification, Capital Structure, and Systematic Risk: An Empirical Investigation, Journal of Accounting, Auditing & Finance 3:1, 19-43.

Bowen, R., A. Davis and S. Rajgopal, 2001, Determinants of Revenue Recognition Policies for Internet Firms, Working Paper, University of Washington.

Demers, E. and B. Lev, 2001, A Rude Awakening: Internet Shakeout in 2000, forthcoming in Review of Accounting Studies.

Hand, J., 2000a, Profits, Losses, and the Pricing of Internet Stocks, Working Paper, Kenan-Flagler Business School, UNC Chapel Hill.

Hand, J., 2000b, The Role of Economic Fundamentals, Web Traffic, and Supply and Demand in the Pricing of U.S. Internet Stocks, Working Paper, Kenan-Flagler Business School, UNC Chapel Hill.

Noe, T. and G. Parker, 2000, Winner Take All: Competition, Strategy, and the Structure of Returns in the Internet Economy, Working Paper, Tulane University.

Rajgopal, S., S. Kotha and M. Venkatachalam, 2000, The Relevance of Web Traffic for Internet Stock Prices, Working Paper, University of Washington and Stanford University Berkeley.

Trueman, B., M.H.F. Wong and X. Zhang, 2001a, The Eyeballs Have It: Searching for the Value in Internet Stocks, forthcoming in Journal of Accounting Research.

Trueman, B., M.H.F. Wong and X. Zhang, 2001b, Back to Basics: Forecasting the Revenues of Internet Firms, forthcoming in Review of Accounting Studies.

Wright, S., 1921, Correlation and Causation, Journal of Agricultural Research 20, 559-585.