UNIVERSITY OF SOUTHAMPTON

Faculty of Physical Sciences and Engineering
Electronics and Computer Science

Modelling Economic Bubbles: Is Web 2.0 Next?

Russell Newman

Thesis for the degree of Doctor of Philosophy

March 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

Faculty of Physical Sciences and Engineering
Electronics and Computer Science

Doctor of Philosophy

MODELLING ECONOMIC BUBBLES: IS WEB 2.0 NEXT?

Russell Newman

The Web 2.0 phenomenon has produced a number of technology companies that, in various rounds of venture capital funding, have attracted very high indicative valuations. Following these rising valuations, Investment Banks took an interest in the sector. However, while the companies concerned seem stable as private entities, their novel approach to business makes their financial characteristics difficult to predict.

Parallels are drawn between the 2001 dot-com bubble and the current Web 2.0 sector. This thesis highlights a dependency between modern web companies, and the established technology sector. It aims to identify the extent to which the contemporary technology sector (encompassing Web 2.0) has exhibited characteristics similar to those of the dot-com bubble.

To that end, this thesis identifies characteristics of modern and historic bubbles, and uses them to formulate a hypothetical set of indicators, in the form of a conceptual model. To determine whether these indicators exist in real data, a novel, repeatable statistical test is developed. It first identifies statistical heuristics representative of bubble circumstances, and then compares other periods to them. Thus, given sufficient data, any period may be tested.

Periods are analysed prior, during and after the dot-com bubble. During the dot-com bubble, consistently strong venture capital activity is observed, and linked to the growth in people using the Internet. This is indicative of the poor decision-making by investors, documented at the time.

In recent periods, patterns in venture capital investment describe an industry that is much more cautious than before, reducing the probability of the formation of a similar bubble. Looking at the past, this thesis observes investor activity that ‘caused’ the dot-com bubble as early as 1995-96, which raises questions about when the bubble started, and the lead-times on market collapses.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Russell Newman declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Modelling Economic Bubbles: Is Web 2.0 Next?

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Russell Newman

17th July 2015

Abbreviations and Terms

ODBC	Open Database Connectivity
OECD	Organisation for Economic Co-operation and Development
VC	Venture Capital

Chapter 1. Introduction

Developed nations have seen consistent growth in the proportion of their population using the Internet, since records began in 1990.

1999 marked six consecutive years of continuously high and increasing growth of the online population. Assured by this, many businesses began to transact business online. Various primordial social platforms were created, as people began to communicate online.

This attracted the attention of investment and venture capital companies, who noticed the level of growth in high-income nations, and the potential for more on a global scale. Companies that wanted to attract investment and funding had to exhibit a means of harnessing the upswing in the online population, and growing their user-base with it.

With a growing online population, some critics envisaged that the majority of trade in goods would move online, signalling the impending “death of the high street”. Share prices were driven upward by the hype surrounding web companies and the strong growth in the online population.

The rise in valuations of Internet companies soon became unsustainable, when investors and web start-ups realised that simply having a large user-base was not sufficient.

Furthermore, many companies had neglected to plan a means of generating income from their websites and services. The market crashed and corrected itself between 2000 and 2001. Valuations collapsed, and many web-based companies found themselves without the support they needed to survive.

However, the growth rate of people using the Internet continued unabated, with growth slowing in 2002 (see Figure 1).

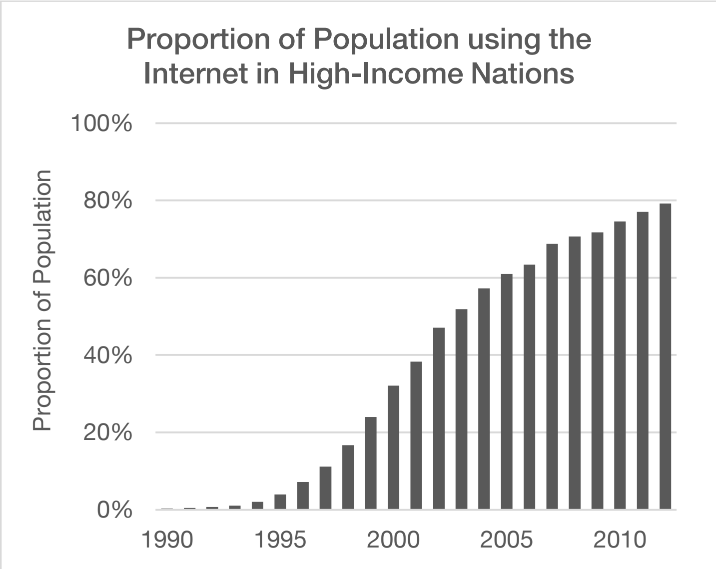

Figure 1 · Internet usage in OECD high-income nations

This was not the first time a new technology or trade opportunity had driven the development of an economic bubble. Examples date from as early as the 1630s.

Since its inception in 2005, the Web 2.0 movement of modern web companies has marked the next great surge in online activity. Web 2.0 has enjoyed a comfortable ecosystem in which to flourish: in 2005, 61% of people in high-income countries were using the Internet, and the proportion has grown in every subsequent year.

Web 2.0 has transformed the way people interact and share information (O'Reilly, 2005), and people spend more time online as a result.

Many Web 2.0 services continue to develop their social platforms, databases, and offerings. Some have attempted to monetise and diversify their properties (Falch, Henten, Tadayoni, & Windekilde, 2009). Some Web 2.0 start-ups have evolved into powerful companies, with great influence online (O'Reilly & Battelle, 2009).

This has led to speculation surrounding the potential public floatation of these companies, and the prices they might fetch. Some Web 2.0 companies have completed initial public offerings (IPOs), and floated on various stock markets worldwide.

1.1 Research Aims

Many onlookers, with both amateur and professional interests, have questioned whether the recent rounds of valuations and IPOs in the technology sector represent a repeat of the 1999-2001 dot-com bubble.

This research aims to identify, quantitatively, the degree to which companies involved with the web sector have exhibited a repeat of the bubble-like state that was observed during 1999-2001.

This is achieved by studying academic discourse and records relating to bubbles, analysing the drivers of their growth, and the enabling factors that have driven the growth of Web 2.0 companies. General characteristics of economic bubbles, and specifically those in the technology sector, were identified from these sources.

The characteristics were formalised into a conceptual model, which provided the basis of a hypothesis of how various metrics interact. A statistical technique was selected to test this hypothesis, and answer the research question.

Ultimately, this research aims to deliver a means of testing periods against circumstances known to be bubbles, using real data and producing quantifiable results.

Herein, a novel technique was developed and tested ‘by hand’ (i.e. exploiting machinery wherever possible, with some manual intervention). An ultimate objective to be kept in mind, albeit not in the scope of this work, was to automate the technique fully such that future market situations may be tested routinely by a machine.

1.2 Structure

This document follows a traditional thesis structure, with some additional chapters where it is helpful to explore detail.

Ch. 2. *Literature Review.*
Examines literature covering historic (1630s – 1840s) and contemporary technology bubbles, with the aim of identifying common characteristics.

Finds that bubbles typically form when investors lack experience of the sector in which the bubble forms. In the case of the dot-com bubble, investors demonstrably left their own sectors to invest in technology.

Notes that the Web 2.0 sector is equally “different” from previous technology, and that few (if any) investors will have experience of it.

Ch. 3. *The Future of Business in Web 2.0.*
Examines practices, business models, and technology trends of modern web companies, with the aim of demonstrating and emphasising that these companies are different to traditional technology companies, and that a mutual dependence exists with the established technology sector.

Finds that, in contrast to the dot-com bubble, some Web 2.0 companies have become established and dominant in their fields.

Ch. 4. *Model and Experimental Development.*
Explores modelling and simulation techniques, with the aim of arriving at an experimental design to answer the research question.

Creates a conceptual model of a market during bubble circumstances, using findings in the literature.

Using indicative data sources and the high-level design of this model, an experiment is designed in which market data shall be combined from many sources and periods. Slices of data *(bins)* for various periods shall be individually analysed using Factor Analysis.

One slice will establish a quantified benchmark of results for a period in time known to be a bubble, which can then be compared to subsequent or prior periods.

Ch. 5. *Methodology.*
Details the process of gathering, combining and analysing the data, with the aim of implementing the experimental design in the previous chapter.

Data sources for the various requisite data are compared, and a source selection is made for each of the three major strands of data.

Following selection, the data is acquired, sanitised (i.e. checked and made usable), and converted into a uniform format, such that it can be combined. The data is loaded into an SQL database, to ease the process of generating different renditions of the data. This enables a more rapid analysis, and eases the testing of different analytical techniques.

Finally, the Factor Analysis technique is applied to the data.

Ch. 6. *Results.*
Presents the results of the Factor Analysis, with the aim of describing the output but not interpreting it.

The raw output is described in terms of the underlying processes of the Factor Analysis, and presented in chronological/process order.

Ch. 7. *Discussion.*
Reviews the results in light of the literature, the conceptual model, and the research question, with the aim of answering the research question.

The research question is answered by first identifying characteristics in the results that concur with findings in literature, for the period of the dot-com bubble. Notably, that venture capital investment was easily obtained during the bubble, and that venture capital investors performed little due diligence upon their investments.

Secondly, the benchmark results are compared to those for other periods. Similar characteristics are not seen in periods following the dot-com bubble, so a similar bubble has not occurred.

Characteristics in the results for contemporary periods are explored. Results show that the nature of venture capital investment changed following the dot-com bubble; the four phases of venture capital investment behaved very similarly pre-bubble, but are clearly disparate post-bubble.

The period prior to the dot-com bubble is also reviewed. This reveals that the characteristics of venture capital investment that caused the dot-com bubble were observed as early as 1996, which raises questions regarding the lead-time of economic bubbles.

Ch. 8. *Conclusion.*
Emphasises key findings, the contribution, and future work, with the aim of exemplifying the work and logical continuations thereof.

This thesis presents a novel technique for gathering, combining and testing market data, to identify bubble circumstances in a technology sector. The technique shows that key characteristics of the dot-com bubble have not been observed subsequently.

The suggestion is made that, as this methodology was created with automation in mind, an automated reporting system could be created to provide results on a quarterly basis.

Several propositions are made for continuation of this work, to answer some follow-up questions:

- Was the dot-com bubble forming prior to 1995? (i.e. what is the lead time?)
- Is contemporary venture capital activity converging on another bubble?
- Did venture capital investment demonstrably leave other sectors during the dot-com bubble?

Chapter 2. Economic Bubbles

To support this research, literature has been consulted on historical and recent bubbles, theories surrounding speculation, the market for venture capital, and bubbles in the technology sector. By analysing a range of bubbles, rather than those in the technology sector, general bubble-principles are also identified.

2.1 Definition

Eatwell et al (1987) define speculative investors as those that are *"interested in profits from trading in the asset rather than its use or earnings capacity"*. As such, non-speculative investors are interested in gaining from the development of a product, or gaining from the earnings of such a venture.

Further to Eatwell's definition, Siegel (2003) adds that this implies the involvement of "momentum" investors, whose aim is to sell to other investors at a higher price, as quickly as possible.

Under this definition, almost any investment targeting a capital appreciation may be classed as a speculative one. Many of the investment activities examined in this section are speculative, taking place due to publicity, popularity or the "momentum" of a scheme. These factors generally contribute to increased risk of such investments.

In this context, an economic "bubble" is defined as a period in which speculative investment leads to an overvaluation of securities within a particular sector (Siegel, 2003). Economic bubbles may "burst" when investors realise that the industry within the bubble is not as profitable or sustainable as they first thought. At this point, valuations of the companies and securities involved descend rapidly to pre-bubble levels. Many bubbles are only categorised as such after they have "burst". Scientifically defining the term "bubble" is a subject of some debate, particularly defining a bubble before it has collapsed (O'Hara, 2008).

2.2 Historic Bubbles

A number of economic bubbles occurred during the last 400 years. Several have been studied for this review: the dot-com bubble, the Winchester Disk industry, the "Railway Mania", the South Sea Company and the "Tulip Mania".

2.2.1 The "Tulip Mania" (1634 - 1637)

The Tulip Mania involved the establishment of a futures market in the Netherlands, in which contracts of sale were traded for tulip bulbs before the end of the growing season. Tulips had to be traded as bulbs, as it was not feasible to transport the live plants.

Tulips had been introduced to the Western Europe in the middle of the 16th century and were becoming a status symbol of successful merchants (Garber, 1989). Until 1634, bulbs were grown and sold solely by professional growers.

After 1634, a broad group of amateurs also began growing bulbs. This led to the emergence of new and vivid tulip variants, which were particularly sought-after. Prices for bulbs of these novel variants rose with their popularity.

Each tulip plant may, over three to seven years, produce a number of additional bulbs and seeds. These may be removed from the plant and traded. The simple mathematics of purchasing one bulb, and being able to sell several spawned bulbs a few years later caused some to see the plants as a growth investment (Garber, 1989).

The seasonal nature of the tulip commodity meant that traders exchanged contracts on tulip bulb futures. That is, contracts were made to buy and sell (at a specified price) quantities of tulip bulbs that were not yet ready. Traders had to forecast demand, and purchase futures accordingly.

A formal futures market was established in 1636, making it easier to trade the contracts. The price of tulip bulbs subsequently rose steadily as traders assumed that wealthy foreign individuals would always purchase bulbs of the novel vivid varieties, regardless of the price. This assumption seemingly arose due to the popularity of bulbs amongst wealthy Dutch families.

Growers often created new variants, fostering a perception of fashion in tulip varieties and subsequent price rises. Traders believed they could always make a profit due to the perceived desirability of their constantly evolving product.

However, by February 1637 (three years after the introduction of novel varieties), traders realised that no-one would actually pay such great prices for tulip bulbs. The price quickly declined to its pre-bubble state (Garber, 1990). Many had purchased expensive tulip bulb futures, speculating that they could re-sell the spawned bulbs three to seven years later.

This bubble appears to have formed due to…

- **Excitement and hype surrounding the commercial viability of a new product, recently introduced to the continent.**
  Fuelled by continual product development, fostering the creation of new varieties, perceived as desirable.

- **Over-estimation of demand for the product.**
  Traders failed to anticipate the factors that limited demand, assuming growth would come from overseas as easily as it had domestically.

- **The notion that the product represented a viable growth investment.**
  In reality, the mechanics of supply and demand ended the bubble before any purchased tulips would have spawned new bulbs.

2.2.2 The South Sea Company (1720)

The South Sea Company was formed in London as an intricate method of providing funding to the government following the War of the Spanish Succession (Garber, 1990). The scheme was similar to one that was run in France by John Law. In 1705, Law published an economic theory that ultimately led to the establishment of national banks and paper currency, as opposed to physical gold and silver, described as “unemployed resources” by Law (1705).

Law also suggested that to raise finance for a venture, an entrepreneur need only make bold claims about their undertaking, and sell shares in the scheme at increasing prices. The revenue-generation of the venture was purported to eventually raise public confidence in the shares, stabilising the price (Garber, 1990).

The South Sea Company is an example of one such venture. Individuals holding £9.47m (approximately equivalent to £1.89tn in 2014[1]) of short-term government bonds were convinced to exchange them for shares in the South Sea Company, effectively writing off the government debt (Temin & Voth, 2004). In exchange, the government granted the Company a monopoly on trade via the “South Sea” (i.e. to South America) and paid the Company an annuity of 6%. The annuity was intended to provide the company with ample revenue to fund its ventures.

Until 1720, the Company issued shares so it could fund further government debt acquisitions, fulfilling its purpose to write-off government debt. At this point, the company began promoting the potential of its South American trade monopoly. This, and the government’s decision to allow the company to autonomously set share prices on stock issues, inflated share prices from £130 (per hundred shares) in January 1720 to £950 in July of the same year. This is shown in Figure 2, which is adapted from Temin & Voth (2004).

Many politicians were persuaded to invest in the venture, and were offered exclusive generous share schemes (Garber, 1990). This served to secure high-profile political support for the venture, further increasing publicity and increasing the share price.

[1] Calculated by the Bank of England, where records limited the calculation to start from 1750.

Figure 2 · Share price of the South Sea Company during 1720

Many shares in the company were sold as *subscription shares* to make the scheme feasible for a wider audience. Under the South Sea Company's subscription scheme, investors provided a down-payment and scheduled instalments in exchange for a specified number of shares. Investors received a fraction of a real share per instalment.

If the value of the company rose during the subscription, then the investor will gain shares worth more than he paid for them, and have the convenience of spreading their payments. The company will have gained the down-payment and the subscriptions, making this scheme instrumental in raising capital quickly and providing reliable, regular income in the medium-term.

However, Shea (2004, 2007) hypothesised that owners of subscription shares perceived them as a type of call option; they were obliged to pay the subscription, but could choose to default in the future if the subscription did not appear worth paying. This would be a cost-effective strategy when the share price of the company was sufficient lower than the price of the subscription as to make the subscription a more expensive means of purchase.

Furthermore, the Company had begun a programme of offering investors cash loans, taking Company shares (regular and subscription) as collateral. The cash was sourced from subscription share instalments. This type of arrangement would become particularly expensive for the Company if the value of the shares held as collateral devalued so much that the recipient of the loan was unable to afford or repay it (Shea, 2007).

Prices of shares in the South Sea Company began to fall during Q3 1720, when instalments were due for subscription shares. Many shareholders found themselves unable to afford the subscription, unless they sold their shares. Many defaulted, reducing cash flow from the subscription shares scheme. Insufficient records are available to attribute the decline to the loans programme definitively, although it is likely.

As an instrument to control inflating share schemes, the Bubble Act 1720 was introduced. This legislation prohibited companies from trading shares without permission from the government. Coupled with similar events occurring internationally, these factors caused the share price to fall rapidly (Smant, n.d.). As the price fell below that of subscription share purchase prices, many investments were rendered worthless. While the subscription share scheme effectively allowed the South Sea Company to set its own share price, the price of a share is representative only if someone is willing to buy at that price. Thus, the share price of the South Sea Company had to fall, due to a collapse in demand for them.

While the actions of the company directors may be regarded as irresponsible, in particular the methods by which shares were structured and marketed, no illegal activities took place. In contrast to the “Tulip Mania”, this bubble appears to have formed due to speculation surrounding the potential of a new venture (rather than a new product) about which the potential investors knew very little. This speculation was catalysed by people in the company aiming to increase the share price artificially.

2.2.3 The “Railway Mania” (1840-1846)

This bubble occurred in the UK during the 1840s, shortly after the Industrial Revolution. The first railways were demonstrated as an effective method of transporting passengers and goods, and appeared to be a key industry in a time when the country was increasing output of manufactured goods (McCartney & Arnold, 2003).

The Industrial Revolution had also created wealth for many middle-class families. The Bubble Act 1720, enforced following the South Sea Bubble, had recently been repealed. This enabled railway companies to sell shares without Government oversight, and many middle-class families were financially capable of making the investment (McCartney & Arnold, 2003). Many companies were provisionally registered with the prospect of building various railways. Figure 3 shows how few of these provisional registrations were actually completed.

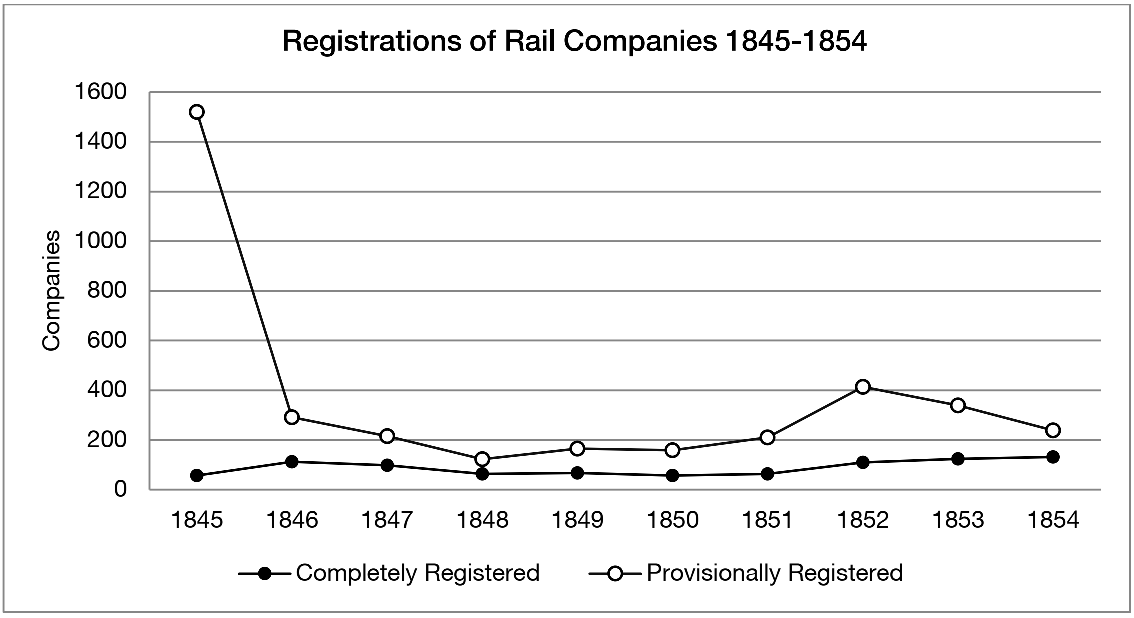

Figure 3 · Provisional vs. Complete Registrations of Rail Companies 1845-1854
Drawn using data from Taylor (2006)

Some of these companies sold shares in the project in exchange for a 10% deposit, whilst retaining the right to demand the remaining 90% at any time. The strong success of early railway projects, combined with the strong marketing for further rail projects, led many middle-class people to invest in these schemes, and sometimes more than they could afford.

Whilst liberated from the limitations of the Bubble Act 1720, Railway construction projects still had to petition the government to approve their route and permit land purchases (Odlyzko, 2010). The proposals of many companies were passed with minimal intervention because various MPs had invested in the projects themselves.

However, the difficulty in constructing railways was realised only when several rail construction projects actually started. This rendered many of the rail projects demonstrably unviable, so the speculation ended and many projects collapsed.

This bubble appears to have occurred due to investor speculation surrounding the unexplored potential for railways. The speculation was driven by rail companies and other investors, who believed that rail technology could transform transport whilst being profitable. However, in its infancy, the limits and returns of rail technology were unknown, so all the investors would have been operating in an unfamiliar environment, reducing their ability to evaluate adequately the rail projects prior to making an investment.

2.2.4 Discussion

The bubble situations described above occurred in different centuries and very different industries. However, they all developed due to investors having higher expectations of a new technology or product than was actually deliverable. The primary driver for their growth was speculation surrounding the companies or industries involved. In some cases, this led investors and those involved in the ventures to perform insufficient *due diligence* – taking time to *critically* analyse an investment or investment opportunity to find whether it is aligned with the investor's objectives and attitudes to risk (Pack, 2002). This concept is discussed further in section 2.3.4.

In some cases, the "inflation" of bubbles appears to have been due to corrupt and/or unethical activities by stakeholders. During the Railway Mania and South Sea bubbles, for example, companies ensured that certain government officials had a personal interest in their venture (Garber, 1990; Odlyzko, 2010), making it easier to gain government acceptance and support for their ventures. While this may be the case, they were legitimate operations according to the legislation of their time. In the cases of many historic bubbles, legislation was adapted to prevent similar events from occurring again (Garber, 1990; Taylor, 2006).

However, this is not the only factor affecting the growth of early bubbles. These ventures depended upon the showmanship of their promoters, their ability to energise potential investors about novel ideas, and their ability to inflate share prices.

It may be argued that an appropriate process of critical evaluation through due diligence should be the panacea to showmanship or speculation surrounding a venture. However, showmanship and marketing tactics may lead a potential investor to think less critically, and neglect some due diligence.

Nevertheless, in *all* the cases described above, showmanship and/or speculation surrounding the novel product contributed to the growth of a bubble. This indicates that investors have difficulty performing due diligence upon ventures that operate in areas outside of the investor's expertise.

Table 1 provides an overview of the bubbles and themes that have been discussed in this section. The technology sector bubbles are covered in the next section.

Table 1 · Overview of Economic Bubbles

Bubble	*Date*	*Location*	*New Product or Venture*	*Product or Venture*	*Investor Classes*	*Gov't Corruption*
"Railway Mania"	1840s	UK	Yes	Trade	Middle and Aristocracy	Yes
South Sea	1720	UK	Yes	Trade and Gov't Debt	All	Yes
"Tulip Mania"	1630s	Holland	Yes	Trade	Middle and Aristocracy	No

The economic bubbles detailed in this section appear to have exhibited positive feedback loops as a mechanism of their growth. Many investments were made in these schemes without due diligence because the share or product price was rising, and investors speculated that they could make a capital gain. However, the act of purchasing products or shares increased the unit price, which fuelled speculation amongst those who had invested. This in turn led to further investment, and represents a positive feedback loop.

Investment in these schemes represents an abstraction of involvement in the underlying businesses and products. Had investors considered the effects of supply and demand in the Tulip or Railway Manias, then the bubbles may have been less severe. Promotion, hype, and abstraction from the underlying product may have inhibited investors' rational consideration in these situations.

For the purposes of summarising the key drivers behind each bubble, the following stages of bubble development are defined.

- **Initial Development.**
  Covering very early development of the bubble, before it exhibits significant growth.

- **Growth and Performance.**
  The stage at which valuations of securities within the bubble rise beyond rational levels. A "rational" valuation may defined as one obtained through a calculative process of valuing the security, ignoring the influence of factors that may provoke speculative investment.

- **Collapse.**
  When the bubble "bursts", causing valuations to return to rational levels.

Table 2 shows a summary of the key drivers behind the three bubbles described in this section, according to the stages above.

Table 2 · Summary of bubbles described in this section

	Tulip Mania	*South Sea Co.*	*"Railway Mania"*
Initial Development	Speculation surrounding a novel product.	Speculation surrounding the prospect of lucrative trade.	Speculation surrounding the prospect of novel rail trade/transport.
Growth and Performance	Speculation surrounding prices for the product.	Schemes that permitted anyone to invest in the venture, and marketing.	Schemes that permitted anyone to invest in the venture, and marketing.
Collapse	Realisation that speculated prices were irrational.	Lack of actual trade; unsustainable financial situation.	Realisation of technological limitations, costs, and actual requirements.

2.3 Technology Bubbles

Many technology companies are (or have been) supported by venture capital companies (Hellman & Puri, 2002). Technology companies (particularly those producing software products) have a shorter development cycle. As such, technology start-ups may find themselves in competition with other start-ups before they have shipped a product, making investment in the sector risky (W. Sahlman, 1997). Venture capitalists supporting these companies typically expect a percentage of their portfolio companies to fail completely and a relatively small percentage to achieve significant success (Cochrane, 2005). The remainder are expected to achieve "success" (i.e. a return on investment) but not *significant* success.

Agility has been mentioned as a beneficial trait to be possessed by a technology start-up. Companies that fail to adapt within the fast-paced technology sector can find their products quickly become unpopular. This necessitates shorter release cycles, which are potentially more expensive. Musser & O'Reilly (2006) note this as a best practise for Web 2.0 companies, referring to the "perpetual beta" as evidence. A Web 2.0 service running a perpetual beta release cycle offers their users services that are not fully tested, on the understanding that some components may not work as expected. This enables new features to reach the market sooner, rather than undergoing a lengthy period of internal testing (Musser & O'Reilly, 2006).

The tendency to release new features as early as possible may be reflective of the "get big fast" strategy observed in original dot-com bubble companies. Under this strategy, companies aimed to get their offerings to market as quickly as possible, with the assumption that being first-to-market improves uptake by users (Oliva, Sterman, & Giese, 2003).

Competition and rapid development within the technology sector may have repercussions for investors: Sahlman and Stevenson (1985) suggest that the collective behaviours of investors, particularly venture capitalists, may be destructive for an industry. This is exemplified by the Winchester Disk industry between 1977 and 1984.

2.3.1 The Winchester Disk Bubble (1977-1984)

A “Winchester Disk” is a hard disk that follows the design and operating principles of the first Winchester Disk designed at IBM in the early 1970s. This was a spinning-platter hard disk, and is the principal design for contemporary spinning drives. Through research and development, the Winchester Disk industry rapidly improved the speed, capacity, and reliability of their proprietary offerings. The later production of industrial standards for these drives led to the de-facto design for magnetic hard disks, which are still used in some computers today.

Many companies in the Winchester Disk industry were start-ups and typically required large amounts of capital to run their R&D operations, without which they would quickly loose competitive advantage. Despite being costly start-ups, one Winchester Disk company was reported as a *significantly successful* investment, when the value of the company rose 31.75 times following its Initial Public Offering (IPO).

Enthused by these results, other venture capital companies were keen to see their own portfolio companies pioneer technological breakthroughs and produce similar performance at IPO. This potential for exceptional returns ensured that high levels of R&D investment continued in the sector until 1984 (W. A. Sahlman & Stevenson, 1985).

However, the rapid innovations in drive design and performance made it very difficult for any company to maintain market leadership for long enough to exploit their position. Venture capital companies eventually realised this and lost confidence in the industry, causing the collapse of many disk manufacturers. Of the 100 companies that existed during the bubble, five major manufacturers remain today.

Sahlman and Stevenson attribute the collapse of this industry to *Capital Market Myopia*; a phenomenon where investors become so engaged with the *companies* they support, they fail to realise the wider-reaching implications of their collective activities upon the *industry* (W. A. Sahlman & Stevenson, 1985). This was catalysed by the initial success story, and the perception that the product was destined to (a) become a popular necessity and (b) remain a high-price item.

Under this theory, the industry collapsed because competition between *investors* resulted in mutually destructive competition between their portfolio companies. Once again, the effects of supply and demand upon the entire market appear to have been overlooked.

2.3.2 Dot-Com Bubble (1995-2000)

The dot-com bubble began in 1995, shortly after developed countries gained access to the World Wide Web, and ended in 2000. It comprised many young publically traded internet companies, funded by venture capital, at a time when online trading was a novel concept. Many of these companies operated at a loss deliberately, providing services for little or no money to increase their market share. Once they had a suitable market share, they planned to exercise techniques to monetise it.

Hawkins observed that growth of the bubble prompted the publication of "many articles and books" about business models, a subject that had rarely been discussed or studied as extensively before in either the mainstream or academic press (Hawkins, 2004). This observation coincides with the desire to grow market share, exhibited by many dot-com start-ups at the time. Ovila (2003) describes the attitude of the time as a 'get big fast' strategy, in which retailers would attempt to be the first to the online market, and discount heavily in order to gain customers.

Further observation of academic discussion shows a tendency towards the notion that the technological capability of a company should be balanced with organisational capabilities (Wheale & Amin, 2003). This discounts the then-popular notion that new markets could be created online, given sufficient capital investment and the technological capability. The flawed notion that new markets could be created in this way arguably contributed to the speculation surrounding dot-com companies, and therefore the inflation of their share prices.

Ultimately, investors lost confidence in internet companies because share prices had risen without an accompanying growth in profit, signalling poor long-term investment prospects. The projected monetisation opportunities failed to materialise for these ventures, many of which were making great losses due to their expensive market share growth strategies (Wheale & Amin, 2003).

2.3.3 The Price and Quality of Venture Capital

A recurring theme in the technology bubbles reviewed above is the ready availability of venture capital funding for technology start-ups that have a novel or seemingly technologically superior product idea. This warrants a review of research into the varieties of Venture Capital available to such start-ups. In this part of the review, two types of venture capital are analysed: traditional venture capital and *corporate venture capital*.

Several studies have reviewed the quality and "price" to entrepreneurs of various venture capital firms, where "price" relates to the share of the company exchanged for investment and services from the venture capital company. They are discussed below.

Sahlman (1997) asserted that "*from whom you raise capital is often more important than the terms*". Hsu (2005) supports this, saying that venture capital firms have varying specialisations, industry connections, and levels of experience. He finds that prices for support and affiliation with a venture capital company fluctuate based on these factors, and especially upon connections to other relevant businesses that may support the entrepreneur.

A study by Ivanov and Xie (2010) found similar results. However, this study was subtly different – the focus was upon corporate venture capital. A corporate venture capitalist operates similarly to a non-corporate one, but is a subsidiary of a large corporation rather than being an independent organisation. The corporate venture capitalist is then able to provide both generic and domain-specialised services and support from the parent corporation. This type of relationship also changes the fundamental objective for corporate venture capitalists; rather than aiming for financial returns, they aim to achieve strategic benefits for their core business. Thus, a corporate venture capitalist is likely to evaluate candidate start-ups based on how they may benefit the corporation overall, rather than principally on potential returns.

It appears that a market exists for affiliation with venture capital firms, and that many prefer to support start-ups from a particular sector in which they have expertise or experience.

It has been noted that venture capital was a popular source of investment during some of the bubble periods. Research by Hsu (2005), as well as research by Ivanov and Xie (2010) shows that support from a venture capital company with relevant industry expertise makes a positive contribution to the success of a start-up. Likewise, a venture capital company operating in an unfamiliar sector may be detrimental to the performance of its portfolio companies.

Valliere and Peterson (2004) based their research upon interviews with 57 venture capital investors. They found that, during the dot-com bubble, some venture capital companies were prepared to depart from their sector of expertise to work with Internet start-ups, and that there was competition between venture capital companies to do so. Therefore, many venture capital companies involved in the dot-com bubble may have lacked sector experience or expertise and been unable to support the companies in their portfolio. Even experienced venture capital companies may have struggled to provide effective assistance to web based start-ups because the web, and the nature of doing business on the web, was such a novel and unexplored concept.

Investors' success criteria are informed by their objectives. Some may target value and growth in the form of capital appreciation, while others may target regular returns from their sustained investment in the company. Capital appreciation investors may be able to achieve their objectives in the short or medium term. Those targeting regular returns will take a longer-term view. Venture capital investors typically target value through capital appreciation, as their strategy involves exiting (i.e. selling) the investment when it is profitable to do so. Lower-risk, longer-term investors may target revenue payback.

Investors will attach differing qualitative values to the company depending on their objectives. For instance, a company that offers regular, steady returns may be of little value to an investor seeking rapid capital appreciation. As such, company valuations (qualitative and quantitative) are quite subjective.

In the case of the Winchester Disk industry (incorporating Capital Market Myopia), investors appear to have aimed for short-term capital appreciation in their portfolio companies. This was at a time when the industry required ongoing term support, to establish itself.

2.3.4 Due Diligence and Speculation

The previous sections found that bubble investors are often willing to leave their area of experience or expertise. This section introduces due diligence; a conceptual process that may be used to evaluate an investment opportunity to assist in making a rational, informed decision.

The financial sector safeguards itself from potentially unsafe investments through an evaluative process of analysis and due diligence, performed upon any new investment opportunity (Sudek, 2006) or potential acquisition. Factors such as the investors' appetite for risk and objectives are taken into account, and the potential investment evaluated against them. Due diligence is not limited to one particular technique; the techniques employed depend upon the nature of the investment or acquisition (Pack, 2002). An example technique for performing due diligence in technology operations is shown as an illustrative example.

The Gartner Hype Cycle is a high-level model used to describe and analyse technologies as they enter a market and become mainstream (Smith, 2003). When used in Gartner reports, it graphically shows the *maturity* of various technologies. "Maturity", here, is defined as the progression of a technology through the distinct stages of industry adoption, as defined by Gartner and discussed further below. Note that the cycle describes *technologies*, such as the concept of cloud computing, rather than *products* of the technology, such as Amazon's EC2 cloud computing platform.

According to Gartner, it is a tool that may be used by management and IT decision makers to perform due diligence upon IT investments which they do not fully understand (Linden & Fenn, 2003).

Gartner reports plot contemporary technology product offerings on this chart, to aid decision makers in deciding whether a particular technology is sufficiently mature or reliable for their business.

The cycle proceeds through several distinct stages, as shown in Figure 4:

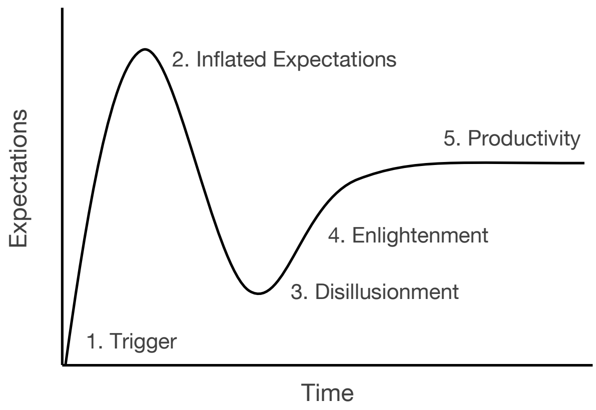

Figure 4 · Gartner Hype Cycle
Derived from Linden & Fenn (2003)

1. In the **“Technology Trigger”** phase, a new *technology* is discussed in the media, and it gains publicity.
2. The **“Peak of Inflated Expectations”** occurs when the technology is most visible in the media due to success stories and other coverage.
   At this phase, some early adopters may begin using the technology.
3. The **“Trough of Disillusionment”** follows, where the inflated expectations are not realised in practical application of the technology.
   At this stage, the technology may be updated or repaired to resolve issues found by early adopters.
4. The **“Slope of Enlightenment”** arises as people begin to understand and accept the limitations of the technology.
   At this phase, many interested parties consider using the technology, create implementation plans, and integrate it into their business.
5. The **“Plateau of Productivity”** phase begins when the technology and the resultant products gain broad commercial acceptance.

Despite its widespread use in Gartner reports, discussion surrounding the Hype Cycle discounts its use as a valid metric or instrument (Smith, 2003). Critics in the scientific community find particular fault with its naming as a “cycle”; Gartner does not describe the model as being cyclic, so these critics describe it as the Hype Curve. Furthermore, the placement of a technology on the curve does not appear to be backed by quantitative data (Fenn, 2010), and is instead an arbitrary and often unexplained placement made by a Gartner analyst.

During a discussion at the AYE Conference for Human Systems, Jerry Weinberg voiced criticism of the model, noting that the circumstances or action required for the initial “Technology Trigger” is not defined by Gartner, so it is impossible to know when the model should be applied to a technology (Smith, 2003). He notes that by analysing several of Gartner’s own annual report publications, technologies vanish and appear at various stages on the model over a period of years. Thus, Gartner presents the Hype Cycle as an instrument that can be beneficial to decision makers, but simultaneously disproves its efficacy by arbitrarily moving technologies on the curve year-by-year.

A further argument is that Gartner consistently place technologies directly *on* the curve; if the model was backed by quantitative data, a more representative placement would be possible, off the curve (Smith, 2003). The curve could then be plotted according to the data, and proven thereby. This deficiency suggests that technologies are forced into a position on the nominally-defined Hype Cycle curve, rather than the curve actually being a trend line that is the product of on-going quantitative analysis of various technologies.

In its current form, the cycle implies that a technology may always reach the “Plateau of Productivity” stage. In reality, it is feasible that a technology may never emerge from the disillusionment stage.

Nevertheless, the cycle does contribute value to Gartner reports, providing an executive overview of technologies for people who do not have time or resource to investigate the technologies themselves.

The hype cycle illustrates the nature of a bubble, albeit a bubble of inflated expectations rather than actual valuations. At the early states, while early adopters are using the product, expectations exceed the product’s capabilities. Expectations “crash” when users realise that the products does not deliver on them all. Finally, a correction occurs, which brings the expectations of the product in line with its capability. In all these stages, the difference between expectations and actual deliverable capability is analogous to the market valuation of a security, versus the actual underlying value.

This section has aimed to demonstrate that tools and techniques for due diligence are available, but may require an expert to interpret the results. In the Hype Cycle example, a company’s attitudes to risk and adoption of premature technologies may require a more specialised interpretation of the data. Furthermore, a specialist would be capable of analysing other technologies (analogous with other investments or opportunities) that lie in a similar place in the results, and draw further insights and conclusions from those associations.

Social Networks and Cloud Computing have featured on Gartner Emerging Technology reports and on hype cycle charts therein. The following section focuses on how these technologies have enabled innovation in Web 2.0 services, and how they are enabling new business models.

2.3.5 Discussion

In section 2.2, it was suggested that economic bubbles might form when investors leave their area of expertise to participate in potentially speculative investment activities in unfamiliar sectors. The review of technology sector bubbles confirms this; venture capitalists were willing to leave their area of expertise to become involved in the popular dot-com and Winchester drive industries.

The collapse of the Winchester Disk industry has already been attributed to the concept of Capital Market Myopia, in which the collective competitive actions of investors is detrimental to an entire industry ecosystem. This concept could also be applied to the 2001 dot-com bubble. Valliere and Peterson (2004) show that venture capital was obtained by many web start-ups, and that there was competition between venture capital firms to invest in web start-ups.

Given the outcome of the dot-com bubble, many companies in the sector would have failed due diligence tests. Competition between investors to gain exposure to such a sector is indicative of speculative behaviour.

The author frequently works with clients seeking investment for their new companies, and some who seek investment for their ongoing, established activities. While discussing with them about the nature of VCs and their own investors, and those they have turned down, these clients have described various investors as follows:

- “expensive”.
- “specialised in [sector]” (actual sector redacted).
- “wanting to break into [sector]” (actual sector redacted).

These comments provide anecdotal evidence that there exists a range of VC, in terms of price and sector specialisation. It also supports the notion that some VCs attempt to gain exposure in sectors they have never worked in. As such, VCs may pay more than a company is worth, in these situations.

Therefore, the concept of Capital Market Myopia and eventuality of investors leaving their sector of expertise should be taken forward as potential indicators of bubble scenarios.

Table 3 shows a summary of the two bubbles discussed here, and uses the same format as the Historic Bubbles Discussion in 2.2.4.

Table 3 · Summary of bubbles described in this section

	Winchester Disk	***Dot-Com***
Initial Development	Enterprise demand for products of this industry. Over-estimation of demand growth. Over-estimation of attainable prices.	Potential of trade opportunities over the Internet.
Growth and Performance	Competition between companies, by products and product development.	Development and discovery of Internet business models. Competition between inexperienced investors, trying to produce successful portfolio companies.
Collapse	Cost of product development. Short product lifecycle.	Poor profit due to failure or lack of business models.

2.4 Investors Understanding of Novel Technologies

Within the reviews of historic and more recent technology sector bubbles, a recurring theme appears to be the *lack of understanding by investors of the limits or costs associated with novel technologies*. This section explores that notion further, through analysis of literature and the case studies already detailed above.

The widespread use of new and un-tested technology in bubbles has made it difficult for investors to distinguish between companies likely to succeed and those likely to fail. This has been seen in the following bubbles thus far:

- Dot-com
- Winchester Disk
- Railway Mania

The lack of familiarity or understanding of these technologies inhibited investors from effectively evaluating their investment opportunities, leading to poorly researched speculative investment in the bubbles analysed. This is because the investors were unfamiliar with the technology, and the limits of novel technologies can rarely be known until they are met.

For instance, in the Winchester Disk bubble, the rapid pace of technological evolution meant that the position of market leadership was often held by a different company each week, in terms of cost, performance, quality and orders (W. A. Sahlman & Stevenson, 1985). Such rapid changes in a brand new sector will have limited the experience investors could draw upon.

A set of objectives, and plan for “exiting” the investment should form the basis of evaluation for even minor investment opportunities. In the situations described above, again due to a lack of domain familiarity, few investors understood when to exit from their investments. Valuations may have risen to, and beyond, investors’ “success” criteria. Investors chose to pursue further gains, instead of exiting their investments when “success” was achieved. Furthermore, investors had no benchmark or objective by which to measure when an investment had succeeded. By the same logic, the opposite should also be true; investors would have lacked the experience and knowledge to identify when an investment had failed, and should no longer be pursued.

If processes of due diligence had been correctly enforced, the lack of experience and domain knowledge would have reduced such investments.

2.4.1 Investor Attitudes to Technology during the Dot-Com Bubble

In section 2.3, the concept of *Capital Market Myopia* was linked to the Winchester Disk bubble. It may also be feasible to apply this concept to the dot-com bubble of 2001.

To test the applicability of this concept, Valliere and Peterson (2004) interviewed 57 venture capital investors who were active in the Internet sector during 2001, aiming to understand cognitive processes that might have led to the inflation of the dot-com bubble.

The investors were asked about their strategy when investing in unfamiliar sectors prior to the bubble. In these situations, the nature of investment success criteria was understood, but the investor may have had limited or no experience in the sector.

The resulting cognitive model, shown in Figure 5, involves “fundamental venture capital operations” of sourcing, screening, structuring, monitoring, and exiting.

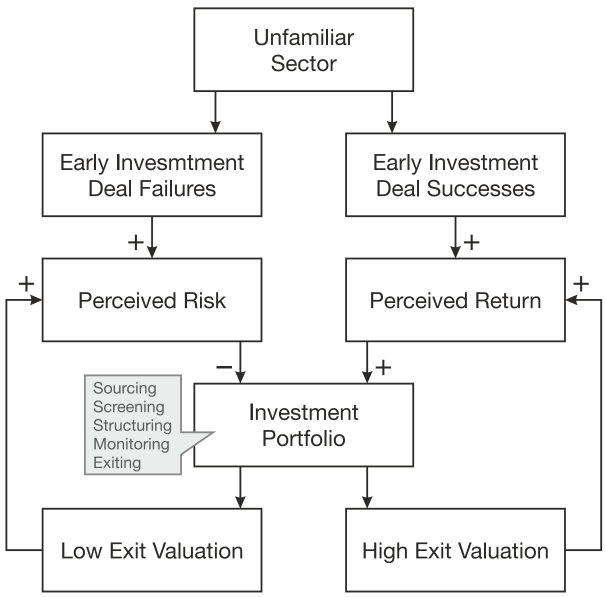

Figure 5 · Cognitive Model of investors operating in unfamiliar sectors Derived from Valliere & Peterson (2004)

While described as a “Cognitive Model”, no formal modelling technique appears to be applied. Nevertheless, the underlying meaning is clear; links between cells represent influences that either positively or negatively affect their target. For instance, a low exit valuation for an investment contributes positively to investors’ perception of risk.

The same investors were asked about their strategy during the bubble, in relation to the unfamiliar internet sector. It was found that:

- Investors had no way of knowing when an internet company was successful, because few (if any) had achieved resounding success at the time.
- Discussions between investors about the seemingly magnificent prospects of internet companies encouraged individual investors to invest speculatively, with little or no due diligence.

Following these findings, Valliere and Peterson (2004) acknowledged that, in relation to investors in the Internet bubble, “forces were operating that were not represented in the simple cognitive model of [Figure 5]”.

Figure 6 shows Valliere and Peterson’s subsequent development of the cognitive model, which incorporates changes to address these behaviours. The original model is shown in white boxes, and additions are shown in shaded boxes.

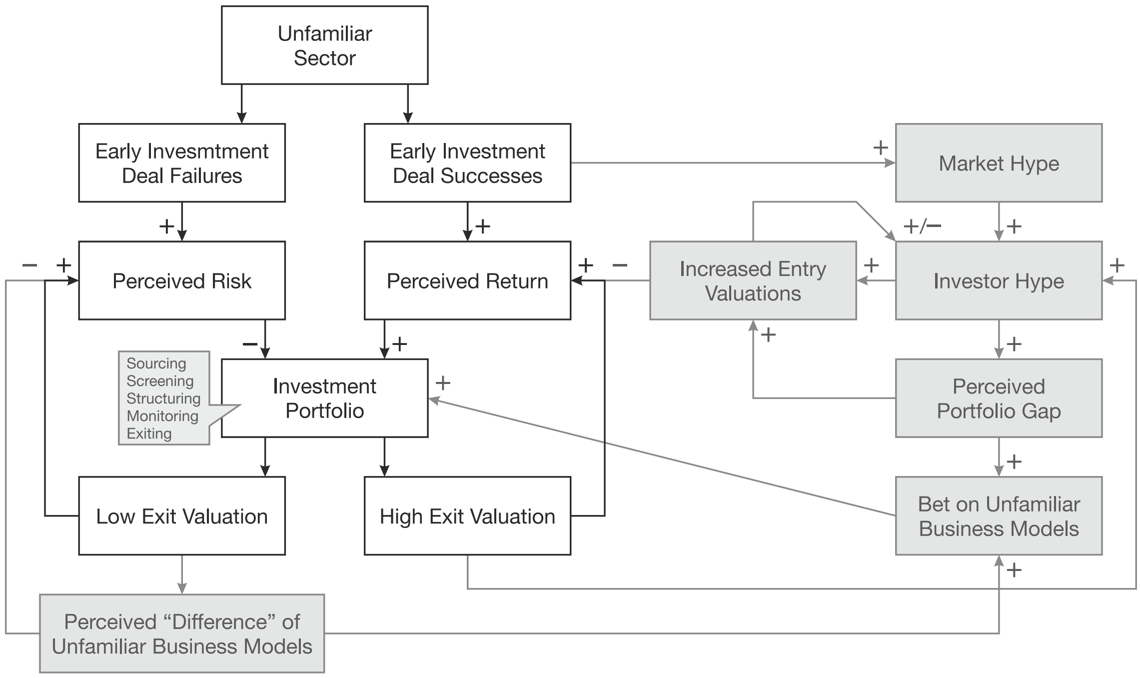

Figure 6 · Cognitive Model of investors during the dot-com bubble
Derived from Valliere & Peterson (2005)

A closed positive reinforcement loop exists within the shaded cells on the right side of the model. Under bubble conditions, this causes a rise in entry valuations for companies and speculative investment through the “hype” components.

Under normal circumstances, the links in the un-shaded model components result in a balanced process. That is, there is no perpetual positive or negative reinforcement.

The “Perceived Difference of Unfamiliar Business Models” cell is noted as decreasing “Perceived Risk” during bubble situations, which may appear contrary to expectations – unfamiliarity should surely *contribute* to a sense of risk. Valliere and Peterson attribute their model to a key finding in their interviews; that investors in a bubble situation assume the content of due diligence must change because they are unfamiliar with the business model or technology employed. Another factor behind this model decision was the assumption by bubble investors that the failure of one business model increased the likelihood of a working one being found in the remaining companies.

In a later paper, Valliere and Peterson develop their 2004 work and respond to questions raised by the community. Their work was constructed using prospect theory: an economic theory for analysing decision-making by individuals in situations that involve risk and reward (Kahneman & Tversky, 1979). The community asked whether alternative behavioural analysis methods might reveal alternative models, suggesting that it could not be applied to *all* investors during the bubble period.

Valliere and Peterson had focused only on venture capitalists in their first work, and agreed that another analytical framework may be more suited to the “*less experienced, but rational, internet investors*” (Valliere & Peterson, 2005). This means that the formation of the dot-com bubble may be attributed to more than just venture capitalists, which may be true when companies grew and some began to file for IPOs.

2.5 Summary

This literature review has examined previous economic bubbles, some of their causes, and identified characteristics in each one.

Over time, investors have learnt the procedure and importance of performing due diligence upon investment opportunities. Nevertheless, when hype and speculative investment surrounds a company or market sector, the findings in this chapter suggest that investors may be willing to reduce due diligence and take unmeasured risks in sectors with which they lack experience.

The untested nature of business models employed by modern web companies means that many investors are fundamentally unfamiliar with the sector. These companies have the potential to generate revenue in novel ways through their user-bases and data, but few current business models have proven this successful. The corporate venture capital investors appear to view larger web companies with optimism, due to the size of their user-bases.

The web sector is developing rapidly due to technological progress supporting the phenomenon of Web 2.0, which has two effects. Firstly, unfamiliar business models are emerging, which are difficult for investors to evaluate. Secondly, there are fewer barriers to entry and overheads in the Web 2.0 sector, which would otherwise reduce possible financial gains.

The following chapter analyses the practises and business models of companies in the Web 2.0 sector, and describes a mutual dependence between small, modern web companies and established technology giants.

Chapter 3. The Future of Business in Web 2.0

The main objective of this research is to determine to what extent the ecosystem of modern web companies represents a bubble. To that end, this chapter explores the drivers of modern web companies, their practises, and business models. It describes these recent changes in terms of implications for the sector as a whole.

The term “Web 2.0” was defined by Tim O’Reilly in 2005, amongst other definitions, as sites and services that rely upon the generation of content by their users, as opposed to editors or dedicated content creators (O’Reilly, 2005).

O’Reilly’s list of indicator characteristics for a Web 2.0 service is instrumental in defining the sector:

- Services, not packaged software, with cost-effective scalability.
- Control over unique, hard-to-recreate data sources that get richer as more people use them.
- Trusting users as co-developers.
- Harnessing collective intelligence.
- Leveraging the long tail through customer self-service.
- Software above the level of a single device.
- Lightweight user interfaces, development models, and business models.

Whilst novel in 2005 and exhibited by just a handful of ground-breaking services, many of these characteristics have become commonplace in services and software. Such rapid adoption has been possible due to a series of low-cost emergent technologies, which are explored later in this chapter.

3.1 The Social Web

Following O’Reilly’s definition, Web 2.0 was later characterised by online “links between people” in addition to the web’s links between documents that characterised it up to that point (Murugesan, 2007).

Social web services have flourished by enabling people to connect with not only friends, family and colleagues, but with events, interest groups, companies, brands and other entities. Enabling people to connect with friends and other entities enables them to receive multimedia updates from those connected entities. It also permits people to freely associate with any other entity they wish, perhaps publically, and build a persona or profile.

Berners-Lee (2010) argues that social web services tie their users into their product offerings. That is, users are prevented from using their existing data on other social web services with ease, thus creating closed “silos” of users’ social data. This is achieved by ensuring that users’ data on a particular service can only be seen and utilised from that service, and not exported to other services. Users are encouraged to depend upon the service as a means of social interaction, as moving one’s data to another service becomes prohibitively difficult.

This "user tie-in" may also be generated as a side-effect of the level of use of a social web service and the breadth of facilities that it offers. For instance, customers are less likely to switch to a competing social web service if all their friends and connections are using their current one. Given that a social web service cannot operate profitably without users, generating tie-in is an important component of social networks' business models (Berners-Lee, 2010).

Berners-Lee made these statements in 2010. In the time since, many social web services have developed methods of interoperating and integrating with other services. These are often manifest in:

- **Federated Authentication.**
  The ability to log into Service B using one's credentials from Service A (Lenk, Klems, Nimis, Tai, & Sandholm, 2009).

- **Cross-service Publishing.**
  Allowing Service B to publish or re-publish some manner of update on Service A, *on behalf of* the user, with their consent (Witek & Grettano, 2012).

These abilities represent some opening of the silos that were described by Berners-Lee in 2010. However, they may also be seen as generating further tie-in. The use of Federated Authentication increases users' dependency upon the originating service, which serves as their federated identity online. Cross-service publishing is another means of amplifying the effect of user-generated content, by allowing it to be re-published on multiple services.

The nature of marketing for both products and services has adapted to capitalise on the network effect of user-generated content. Marketing campaigns now typically direct people to campaign materials on social media sites, whereupon they are asked to post an update to their connected friends regarding the promotion. Such updates appear as a recommendation from a friend, rather than an unsolicited recommendation from the product owner. This encourages people to trust the materials, and pay attention to it (Harris & Dennis, 2011).

3.2 Cloud Computing Services

The usage level of many social web services is dependent upon cycles of human activity and real-world events. The peak usage level of a social web service, for a given geographic area, may be many times higher than the trough (Stone, 2008).

This creates problems for enterprises running social web services; the differences in demand between peak and trough traffic can be great, and providing financially effective capacity for both peaks and troughs can be challenging.

In the past, companies requiring high serving capacities often addressed this problem by employing a farm of dedicated storage and computing servers, capable of handling a given level of peak traffic. However, the full capacity would only be used at peak times, resulting in increased costs during normal operating load.

Furthermore, such great investments sometimes fail to perform during the most critical of times; companies have experienced traffic peaks above the designed capacity of their server farm, and have subsequently been unable to serve some or all of their users.

A couple of examples exemplify this problem.

1. **UK 1901 Census Website Launch.**
   On 2nd January 2002, data from the 1901 census of England was released online (BBC, 2002c). Demand for the site was high, as it provided a tool that anyone could use to find information on their ancestors, with ease. News of the launch was also widespread in the media.

   The first three days of release saw an average of 32 million visitors per day, which was 27 times higher than the designed capacity (BBC, 2002d). Being unable to cope with this peak of demand, the site failed completely, and had to be taken offline.

   Eight months later, in September 2002, the site had been improved to cope with the higher levels of demand and was undergoing testing. However, the media and public interest in the site had passed, and the site never saw the same levels of popularity (Sfetcu, 2014).

   Under-provisioning of server capacity effectively ensured that this project would never meet its objectives, at a time when predicting demand and providing capacity was difficult.

2. **Nectar Loyalty Card Launch.**
   When this loyalty card scheme launched in 2002, it was backed by email and TV marketing, with exposure to an estimated 10 million households. Those signing up online were given bonus loyalty points, in an effort to reduce demand on telephone and postal registration systems (BBC, 2002b).

   Despite media coverage of 10,000,000 households, the Nectar website crashed on the first day of service with just 10,000 visitors per hour (BBC, 2002a).

   The objective of the Nectar Card was to amalgamate individual loyalty schemes for different retailers into a single scheme. An expensive marketing campaign raised public awareness sufficiently. However, the failure of the website meant that Nectar were unable to issue loyalty cards to thousands of people at the most opportune time to do so.

These examples characterise 2002 as a poor year for IT reliability. However, this is representative of the era. Prior to cloud computing offerings, there was no way to provision sufficient capacity for major product/service launches.

The problems created by capacity provisioning leaves both Technology/Information and Finance Directors in an embarrassing position, where their companies can be incapable of effectively performing their core business operations at the most opportune time. Thus, traditional server farms represent a large capital expense that must be paid even before a service has begun functioning, and may not be capable of scaling adequately to demand.

Storage and computing capacity can now be automatically purchased from a Cloud service reseller as it is required, providing a cost-effective solution to the problem of usage spikes. Capacity may be allocated almost instantaneously at peak times, and then released when it is no longer needed (Buyya, Yeo, & Venugopal, 2008). Companies are billed according to what they use, usually by the hour (Armbrust et al., 2010).

Customers using Cloud services are offered the advantages of the server capacity, and spared the tasks of purchasing and maintaining the physical hardware, land, cooling, security, networks, and power. These tasks are the responsibility of the Cloud service, liberating customers from such overheads and setup costs (Armbrust et al., 2009).

Some of the most popular Cloud services are run by familiar companies. In addition to their other operations, Amazon, Microsoft, IBM, and Google all run competing Cloud service platforms. Some companies, such as Rackspace, perform only Cloud service operations and have no other business offerings.

Cloud services remain a growth area (Gartner, 2013).

For companies operating popular social web services, this removes the overhead of a server farm, converting it into a flexible cost that is adjusted according to the usage of (and therefore the revenue generated by) the product. The cloud service profits from this arrangement by signing many customers to their cloud computing platform and sharing the capacity between them (Armbrust et al., 2009).

Cloud services have lowered the barriers to entry for web start-ups. Fledgling companies benefit by spending less of their limited capital on costly land, hardware, and connectivity. These capital expenses, which are typically large and paid up-front, are converted into monthly payments, scaled according to usage of the product.

Some modern web companies effectively resell cloud processing or storage capacity, adding value in the methods of use and application they offer to customers. One such example is Dropbox[2], which resells capacity on the Amazon S3 storage Cloud service, adding value by providing a file synchronisation and sharing service through their software. This is marketed to end-users as a Cloud-based solution to access and share one's files anywhere. A version of the service with more fine-grained security features is marketed to businesses.

[2] www.dropbox.com – a Web Service that automatically synchronises files between computers and mobile devices.

Cloud services emerged based on these two enabling factors:

1. **The Commoditisation of Virtualisation Technology.**
   Traditionally, one physical computer has run one operating system at a time. Virtualisation adds an abstraction layer that allows one physical computer to run multiple operating systems simultaneously, on *virtual* hardware. This allows customers to share hardware (and all the associated costs) transparently.

   A Cloud computing customer can create and start virtualised servers in seconds, boosting their computing capacity according to demand. The provider of virtualisation services must balance the load of virtual servers across physical hardware, in a way that optimises performance, availability and profitability (Foster, Zhao, Raicu, & Lu, 2008).

2. **A Global Improvement in Bandwidth.**
   Cloud services represent a move from websites storing high-volume content on their own servers to a distributed model (hence the "Cloud" name), where storage of high-volume content is outsourced to a Cloud service.

   Cloud services usually own datacentres located at or near population centres. By hosting content in the correct physical location, rapid delivery by the shortest and cheapest route can be achieved. Crucially, this helps avoid the use of costly trans-continental links. Bandwidth availability in the target market is paramount for ensuring that customers have a reliable connection to the service.

   Such high-volume content often requires high bandwidth. The bandwidth of residential and business Internet connections has improved in many countries, enabling many more people to consume "rich" multi-media internet content. Conversely, this means social web services can make their offering available to a greater number of people in a variety of geographic locations (Telegeography, 2009).

3.3 Interdependence in the Sector

Social Web Services and *Cloud Computing Services* have been introduced and described quite separately. They both exist within the Web 2.0 ecosystem, but represent different portions of the sector, and have different customers.

Cloud computing services provide storage and computing capacity to social web services, in a business-to-business (B2B) relationship. In turn, social web services use this capacity to produce a product that is utilised by consumers, in a business-to-consumer (B2C) relationship. This is illustrated in Figure 7.

Figure 7 · Web 2.0 supply chain

During the dot-com bubble, companies purchased and maintained their own computing capacity. Thus, the illustrated supply chain relationship has emerged during the Web 2.0 phenomenon.

Many established technology companies have positioned themselves at the B2B end of this supply chain, preferring to provide infrastructural cloud services to companies that operate closer to consumers. Microsoft, Cisco, Oracle and IBM have all adapted their product offerings in this way.

This chain highlights a dependence between companies at the two tiers of supply. Cloud computing services depend upon the survival of social web services, to ensure that there is sufficient demand to make their offerings and business model profitable. As such, any funding or cash flow problems at the social web services are likely to negatively impact the cloud computing services.

3.4 Is Web 2.0 Entering a Period of Stability?

As of 2009, *"powerful players"* had emerged in several Web 2.0 service markets according to O'Reilly & Battelle (2009). O'Reilly (2010) further reinforced this in his keynote to the 2010 Health 2.0 conference. This could perhaps mark the end of competition for market dominance in these sectors, and a stage of stable dominance by a few companies.

This emergence of powerful companies represents a contrast to previous technology bubbles, which collapsed before powerful players could emerge. For that reason, "stability" is defined here as the emergence of a dominant Web 2.0 company in a particular service sector of Web 2.0.

Following O'Reilly's acknowledgement of emergent dominant companies, articles appearing in the financial press suggest that stable and dominant Web 2.0 companies are becoming a target of speculative investment by increasing numbers of investment banks and private investors (Braithwaite, 2011; Das, 2011a, 2011b; Dembosky & Demos, 2011; Gelles, 2011a). Given the demonstrable risks of investing in young web-based companies, it is understandable why these investors may be focusing on only the companies emerging as stable and dominant.

The same articles show that Facebook and Twitter are two of the largest private Web 2.0 companies targeted by investment banks (Braithwaite, 2011; Das, 2011a; Dembosky & Demos, 2011; Gelles, 2011a). The investment proposals detailed therein have become more generous over time, in terms of amount invested for the company share received (Braithwaite, 2011; Gelles, 2011a). This is illustrated in Table 4 using Facebook as an example. Also shown are recent valuations of Facebook, according to public trading.

Table 4 · Valuations of Facebook, indicative and otherwise

Source	*Trading*	*Type*	*Date*	*Valuation ($bn)*
Secondary Market[3]	Private	Indicative	24/08/2010	33.7
Goldman Sachs[4]	Private	Indicative	04/01/2011	50.0
Google Finance	Public	Market Cap.	13/07/2012	65.7
Google Finance	Public	Market Cap.	12/07/2013	52.9
Google Finance	Public	Market Cap.	11/07/2014	135.3

Facebook is largely supported by advertising revenue and micropayments from social applications, while the majority of Twitter's finance appears to still be sourced through venture capital (Sharespost, 2011). As explained in Section 2.3.3, the desired outcome scenario for a venture capital investor is that the portfolio company shares increase in value and can be profitably sold (Cochrane, 2005). This places Facebook ahead of Twitter in terms of company development, with a negligible dependence upon venture capital support and income from multiple revenue streams.

During the dot-com bubble, many companies offered their products as a loss-leader to incentivise people to use the service. They subsequently depended upon venture capital until a revenue-generation strategy was devised. Some Web 2.0 services also offer free services. The following section examines the novel business models that are emerging within the social web, and investigates whether they are more sustainable than models employed by companies in the 2001 dot-com bubble.

3.5 Emergent Business Models

Various business models are in use by Web 2.0 companies. One such model is "Freemium", a portmanteau of "free" and "premium". In this model, a base service is offered at no price and enhanced services and products (so called "premium" facilities) are offered for a fee. This model is particularly suited to companies that can transform their users' individual data into a larger dataset that is useful commercially (McGrath, 2010). The web (and cloud computing in particular) offers sufficient economies of scale to run a company under this model. LinkedIn is one such example, which allows users to create CV-like profiles for free, but charges for access to the various tools which can assist in recruiting and head-hunting, by analysing users' underlying CV data (Johnson, 2010).

[3] Gelles (2011b)
[4] Braithwaite (2011)

Social “marketplaces” are an alternative business model, which enable users to make purchases within the context of the network. Such purchases may provide a product to use within the network, or within a plug-in (or *app*) that runs within the network. This has been demonstrated extensively on Facebook: players of casual games in Facebook can often purchase items in-game that confer an advantage to them.

3.5.1 Advertising

The delivery of advertising online is one of the oldest and most prevalent methods of sustaining an otherwise revenue-less service (Crain, 2014). Advertising companies run *platforms* that enable website *publishers* to embed adverts dynamically on their webpages. Publishers are paid per impression (the displaying of an advert) and/or per click.

Advertising platforms tailor the ads they display automatically. Prior to the widespread use of social web services, advertising platforms would select adverts for a particular webpage based largely upon the content of the webpage or site.

The data-centric nature of modern web-based services and the prevalent use of user accounts enables advertising platforms to create a detailed profile of their users’ interests, likes, dislikes, habits and other characteristics (Guha, Cheng, & Francis, 2011).

As such, adverts can now be additionally targeted according to the actual *person* that will view them.

This is facilitated by the widespread offering and usage of individual user accounts, so that most traffic on various websites can be traced back to a particular user. This has created widespread concern and debate on the meaning of privacy on the web, and how it conflicts with the advertising industry’s business models (Hoofnagle, Urban, & Li, 2012). The growth of smartphone usage, with the associated increase in GPS/location information, has enabled advertising platforms to capture even more profiling data, such as users geographic locations and commonly frequented places (Dhar & Varshney, 2011).

Advertising platforms compete on two main factors:

- **Accurate Targeting.**
  Advertisers typically plan and identify the target market(s) for their products and advertising materials. By profiling users, advertising platforms enable advertisers to display their materials to the audience that is most likely to respond to it.
  Advertisers often prefer to use advertising platforms that offer more fine-grained targeting, as this increases the effectiveness of each advertising campaign, and may reduce costs.

  For instance, the LinkedIn advertising platform enables advertisers to target LinkedIn users by profession, seniority, job role, capabilities, and interests. This makes it easy to place adverts in front of receptive individuals, particularly those likely to have the authority to purchase the product.

- **Prestigious Placements.**
  Advertising platforms typically have a collection of websites (a "Network") where they may automatically display adverts. The characteristics and popularity of placements available to an advertising platform may appeal to certain advertisers. For instance, an advertising platform may market itself as having placements on a number of the most-visited websites. Alternatively, placements on websites of a particular focus, or with a particular type of audience.

Users of social web services contribute data to their advertising profile through many of their actions. Even passive users (i.e. those who just read) generate valuable data, as the items they view reveal their interests. Users who post updates and contribute to the service may generate a richer profile. This places social web services in a powerful position, as they have particularly rich datasets and advertising profiles.

Furthermore, social web services may employ the network of user connections in their service to deliver targeted adverts. For instance, adverts may be displayed to a particular group of users because another user in their network responded positively to something online.

The use of Federated Authentication, mentioned in section 3.1, also assists in developing an advertising profile. By tracking the websites from which a user makes use of Federated Authentication, the social web service may identify what sort of online content the user consumes.

Both Twitter and Facebook have large user-bases, with rich advertising profiles automatically generated by the day-to-day activity of their users. This enables fine-grained advertisement targeting, which makes both sites popular for advertising placements (Facebook, 2012; Twitter, 2011).

Google's *AdWords* Advertising Platform exploits all the factors described above. AdWords allows advertisers to target users based upon keywords from Google searches, geographic location, interests, and demographic factors such as age, gender, and parental status. Adverts are then into Google Search result pages, and as placements on other sites according to the targeting specifications of the advertiser.

Google have maintained their position as a successful online advertiser by:

- Exploiting their position as one of the most popular search engines.

- Encouraging users to create an advertising profile, by offering free accounts with email and other services. The Google Privacy Policy, which is applicable to all Google products, states:
  "Our automated systems analyse your content (including emails) to provide you with personally relevant product features, such as customised search results, tailored advertising and spam and malware detection." (Google, 2014)

- Developing various products and services that facilitate the growth of rich advertising profiles. For instance, a single Google account ties together data from email, YouTube, search queries and results clicked, and the Google+ social network.

3.5.2 Microtransactions and Social/Casual Games

Microtransaction platforms are payment systems created to allow users to make purchases, typically for small amounts. These systems have been implemented by various online services, social and otherwise, to enable products or services to be purchased with small amounts of real currency.

Purchases may be made directly through the social service, or through any other product that integrates with the online service. For instance, Apple offers a microtransaction platform through the retail of Apps and In-App Purchases. When a purchase is made, Apple automatically takes 30% of the revenue, and the remainder is passed to the developer of the App. Microtransaction platforms typically keep prices low to incentivise purchases and pool transactions together when billing them to a credit or debit card to minimise transaction clearing fees.

Casual video-gaming gained popularity in the mid-2000s. Casual video games are typically playable in repeated short periods. This facilitates minutes of play at a time, in between other activities. The lack of complexity means the games require little time to learn. These games are often provided free to the user, and offer microtransactions in the game to drive revenue.

They may sometimes require the player to wait an amount of time while operations in the game complete (with the option of expediting progress via a microtransaction). This differs from regular PC or console gaming, where participants may play for hours, after taking time to learn how to play (Kuittinen, Kultima, Niemelä, & Paavilainen, 2007).

Social web services have built APIs that allow games and apps to be run inside their site, in the context of the social network. For instance, Facebook allows games and apps to be run “in Facebook”, and to exploit the user’s list of friends, and other attributes.

Social games gained popularity on social networking sites because the underlying network allowed players to play the games quickly, with or against friends, and to see one another’s progress. Prior to this, social video gaming remained the purview of PC and console gamers as mentioned by Kuittinen et al (2007). Arguably, social networking websites have catalysed the spread of these games, leading to the development of social games.

A number of companies publish such games and Radoff (2011) has identified several observations on these phenomena, notably:

1. **Exploitation of Social Connections.**
   Games can exploit data in the social network to instil competition, teamwork, or co-operation between players. This may be in the form of status updates on the network, notifying other users of friends' activities and interactions in-game, and enticing them to join in.

2. **Speed, Competitive Advantage or Aesthetics – at a price.**
   Players may typically purchase an item in-game that provides them with a competitive advantage, or speeds up play, through a real-world currency transaction. Some in-game purchases confer purely aesthetic adaptations, allowing players to apply *"premium"* customisations to their characters. Such aesthetic additions may adapt the appearance of an avatar, provide rare clothing, distinctive millinery works, or accessory items for the player's environmental context (e.g. a dog for their farm).

The combination of social networks and casual gaming has produced several highly-performing companies whose products depend upon social networking services (Vascellaro, 2008). The parent social network typically benefits by automatically taking a proportion of any monetary transaction made through their microtransaction platform.

Companies producing social network games are highly dependent upon the underlying social graph of their parent network, but at the same time capable of generating revenue at little actual cost. While the games themselves are free and thus loss-leaders, the one-off cost of creating purchasable items is low compared to their sale price. Once the production of the original game is paid for, only operating costs need to be met. Furthermore, the games are designed to scale with the growth of the social network, enabling millions of users to engage globally.

Zynga is a company that produces various free-to-play social network games that are tied-in to the Facebook platform. Their revenue is based upon retailing purchasable extras for the games, according to the model described above.

When a venture capital company proposes a private investment deal with a company, this gives an indicative valuation of the company. In the case of Zynga, investments made by venture capital companies valued the company at $15bn-$20bn (Dembosky & Demos, 2011; Forro, Cauwels, & Sornette, 2011). However, at floatation in late 2011, the IPO price valued the company at $9bn (BBC, 2011; Google, 2011).

The value at floatation is based upon what the market is likely to pay for a share of the company, rather than the indicative private valuations that Zynga had experienced before. This may indicate that the venture capital companies overvalued Zynga during its time as a private company. Speculation surrounding Facebook and the social networking sector could have been the cause of this overvaluation.

Since social network games emerged, the rise in smartphone and tablet ownership has created a change in the way such games are delivered and played. Users can now be reached easily through the app stores on their mobile devices, allowing play on a mobile device at any time, instead of having to play through a social networking website.

These modern adaptations still encompass the social aspect of the gaming – users are frequently prompted to post updates about their progress in the game to their social networking services. However, since the games are now based on mobile devices instead of a social networking service, microtransaction revenue will pass through the company that owns the app store (i.e. Apple or Google), and no longer through the social networking site.

3.6 Summary

This chapter described Web 2.0 companies, and explored their practises and business models. Several companies might have become emergent dominant players in their field. This section finishes with an exposition of the revenue generation methods employed to raise these companies to their dominant position.

While it is easy to focus upon several high-profile companies providing social web services, this chapter explored how the technology sector as a whole has facilitated modern web businesses, and modern business models. Traditional vendors of computing equipment are adapting to the demand for Cloud services, and attempting to create differentiated product offerings.

Cloud service providers now depend upon the companies they permitted to develop – those with new business models and services.

The following chapter develops a conceptual model based upon these findings and those in the literature review. The model aims to explain links between key metrics, and results in an experimental design that will answer the research question.

Chapter 4. Model and Experimental Development

Following the literature review and discussion, an indicative model was constructed to explore the domain further. This chapter first outlines some modelling and simulation techniques. The fundaments of these techniques are then used to construct an initial model.

4.1 Modelling and Simulation Techniques

Simulating a system enables analysis of various situations by modelling them, over time, within a computer program (Balci, 1998). This enables the exploration of outcomes by adjusting parameters within the model.

A model is a “representation of an event and/or things that [are] real (a case study) or contrived (a use case)” (Banks, 2009).

A simulation may be run multiple times, to investigate how differing conditions alter the outcome.

The word “system” denotes what, from the real world, is being simulated. A system may be broken down into its components (such as people, machines, and resources) and processes or relationships that connect the entities.

The “system model” is the simulated representation of the real-world system. System models are designed and built in such a way that a computer can perform calculations upon them and effectively run a simulation (Banks, 2009). When building a model, an important consideration is how the entities interact and affect one another (Cellier, 1991).

The actual *method* for designing the model depends upon which simulation technique is employed. Regardless of the technique employed, a model designer must determine the level and areas of detail for the model, known as *scope*.

4.1.1 Scope

Scoping is the process of deciding which components should be included in a simulation models, and at what level of detail, and which components should be left out, simplified or abstracted (Sokolowski, 2009).

Building a model of a real-world system means that some concessions and assumptions must often be made for the sake of feasibility or expediency. For instance, a certain component of a real-world system could be implemented fully in a simulation, resulting in a theoretically accurate simulation, at the cost of including many elements to represent that component. Alternatively, the same component may be modelled using a simpler implementation that is perhaps abstracted or makes some assumptions, without compromising the accuracy of the rest of the model (Cellier, 1991).

A model designer must choose exactly how much detail should be expressed in a model; detail may lead to a more accurate model, but at the same time create potentially unnecessary work in situations where a less detailed model would be sufficient.

The breadth of the model must also be scoped. A system model may contain modelled representations of many external entities that influence the core elements of the simulation. Including more of these may increase accuracy, again at the expense of time and design complexity. Alternatively, excluding more of these may result in a simulation model that meets the objectives, and saves time.

4.1.2 System Dynamics

System Dynamics is a method of quantifiably modelling and simulating complex systems. It was developed by Jay Forrester at the MIT Sloan Management School, which was founded to exploit a fusion of engineering tools and techniques with traditional management. It was initially developed as a means of identifying the factors that make up the success or failure of a corporation or group of people (Forrester, 1961). System Dynamics models were originally processed by hand, but the technique was later adapted to take advantage of computer processing.

To build a System Dynamics simulation, a design progresses through two distinct stages: building causal loops and translating these to stocks and flows. It is possible to skip the first stage, but this would also exclude valuable analysis of the system, which can lead to a higher quality simulation.

Sterman's 2001 work on System Dynamics provides a succinct description of the methodology, which is used to inform the following sections.

4.1.2.1 Causal Links and Loops

At this first stage of design, a causal loop diagram is designed showing the various system variables and influences between each of them (Sterman, 2001). The causal loop diagram does not quantify any of the variables, but does denote whether variables positively or negatively affect each other. This is done through *feedback links*.

A feedback link is represented by an arrow. It signifies that the variable at its origin affects the variable at its destination in a positive *or* negative way relative to changes at the origin.

- A **positive link** means that as the item at the origin increases, the item at the destination may increase as well. If the origin were to decrease, the item at the destination would also decrease.

- A **negative link** means that as the item at the origin increases, the item at the destination may decrease. If the origin were to decrease, the item at the destination would increase.

Figure 8 shows a simple causal loop diagram modelling the causes and effects between variables of birth rate, death rate, and living population. The *links* are represented as thin arrows, with +/– indicating the nature of the link.

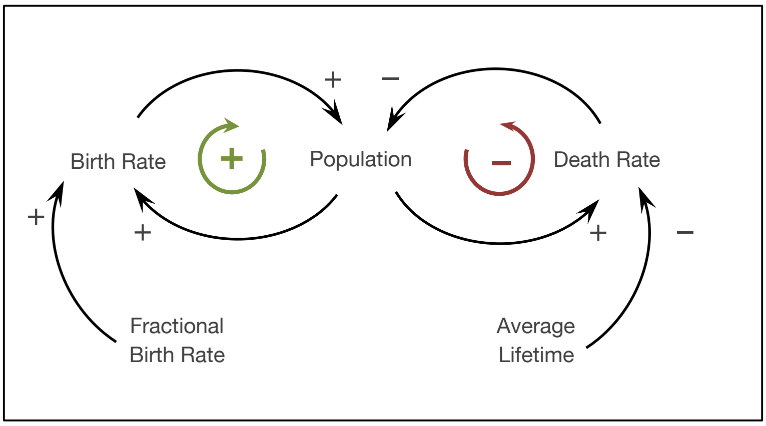

Figure 8 · Population System Dynamics Model
Derived from Fontaine et al. (2009)

Using these constructs, loops commonly feature in System Dynamics designs. A loop emerges when feedback links connecting variables form a closed path. Loops may be categorised according to whether they result in a positive or negative affect after all components of the loop have been evaluated.

Loops are usually highlighted on the diagram in the following way. Highlighting loops is purely denotational, and does not confer changes to the behaviour of the system.

- A **positive (reinforcing)** loop is denoted as a circular arrow containing a + sign (shown in on the left in Figure 8).
  It emerges when, after evaluating all the links in the loop, there is an overall positive effect upon all variables within. This can lead to exponential growth of variables, if not mediated with other links.

- A **negative (balancing)** loop is denoted as a circular arrow containing a – sign (shown in red in Figure 8).
  It emerges when, after evaluating all the links in the loop, the variables inversely affect each other.

For instance, in Figure 8, the positive loop on the left would cause an exponential Population increase, if the Death Rate variable were greater than the birth rate. The negative loop on the right will result in an eventual balance between Population and Death Rate, assuming the Birth Rate does not change.

Once the diagram is complete, it may be evaluated to ensure logical and theoretical integrity.

At this stage, the magnitude of the influences between variables is not important, and nor is the value of each variable. The aim of the exercise is to produce a reliable cause and effect structure for the model, according to reality (Coyle, 1999). Next, the model is adapted with stocks and flows, which enable variables to be quantified.

4.1.2.2 Stocks and Flows

The causal link diagram indicates direct and inverse links between variables, but no magnitude of effect, and no value of the variables. To overcome this problem, the model may be adapted with stocks and flows (Sterman, 2001).

A "stock" is a variable that is annotated with the quantity present at any given time. The quantity may grow or shrink over time, depending upon how it is connected to other stocks in the model.

A "flow" connects stocks and is annotated with a flow rate. This represents the rate of transfer from one stock to the other. Items will always travel through the Flow, as fast as the rate permits (which may be zero) and provided items are present at the source stock.

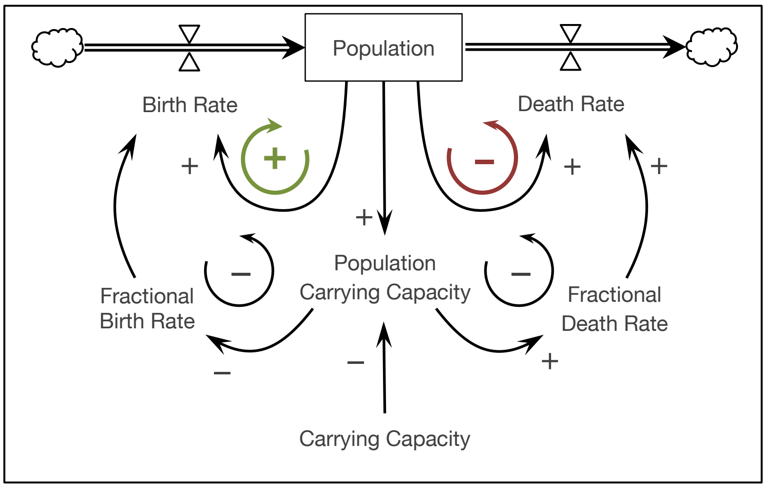

Figure 9 · Stock and Flow adaptation of Figure 8
Derived from Fontaine et al. (2009)

Figure 9 shows a stock and flow adaptation of Figure 8. The principal model variable, Population, is converted to a stock. The positive and negative loops identified in Figure 8 are shown in their new positions within the modified diagram. Some adaptations have been made to the variables. For instance, *carrying capacity* of the overall species has been added, as has the carrying capacity of the Population shown in the stock. The carrying capacity represents the population that can be feasibly sustained by the environment.

Forrester defines two categories of system (Forrester, 1968);

- In **open systems**, the outputs of the system have no effect upon its inputs, meaning the performance of the system does not result in any changes for the system.
- In **closed systems**, the outputs of the system do affect the inputs, so the performance of the system determines how it will behave in the future. The population models in Figure 8 and Figure 9 are closed systems.

4.1.3 Summary
This section has described System Dynamics as a method of simulating complex systems that change over time. The process of developing a System Dynamics simulation model involves stages of initial design on paper to ascertain the fundamental layout and connections of the system. Systems are later converted to a fully quantified digital model, containing logical operations that define how components interact.

4.2 Initial Model Development
The previous two chapters presented an overview of findings from previous bubbles, and some indications of trends in the Web 2.0 sector. This section aims to produce a conceptual model based upon those findings, built using the principles of System Dynamics. The purpose of this model is to consolidate findings, as a means of developing a quantified technique for answering the research question.

4.2.1 Model Specification
The model was built to satisfy the research question for this work; to find the extent to which the Web 2.0 sector has represented a bubble during its lifetime.

To achieve this, the model had to represent bubble scenarios. As the most recent bubble analysed, literature on the dot-com bubble was used as the theoretical basis of the model. However, the model also had to encompass other scenarios that may lead to a bubble, such as those discussed in Chapter 2.

The following two bubble characteristics, identified in Chapter 2, were taken forward as indicators the model should support.

- **When a "rapid increase in investment volume [is] not based on a corresponding increase in market knowledge or corresponding decrease in investment risk".**
  Source: (Valliere & Peterson, 2004) and section 2.3.3 (The Price and Quality of Venture Capital)

  "Investment volume" may be measured directly against the records for a public company, or by the quantity of venture capital supplied to a private company.

  "Market knowledge" cannot be directly quantified. When interpreted in terms of investors' experience and expertise within the market, it could potentially be measured in terms of the age of the market and the number of successes or failures.

  "Investment risk" cannot be directly quantified. Market analysts typically use historic data, amongst other metrics, to estimate potential risk in the future. However, no such historical data exists for start-up companies, so other factors must be used to generate an indicator of survival, profitability and, therefore, risk.

- **Continued <u>investment</u> despite the lack of a business model or <u>market dominance</u> (and therefore profit).**
  Source: Section 2.3.5 (Technology Bubbles: Discussion)

  "Investment" may be measured in the same way as above.

  The presence of a business model is difficult to quantify, but the effective level of profit generation heuristically represents the existence of a business model.

  "Market dominance" may be measured in terms of people using the company's product, relative to the total number of people available in the market.

4.2.2 Primary Model Variables

The model is intended to represent a sector of industry. At an abstract level, it must therefore represent multiple companies within the sector and sources of financing for them.

Table 5 shows the primary variables that will be used to construct a model. They are "primary" because they contain data that drives the model.

Table 5 · Primary Model Variables

Stratum	***Entity***	***Description/Purpose***
Sector	***Technological development***	Sector-wide progress in technological resource that individual companies may choose to employ. In the dot-com bubble, for instance, companies were racing to integrate the latest available technologies in their products. (Wheale & Amin, 2003).
Sector	***Demand for Product***	Represents the volume of consumers in the sector. The dot-com bubble was a time of rapid growth in people using the internet, which could be a driver for investment. Rapid expansion and the acquisition of users has been described as a component of the 'get big fast' strategy of early internet companies. (Oliva et al., 2003).
Sector	***Capital Available for Investment***	The amount that investors have available to invest in companies within the sector.
Company	***Venture Capital Investment***	The availability of venture capital finance, or desire of such companies to make investments, is an indicator of speculation in a sector. (W. A. Sahlman & Stevenson, 1985). Potential sources: Eurostat, Financial Databases.
Company	***Public Investment***	As above, but for investment in a publicly traded company, through a stock exchange. Potential sources: Public investment records.
Company	***Valuation***	Calculated by number of shares multiplied by share price at the given time (i.e. capitalisation, but calculated wherever possible for non-publically-traded companies). Included to test whether valuation affects investor decisions. Potential sources: Public investment records.
Company	***Product Development***	Represents a company's technological capabilities, and ability to deliver products. May be sympathetic to overall sector technological development, although some companies will have a relative advantage/disadvantage. Product development may lead to tie-in for users (Berners-Lee, 2010), attract more users to the product and potentially hype the company.
Company	***Product Adoption***	Number of people using the product(s) of a company.

4.2.3 Derived Model Variables

Some items that would be desirable as variables of the model are inappropriate to build into the model structure as primary variables. This may be due to the intangible nature of the item, or because it is a function of some of one or more other variables. Thus, a derived variable is one that is wholly dependent upon others, and can serve only as an indicator or output of the model.

Table 6 shows the proposed derived variables.

Table 6 · Derived Model Variables

Stratum	*Derived Entity*	*Description/Purpose*	*Function of*
Company	***Investment Volume***	Investment volume is measured with different variables for private and public companies. This variable will be a compound of both, to simplify the model architecture.	• Investment (Venture Capital) • Investment (Public)
Company	***Market Share***	Proportion of consumers using a company's product.	• [Company] Product Adoption • [Sector] Demand for Product
Company	***Perceived Risk***	The risk of investing in a company, as perceived by investors.	• [Company] Market Share • [Company] Valuation
Sector	***Market Share Fragmentation***	Indicator of whether the market comprises several competitors of similar performance, or whether a dominant company has emerged.	• [All Companies] Market Share
Sector	***Perceived Sector Experience***	The degree to which investors think they understand the behaviour and business models of the sector they are working in.	• [Sector] Technological Development • [Company] Product Development • [All Companies] Valuation
Sector	***Speculation***	Momentum of investors involved in a sector they do not fully understand, or otherwise neglect/ignore due diligence.	• [Sector] Demand for Product • [Sector] Market Share Fragmentation • [All Companies] Perceived Risk • [Sector] Perceived Sector Experience

4.2.4 Model Design

The following model design (Figure 10) is based upon the specification and variables noted so far in this chapter.

The model is split into three sections, representing the overall Market or Industry, Investors and Companies.

Variables have been linked in the diagram using the following notation method:

- A line with a + indicates a direct (positive) relationship between variables. The destination variable will increase or decrease with the origin variable.
- A line with a – indicates a negative (inverse) relationship between variables. The destination variable is inversely affected by the origin variable.
- A line with “Fn.” Indicates a more complex relationship between variables. The effect of the relationship may depend upon rates of change at the origin variable, for instance.

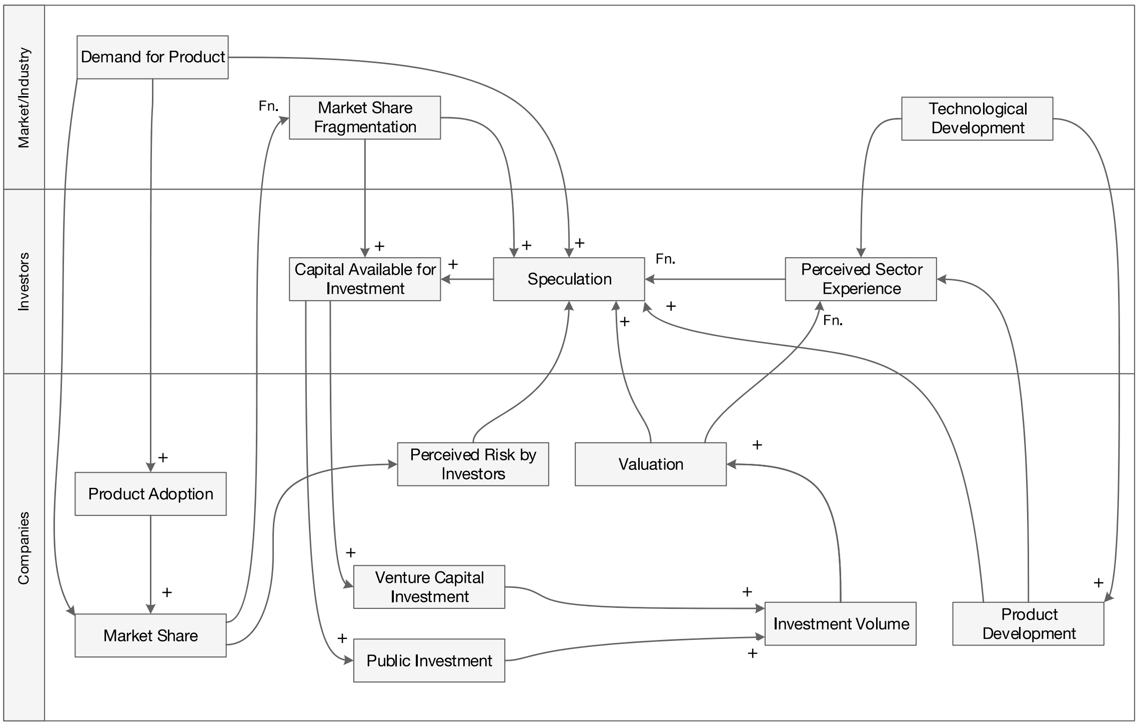

Figure 10 · Indicative Model Design

Figure 11 shows the same model, with annotations to show System Dynamics loops. Interestingly, only one loop appears; a positive loop around the noted Speculation/Investment/Valuation components. The causal links involved in the loop are emphasised with dashed lines.

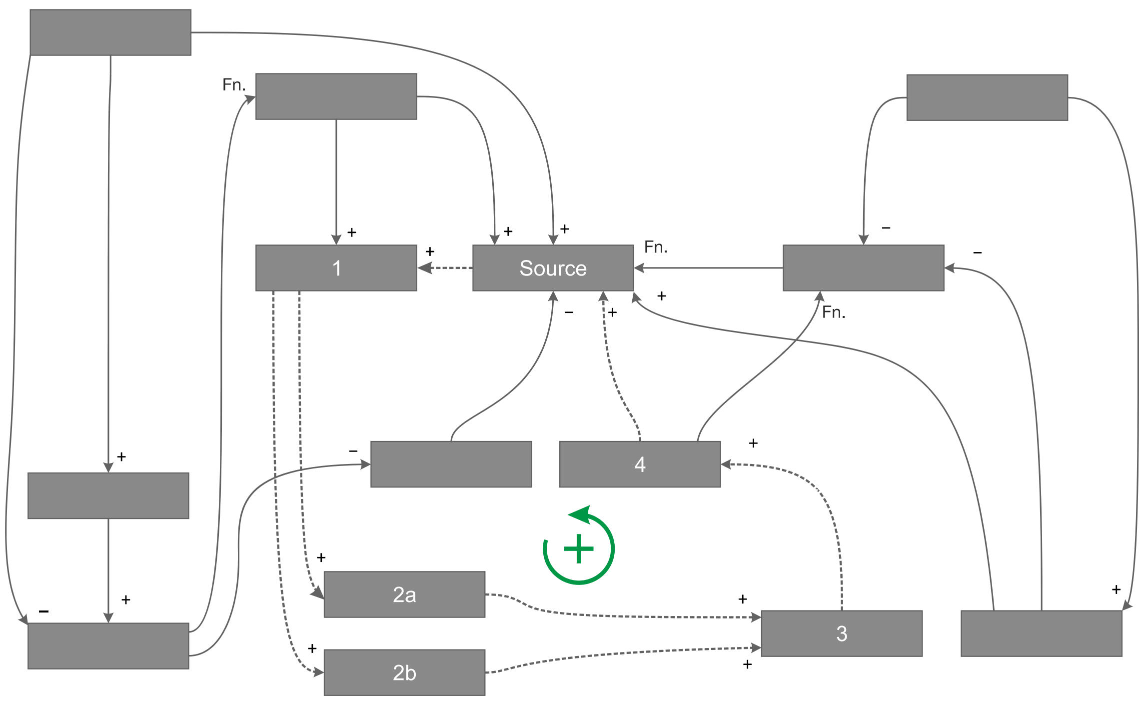

Figure 11 · System Dynamics Loop in the Indicative Model Design

4.3 Experimental Design

At this point, it is helpful to remind ourselves of the overarching research question:

> *This research aims to identify, quantitatively, the degree to which companies involved with the web sector have exhibited a repeat of the bubble-like state that was observed during 1999-2001.*
>
> *Excerpt from Section 1.1.*

The model shown in Figure 10 was a tool to aid development of this research. It presents an interpretation and combination of findings from the various literature that has been consulted for this work. Some elements are based on actual data, while others are conceptual and not quantifiable. This model is, intentionally, not computable and cannot answer the research question *directly*.

The literature review has revealed key metrics that may aid in the identification of a bubble state, and these are illustrated in the model. This research hypothesises that if the model is constructed from findings in literature, then relationships within the model should be reflected in a corresponding analysis of representative real-world data. Some relationships that will be reviewed are:

1. Technological Development positively influences Investment Volume.
2. Investment positively influences Valuations.
3. Market Share of a company positively influences Investment in the company.

The method and datasets required to test these relationships are discussed in the following section.

4.3.1 Conceptual Design

This section aims to establish a potential approach to answer the research question. It works at a high level, and an actual technique to implement the idea is described in the next subsection.

The research question asks whether a repeat of the 1999-2001 bubble-like state has ever occurred. Thus, a benchmark of "the 1999-2001 bubble-like state" is required. Subsequent periods may then be compared against this benchmark, to report (quantitatively) how similar a period is to the benchmark.

Broadly speaking, the process was as follows:

1. Collect data.
   Actual metrics to be derived from model.

2. Run analysis to establish a quantified benchmark for the dot-com bubble period.

3. Run the same analysis on a subsequent period.

    a. Compare results for this period to the benchmark created in step 2. Any similarities/differences in the results may be indicative of the presence/absence of a bubble state.

4. Return to step 3 for additional time periods *(optional)*.

This approach offered the potential to explore results for more than one period after the bubble. By analysing two or more periods after the bubble, a convergence/divergence from the benchmark may be observed.

The metrics employed to do this, time periods analysed and interpretation of results are subject to merits and limitations of the selected analytical method.

The number of companies surviving a bubble might be one such metric that could be produced from the data. When compared to the number of companies that entered the bubble, the resulting ratio may give an indication of the severity of the bubble.

While this approach is feasible, it has some drawbacks. A bubble has to have grown and burst within a known period, so that the number of companies before and after may be compared. The nature of the research question, in concurrence with mainstream literature, assumes that we do not yet know whether a bubble has occurred in the Web 2.0 era. As such, this metric cannot be implemented prior to the emergence and confirmation of a bubble.

Furthermore, producing this metric requires the model to take a high-level evaluation of the data. The model herein has a finer-grained scope that is unsuitable, working per-year, and per-company.

4.3.2 Real Design

The previous section proposed a conceptual design for an experiment that would answer the research question. This section aims to identify a suitable technique to convert the conceptual design into a feasible experiment.

The conceptual design and model make some implications of the data that will be analysed:

1. One axis of data (probably rows) will represent instances in time.

2. One axis of data (probably columns) will represent metrics of the period. These will reflect the metrics shown in the model.

3. The relationship between metrics in each analysed period will be the primary output.

Factor Analysis lent itself particularly well to these requirements. Given metrics, a Factor Analysis will try to form “components” around these metrics. Components represent groupings of metrics with common correlative properties within the analysed period. Metrics are given a value of “loading” to each component. As a method of dimension reduction, this analysis will reveal underlying commonalities between the metrics.

The Factor Analysis may be run in a binned fashion. That is, an initial analysis may be run for the dot-com bubble period, and subsequent analyses may be run for other periods.

The results from different periods will be comparable. Changes in the structuring of components within each period will enable detailed analysis of the results.

This approach implies the following schedule of tasks:

Gather data for analysis
Based upon the model, and the primary/derived variables presented in Table 5 and Table 6, data for the following variables will be sought.

Table 7 · Input variables for Factor Analysis

Metric	***Source***
Venture Capital Investment *(by phase if available)*	Bloomberg *or* Eurostat *or* Datastream *or* NVCA
Online Population *(by geographic area if available)*	Datastream *or* World Bank
Company Share Price	Datastream
Company Volume Traded	Datastream
Company Sales	Datastream
Company R&D Spend	Datastream
Company Market Capitalisation (i.e. valuation)	Datastream
Company Market Share	Datastream

The minimum reporting period sought for each variable in Table 7 shall be one year. If more detailed data is obtained, it was retained and aggregated into annual representations. The actual analysis periods will be the lowest common denominator of available reporting periods across all required data.

Wherever possible, venture capital and company data will be limited to activity within the technology/web sector.

Bin Data by Period
As explained earlier, a benchmark is required of the dot-com bubble period. This may then be tested against subsequent analyses for other periods.

Four bins are defined as thematic ranges of years, so that each bin represents a different period within 1995 – 2012.

Figure 12 shows a graph of the NASDAQ index for the relevant period, and illustrates the proposed binning periods. Gaps of one year are left between each bin to ensure greater contrast between them.

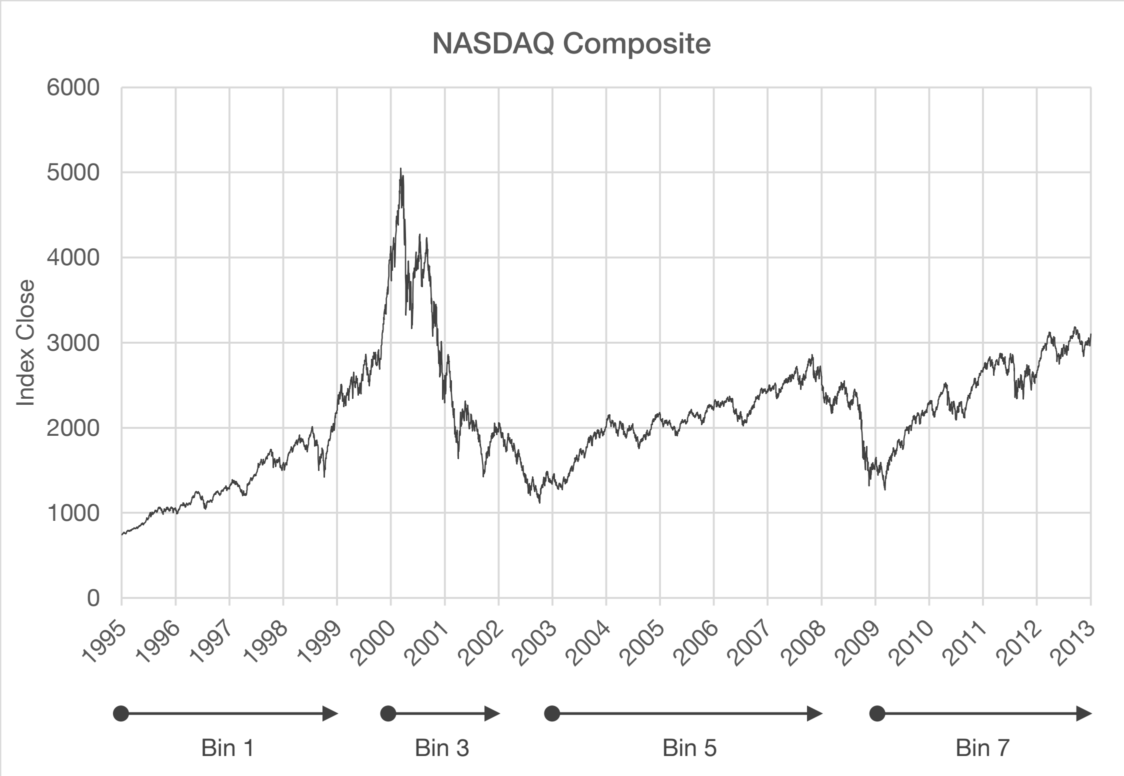

Figure 12 · NASDAQ Composite index annotated with Data Bins

Detailed dates for the same bins are presented in Table 8.

Table 8 · Binning Periods

Bin	*Beginning*	*End*	*Description*	*Included*
1	01.01.1995	31.12.1998	Nominal activity prior to bubble	Yes
2	01.01.1999	31.12.1999		
3	01.01.2000	31.12.2001	Bubble Collapse	Yes
4	01.01.2002	31.12.2002		
5	01.01.2003	31.12.2007	Recovery	Yes
6	01.01.2008	31.12.2008	Subprime market crash	
7	01.01.2009	31.12.2012	Contemporary	Yes

If data is available as early as 1995, Bin 1 will present results prior to the bubble collapse. Two post-bubble bins are designed, as 5 and 7.

Run Factor Analyses

A factor analysis will be run once for each included bin. A Varimax rotation will be applied to the result. This adjusts the axes of the initial results to optimise loadings and simplify the final factor structure. It assists in emphasising relationships in the data, and outputs “factors”.

Factors are similar in appearance to components, and retain the same characteristics of metrics and loadings. However, as they are generated through the rotation process, they are not directly comparable to the component values (hence, the different name).

Compare Factor Analysis Results
The factors extracted from each bin will be compared side-by-side. Results for the bin during the bubble will be compared to those after it. Particular attention will be paid to the structuring of the factors in each bin, as this indicates how the underlying relationships between the metrics changes over time.

If a bubble state has occurred since the dot-com bubble burst, then similar factors with similar weightings will be observed during and after the bubble.

4.3.3 Dataset Inclusion Criteria
The experimental design requires data from multiple sources. This necessitates the definition of benchmark criteria to ensure broad compatibility between the datasets.

1. **Earliest reporting period: 1995**
   All data sources should provide data from 1995 onward. This date is four/five years prior to the dot-com bubble bursting, and is shown on the annotated graph of NASDAQ value in Figure 12.

2. **Minimum reporting frequency: Annual**
   All data sources should provide data on at least an annual reporting frequency. Sources that provide data on a more frequent basis (e.g. quarterly or monthly) may be aggregated into annual figures.

3. **Data for Current and Defunct Companies**
   Several metrics described above require the sourcing of company-specific data. These should be sought for a variety of technology companies, and not just companies that are currently trading. The list of technology companies should include those that are current, defunct, old (i.e. established pre-1995), and new.

4.4 Summary
In this chapter, a conceptual model has been specified and designed. The model is designed and built using insights from the literature review. As a conceptual model, it is not intended to be computable. As a tool for furthering this research, it hypothesises ways in which data may be related. This informed the selection process for techniques of answering the research question.

Following the model design, an experimental design identified several metrics that should be analysed, and nominated Factor Analysis as the means of analysing them to answer the research question. At a high level, Factor Analysis will identify correlative relationships between key metrics before, during, and after the dot-com bubble. The comparison of outputs from these periods should provide an answer to the research question.

The following chapter presents a detailed account of how the metrics/data were gathered and prepared for analysis, up to the point of running the analysis.

Chapter 5. Methodology

Datasets from multiple sources were required to implement the experimental design detailed in Chapter 4. This necessitated the development of a method to clean, normalise, and combine the datasets, such that they could be easily imported into analytical software in any required format.

The data was acquired from multiple sources, sanitised, normalised and combined into a relational structure. The data was then loaded into whichever analytical tool necessary, by running an SQL query through an ODBC connection.

Performing this pre-flight on the data permitted an agile approach to analysing the data in various dimensions, by simply amending the SQL query used to access it.

Once loaded into analytical software (SPSS), the datasets were analysed according to the experimental design detailed in section 4.3.

5.1 Sourcing Venture Capital Activity Data

The conceptual model called for venture capital activity data on a **per-company** basis, so that relative levels of investment could be compared.

Bloomberg do not hold data on per-company venture capital deals, because many companies receiving this support are private and therefore not required to publish this information.

Following discussions with specialists at Bloomberg, it was found that this data *could* be ‘scraped’ from Bloomberg News articles, by checking for certain headlines and keywords, then extracting the venture capital data.

However, this approach has some major drawbacks that mitigate its practicality and value:

1. Not all venture capital deals may have been announced publically, and therefore are not reported by the Bloomberg News service.

2. Scraping the Bloomberg News service for articles related to venture capital would require writing software to perform this task, which is outside the scope of this work.

3. As an information service (i.e. *not* a data service), Bloomberg News could contain duplicated news stories, resulting in duplicated data.

4. Any data scraped from Bloomberg News would require considerable manual sanitisation and verification.

It is understandable that Bloomberg would be unwilling to publish data on this subject if there was a chance the data could be incomplete. Likewise, and for the reasons above, this data cannot be sourced in a sufficiently reliable manner.

Since per-company venture capital data cannot be obtained in a sufficiently reliable manner, the possibility of sourcing this data at a higher level was explored.

Three potential sources were found:

1. **UK Data.**
   NESTA's (National Endowment for Science, Technology and the Arts) 2010 report "Venture Capital Now and After the Dotcom Crash" (Pierrakis, 2010).

   NESTA has provided data on venture capital activity in the UK in their Venture Capital report. This report is compiled using data from the Thomson One service, but lacks the volume and depth required.

2. **European Data.**
   European Venture Capital Association (EVCA) quarterly reports.

   The EVCA publishes quarterly pan-European reports on venture capital activity, broken down by sector (EVCA, 2014). Reporting is suitably frequent (quarterly), and covers the European area only.

3. **US Data.**
   The PriceWaterhouseCoopers (PWC) and National Venture Capital Association (NVCA) MoneyTree report (PriceWaterhouseCoopers & National Venture Capital Assciation, 2014).

   The PWC MoneyTree report is a free service that uses data sourced from Thomson-Reuters DataStream to compile a quarterly report on venture capital activity in the US.

   This report is prepared on a quarterly basis by PWC, on behalf of the US NVCA. The data in this report is further grouped by US state, market sector and funding stage.

Much of the data required to validate the model is available in sufficient detail from the US, and not from sources in the UK.

An overview of the data sources is shown in Table 9, below, followed by a short discussion.

Table 9 · Comparison of Venture Capital Data Sources

	NESTA	EVCA	PWC MoneyTree
Subject Zone	United Kingdom	Europe	United States
Reporting Period	Annually	Quarterly	Quarterly
Oldest Data	Q1 2000	Q1 2005	Q1 1995
Most Recent Data	Q4 2009	Q3 2013	Q4 2013
Dimensions	Volume Num. Companies	Volume Num. Companies	Volume Num. Deals
Funding Stage Grouping	Yes	Yes	Yes
Sector Grouping		From Q1 2013	Yes
Geographic Area Grouping			Yes
Format	PDF	PDF	Excel
Data Definitions	Basic	Basic	Detailed

The data sources are clearly differentiated by resolution and target market.

Reporting Period

Yearly reports would be adequate, but higher resolutions provide the ability to improve detail if necessary or desirable.

PWC and EVCA offer the highest resolution data, with quarterly reports.

Historic Data

The subject of this research inherently sets the timeframe for which data should be acquired: ideally starting from a period preceding the dot-com bubble period (1999-2001), and ending as close to the current day as possible.

Only PWC offers data from before the dot-com bubble, which shows venture capital activity before, during, and after the bubble.

Dimensions

All sources provide venture capital data quantified by volume of capital invested. NESTA and EVCA provide a further dimension of the number of companies that received investment. Instead of this, PWC provides the number of venture capital deals. The number of deals (PWC's metric) is an appropriate measure, as it will account for companies that received multiple investments within the period, and the other sources will not.

Funding Stage Grouping

All sources group their data by funding stage. That is, venture capital investment volumes are broken down by the stage in the pipeline (seed, expansion, etc.).

Sector Grouping
Only PWC has consistently grouped their data by market sector, since their records began in Q1 1995. This will enable the study to ignore venture capital activity outside of the software/internet sector.

Sector grouping is not present in the NESTA analysis of venture capital activity, and the EVCA has only begun recording sector-based data since Q1 2013. By grouping data by sector, the PWC data clearly exposes funding trends and weightings towards certain sectors.

While PWCs data is US-centric, a cursory analysis of their per-sector funding data reveals strong weightings towards certain sectors. Thus, it would be inaccurate to rely on a dataset that does not break down its data into sectors.

Geographic Area Grouping
Both the NESTA and EVCA data applies to the *Subject Zone* as a whole. The PWC data is grouped by US state. However, the relationship of geographic area to venture capital activity, at a state-level of detail, is not the subject of this work.

Format
Only the PWC data is available in an immediately machine-readable format. Both the NESTA and EVCA data would require manual input, remedial work, and verification to convert into a machine-readable format.

Data Definitions
The PWC MoneyTree service clearly defines the groupings and categories that are used to generate and present their data. The website also defines the methodology that is used to compile the quarterly report, and declares types of data that are specifically excluded from the report. PWC provides sufficient detailed definitions that their published data could be reproduced from base data.

NESTA and EVCA do not provide such detailed information on the sourcing and processing of their data, so one cannot be assured of the origins and reliability of these datasets.

5.1.1 Selection and Justification
PWC MoneyTree was used as the source of venture capital activity data for the purposes of this work. The discussion above demonstrates that it is the richest data available, the best defined, and requires the minimum effort to begin working with.

The grouping of venture capital data by sector is a key differentiator for the data sets. This affords the ability to focus upon the industry sectors relevant to this work, eliminating “noise” from other sectors.

The use of US data, through PWC MoneyTree, represents a divergence from the assumption that this work would be UK-centric. Such divergence may be advantageous, though, as many of the web-based companies that could be subject to analysis in this work are indeed based in the US.

The result is a more coherent analysis, where data subjects are based in the same geographic area and subject to the same economic conditions.

5.2 Sourcing Online Population Data

Online Population Data is available from several sources. It is published by the International Telecommunication Union (ITU), and re-published by Datastream (discussed for Company Data, in section 5.3) and the World Bank.

A summary of key differentiators for these sources is shown in Table 10.

Table 10 · Comparison of Online Population Data Sources

	ITU	*Datastream*	*World Bank*
Oldest data	2005	1991	1991
Most recent data	2014	2014	2014
Data Definitions	Basic	Basic	Detailed
Republished From		ITU	
Subject Zone	World		
Reporting Period	Annually		
Dimensions	Country/Region (by UN M49 Regions) % Population Online		
Format	Excel		

Online Population Data was sourced from the ITU (2014), this being the source closest to the original data. The ITU publishes annual statistics on various telecoms-related matters, including the number of fixed and wireless broadband subscriptions, and the proportion of individuals using the Internet (% Population Online).

The data is distributed in a simple Excel file, presented on a per-country basis, coded against the UN M49 standard country/region codes (United Nations Statistics, 2013). This provides codes and data for geographical regions, such as *"Developed Countries"* or *"The Americas"*.

Data was required from 1995 onwards. The ITU publishes data only from 2005. The World Bank, *which re-publishes ITU data*, provides the data for these earlier periods (World Bank, 2015). The data is available on the Microsoft Azure Data Marketplace, which is a marketplace for premium datasets (Microsoft, 2014). This loads the data into Excel, based on a query sent to the World Bank data service, and enables rapid transformation of the dataset.

Datastream also republishes ITU data, but with more basic definitions.

The ITU and World Bank datasets were combined to produce a coherent dataset for Online Population, categorised by UN M49 regions, for 1995 onwards.

The datasets provide only the percentage of the population that is online in any particular region. To convert this to a real number, population statistics for each region were downloaded from the World Bank (via Azure, as before) on a per-year basis. These were combined with the online population data and used to calculate the actual number of people using the internet.

The resolution and dimensions present in the data makes it suitable for this analysis, with minimal adjustments. As venture capital data for the US was sourced in the previous section, Online Population data specifically for the US may be included in the analysis. Data for other regions, such as *OECD Developed Nations* and *The World* may be included in the analysis, to explore relationships.

To allow analysis to be as flexible as possible, *all* the Online Population data will be saved, for all regions and years. This will allow rapid selection of appropriate data from the larger set, should the analysis require it.

5.3 Sourcing Company Data

Company and Security data was sourced from Thomson Reuters Datastream, an economic research and asset analysis tool used by investment managers and analysts. Datastream is a database that is populated by Thomson Reuters by combining quantitative data from annual/quarterly company reports, and other industry sources. It is a single entity, where data from otherwise-disparate sources can be queried and compared.

Datastream structures financial data around “levels”, representing various entity types, in what is called the “Worldscope”. For instance, Worldscope levels exist for *Company* and *Security*. This separates data about a given Company from that of (possibly multiple) underlying traded Securities.

Datastream also features “lists” of Companies and Securities, curated around various themes. This allows users to retrieve data for a group of related Companies/Securities quickly. For the purposes of this work, the “US Software Companies” list was used, which enumerates 105 Companies (listed in Appendix A). This list meets the criteria set out in section 4.3.3; namely, that the list includes companies that are “current, defunct, old (i.e. established pre-1995) and new”.

5.3.1 Datastream Querying Technique

To access the Datastream product, a graphical mnemonic-based querying tool is provided, manifested as an Excel plugin. The plugin delivers results directly into a spreadsheet.

Datastream supports various query types. *Static* queries retrieve data that may not be compared temporally, such as a list of directors of specified Companies. This research used *Time-Series* queries, that retrieve quantitative data for given periods.

A typical Datastream Time-Series query comprises the parameters found in Table 11:

Table 11 · Datastream query parameters for Time-Series queries

Parameter	*Description*
Series (x axis)	One or more subject entities for which the query should return data. These are mnemonic-based representations of Companies, Securities and other Worldscope levels.
Start/End Dates (y axis)	The period for which the query should return results. Specified as explicit dates, or relative dates (e.g. "Start of Year").
Datatypes (z axis)	Dimensions of the Series that the query should return, selected from a searchable list. These are mnemonic-based representations of 16,432 financial indicators. Each Datatype has a Reporting Period, which specifies how frequently the Datatype is updated. Highly dynamic Datatypes, such as Share Price (mnemonic P) are updated at least hourly, whereas the Net Sales (mnemonic WC01001) are sourced from annual Company Reports, and thus have an annual Reporting Period.
Frequency	The time frequency upon which results should be returned. (e.g. Yearly, Quarterly, Monthly, Weekly, etc.).

Some Datatypes (as described in Table 9) are applicable at certain Worldscope levels, whilst others are not. Queries must be constructed with care to ensure that relevant Datatypes are selected, given the level of the Series requested. Effectively, one must ensure that the queried attributes are actually possessed by the desired entities.

Frequency should be specified with attention to the Reporting Periods of requested Datatypes. If an Annual Datatype is requested (e.g. *WC01001*) and a Quarterly Frequency specified, the single annual value will be duplicated for each Quarter in the results.

Results are returned as a crude crosstab within an Excel spreadsheet, with Time-Series axes, as the query name would suggest. By default, rows represent periods, formatted according to the query parameters. Columns represent first the requested Series, and below those, the Datatypes. In effect, this concatenates all the requested Datatypes into one table.

Table 12 shows an example Datastream query to find the Research & Development Spend datatype (mnemonic WC01201) and Trading Volume datatype (mnemonic WC08006) of Google and Yahoo for the three weeks between 31/12/2004 and 14/01/2005.

Table 12 · Example Datastream query

Parameter	*Input*
Series	@GOOG, @YHOO
Datatypes	WC01201, WC08006
Start Date	31/12/2004
End Date	14/01/2005
Frequency	Weekly

Table 13 shows the output from this query, and exemplifies why care must be taken when designing queries; some of the retrieved data is duplicated because *Annual* Datatypes were requested against a *Weekly* frequency. The query should be re-formulated, using either an Annual Frequency, or different Datatypes.

Table 13 · Datastream crosstab output

Name	*GOOGLE INC. - RESEARCH & DEVELOPMENT*	*GOOGLE INC. - TRADING VOLUME*	*YAHOO! INC - RESEARCH & DEVELOPMENT*	*YAHOO! INC - TRADING VOLUME*
Code	@GOOG(WC01201)	@GOOG(WC08006)	@YHOO(WC01201)	@YHOO(WC08006)
31/12/2004	214289	343760368	368760	153826027
07/01/2005	483978	831408175	547137	178197917
14/01/2005	483978	831408175	547137	178197917

Querying in this way reveals two shortcomings that must be overcome:

1. **Datastream limits the size of returned resultsets.**
   If a query contains too many dimensions, the returned data may be truncated. The truncation manifests as a query error code in the affected cells.

2. **Datastream does not return data in a true crosstab.**
   This is evidenced by Company name duplication in the column headings of Table 13. This prevents quick and easy access/analysis of the data in Excel, once Datastream has returned its results. It necessitates the creation of a workflow to convert Datastream output into a truly machine-readable format.

Shortcoming #1 was mitigated by anticipating and controlling the expected size of resultsets. In all cases, this was achieved by breaking down larger queries into several smaller ones, thus limiting the number of Series and Datatypes requested within each query. The resulting datasets, which were paginated, were combined in Excel to reproduce the overall results.

Shortcoming #2 was partially mitigated by the resolution to Shortcoming #1; by querying for just one Dataset at a time, the crosstab may be reduced to a two-dimension matrix. This removes the potential for ambiguity in some of the crosstab output headings.

5.3.2 Querying Datastream

Table 14 shows the Datatypes that were extracted from Datastream for the purposes of this research. As per the solutions to the Datastream shortcomings, described above, each Datatype was fetched in an individual query.

Table 14 · Datatypes extracted from Datastream

Datatype Mnemonic	***Name***	***Reporting Period***	***Worldscope Level***
NOSH	Number Of Shares	Hourly	N/A
WC01001	Net Sales Or Revenues	Annual	Company
UP	Unadjusted Price	Hourly	N/A
MV	Market Value (Capital)	Hourly	N/A
P	Price (Adjusted - Default)	Hourly	N/A
WC01201	Research & Development	Annual	Company
W08006	Trading Volume	Annual	Security

Table 15 shows a sample of the output for the Trading Volume (mnemonic W08006) query. The four Companies shown in the table were selected to exemplify the various types of output that occur. Company mnemonics are shown in bold text, in the header row.

Table 15 · Sample of output from Datastream query.

	@AKAM*(W08006)*	***U:DDD****(W08006)*	***@ADBE****(W08006)*	***@ALTR****(W08006)*
1990			4895877	363212
1991		13455	6620124	1054756
1992		12331	7975639	1856235

This output shows data that reflects three different company circumstances:

1. **@AKAM.**
   This company was not publically traded during the requested period, and so has no data.

2. **U:DDD.**
   This company became publically traded during the requested period, so data appears from 1991.

3. **@ADBE and @ALTR.**
   These companies were publically traded throughout the entire requested period, so a complete dataset is returned.

Additionally, a company may cease trading publically during the requested period, causing the data to be unavailable following the date the company was made private.

5.4 Structuring and Combining Data

The data sourcing exercise generated nine distinct datasets. These had to be combined into one dataset to enable analysis. The work described in this section was carried out entirely in Excel.

Seven of nine sets were generated by similar Datastream queries, where only the Datatype was changed. This enabled the sets to be combined by identifying matching company mnemonics and years in each set. By combining each dataset individually into a destination dataset, a combined dataset of all Datatypes was produced.

To combine the sets, Datastream's seven crosstab outputs were converted into Pivot Tables. The following workflow in Excel was found to be most effective, and was applied to all seven datasets:

1. **Isolate the Company mnemonic.**
   Using Excel functions, extract the Company mnemonic (as shown in red in Table 15) from cells in the heading row:
   `=LEFT(A2, FIND("(", A2)-1)`

2. **Produce a multiple consolidation range Pivot Table of the data.**
   This reproduces exactly the same table structure, but in a "live" Pivot Table.

The last dataset to be processed (number 7) was additionally flattened into a linear table, by following Step 3:

3. **Flatten the Pivot Table to a linear table.**
   Double-click the Pivot Table's *Grand Total* cell.
   This creates a new sheet, containing a flattened, linear version of the data in the Pivot Table.
   Table 16 shows sample output from this step, using data in Table 15 as an input.

Table 16 · Flattened/Linear Datastream output

Row	*Column*	*Value*
1990	@ADBE	4895877
1990	@AKAM	
1990	@ALTR	363212
1990	U:DDD	
1991	@ADBE	6620124
1991	@AKAM	
1991	@ALTR	1054756
1991	U:DDD	13455
1992	@ADBE	7975639
1992	@AKAM	
1992	@ALTR	1856235
1992	U:DDD	12331

This exercise produced a flat listing of each company-year (see *Row* and *Column* columns in Table 16), plus the *Value* of Datatype W08006, from the last Dataset. The Pivot Table data of datasets 1-6 was then imported into this flat listing, by appending columns containing GETPIVOTDATA functions similar to the following. Lookups on the 'Date' and 'Company' dimensions are shown in red.

```
=GETPIVOTDATA("Price", 'Share Prices'!$A$3, "date", $A4, "company", $C4)
```

This resulted in one flat table, containing rows that represent company-years, and one column per dataset. A truncated sample of four datasets is shown in Table 17.

Table 17 · Flattened and Combined Datastream output

Year	*Company*	*MV*	*NOSH*	*UP*	*P*
1990	@AAPL	4430.25	125681	35.250	8.8125
1990	@ADBE	417.23	20604	20.250	1.2496
1990	@ADI	458.25	47000	9.750	1.6252

The resultant two-dimensional table contains all the output from Datastream. Being two-dimensional, it is suitable for storage in a single table, in an SQL database.

Venture capital and online population data was provided in a more usable format. The three datasets all feature annual data, so a column representing *year* was extracted from the remaining datasets.

The online population dataset contained data on a per-region per-year basis. Some regions were actual countries, and others represented groups of countries (e.g. *"European Union", "High Income", and "World"*). A list of unique regions was extracted from the dataset, and coded according to whether each row represented an individual country or a region.

The Venture Capital dataset contained data on a per-phase per-quarter basis, where phase represents phases of the venture capital process (i.e. *"Startup/Seed", "Early Stage", "Later Stage", and "Expansion"*). A list of the four phases was extracted from the dataset. The data was provided on a per-quarter basis, so a *year* column was added and populated from existing data.

A conceptual diagram is shown in Figure 13, describing how the common *year* fields are used to join the three datasets. Only a sample of columns/attributes are shown in the diagram.

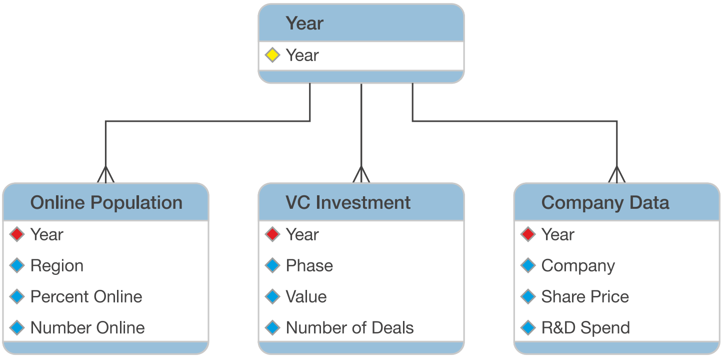

Figure 13 · Conceptual Database Design
Showing Dataset sample data columns and Joins on Year columns.

5.5 SQL Database

A database schema was devised to concert all the datasets into one set, for easy querying and retrieval. The schema was created using MySQL Workbench, and forward-engineered to a MySQL server.

The rationale for loading data into SQL was:

- **Faster/easier adjustments to data presentation and layout.**
  Data may be presented in different ways rapidly by adjusting SQL queries and/or writing new ones. Achieving the same in Excel often requires undesirable duplication of the source data, and multiple complex functions to re-format data.
  This argument is essentially one of using code/commands to render data, versus manual manipulation in a GUI application.

- **Faster/easier transformations and adjustments to data.**
  SQL queries may be used to preview potential adjustments to table data, before committing them to storage. Such changes may not need to be committed at all, and may exist as calculation steps in an SQL query.

- **Easier to import data into various analysis applications.**
  Most data analysis applications include ODBC support, which may be used to load data directly from the SQL database, as either direct table outputs or the results of a more complex query on the underlying data.

The design, shown in Figure 14, complements the flat Excel table produced in section 5.4, by matching the column types and names. This can be seen as the ‘annualReport’ table in the schema. As a result, the Excel tables may be imported directly into the database.

The groupings of tables shown in the diagram represent storage for the three main datasets that have been described.

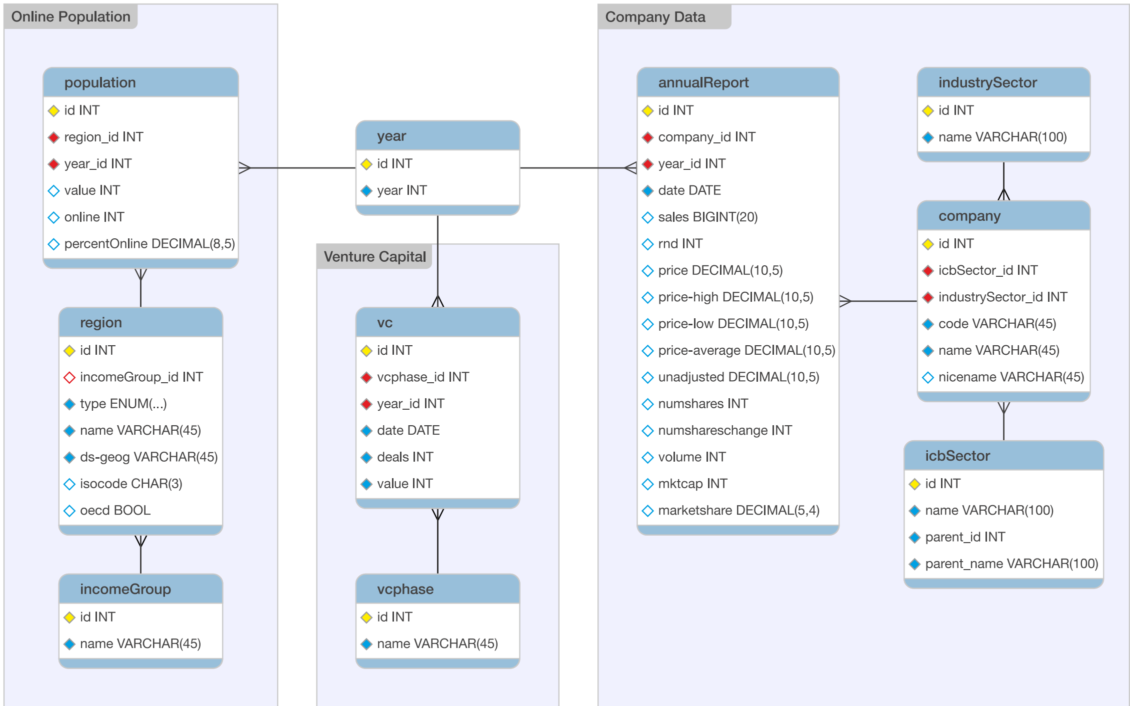

Figure 14 · Database Design
Showing groups of tables for the three datasets

1990-2014 was the date range for data fetched from all data sources. The ‘year’ table was populated manually, by creating a record for 2014 and re-running the following SQL query until 1990 was reached. Each time it is run, the query finds the minimum year in the table, and adds a row of the year before it.

```
INSERT INTO `year` (id) VALUES ((SELECT MIN(y2.id) - 1 FROM `year` y2));
```

The ‘year’ table is used to join the three main datasets, so they may be compared and queried together. This is a technique commonly seen in medium- to high-level Business Intelligence systems, and adapted for the purposes of this work.

For convenience, the ‘id’ field of the year table is coded to the number of the year. This saves the repetitious work of coding years in the subsequent import.

5.5.1 Importing and Loading Data

To import the Datastream data, the Companies table was first populated with the companies targeted for this work, storing their real name and Datastream mnemonic. Each company was assigned a unique ID in the table. This was achieved by exporting the data from Excel to CSV, and importing it into MySQL using the Sequel Pro DBA utility.

To import the remaining Datastream company data, the *'company'* table was imported into Excel using ODBC. The imported dataset related the new database IDs to the existing Datastream mnemonics, and other company metadata. The remaining Datastream data in Excel was then coded with the company IDs, using Excel's VLOOKUP function. This enabled the dataset to be imported into the *'annualReport'* table, and related back to the subject company.

Online population data was imported into the *'region'* and *'population'* tables. The *'region'* table was populated first, so that each unique region from the online population dataset was assigned an ID in the database. The same ODBC re-importing technique as above was used to annotate the new region IDs onto the actual population values, before importing them into the database as well.

Venture Capital data was imported into the *'vc'* and *'vcphase'* tables. The *'vcphase'* table was populated with the four venture capital phases noted in the PWC dataset. These were re-imported into Excel via ODBC, with their newly generated IDs. The actual venture capital data was then imported, using the corresponding phase IDs.

5.5.2 SQL Views for Analysis

Data can be loaded into analytical tools (such as SPSS) using ODBC. However, this necessitates a rather large query being written in various client applications, repetitiously. To simplify this process, an SQL view was defined on the database server. This view uses a query to join the three datasets, and present their data in a single table. The single table can then be imported into analytical tools in one operation.

A *Market Share* column was added to the *'annualReport'* table. For each company-year row, this stored the percentage market share by revenue. This column was calculated by summing the revenue amongst all companies in a given year, then calculating individual companies' shares of it.

The full query used to generate the SQL View is shown in Appendix B.

5.6 Cleaning Data and Checking for Normality of Distribution

Before performing any statistical analysis, it is prudent to check that input data is correct and consistent, and that no data has been compromised during import or transformation. The process of importing data into the database was verified by:

- Taking samples of data in the database and comparing against the raw output from DataStream.
- Comparing columns to one another, to ensure that data had been imported into the correct locations, and was not erroneously duplicated in other columns.

These verification tasks revealed a pair of datasets with several matching values. This was traced to a human error while querying Datastream, resulting in one page of results in a dataset being for the wrong data series. The erroneous data was identified and replaced with the correct data from Datastream. Both the original and partially cloned datasets were then closely verified.

As a further point of verification, the data should also be checked for normality of distribution. This is usually accomplished by producing a histogram of the data, and comparing it to a normal distribution curve.

Figure 15 shows how raw data from two of the Datastream datasets appear when drawn as a histogram. The distributions are heavily skewed to the lower end of the scale, with long tails, and not normally distributed. These results are representative of all the Datastream datasets. Any statistical analysis performed upon data in such a state may yield inaccurate results and be deemed unreliable. This is due to the wide ranges in the data, and differences of magnitude between the largest and smallest values.

To address this issue, a logarithm was calculated for each point in the Datastream datasets. Results were better than expected, and are easily comparable to a normal distribution.

Figure 16 shows histograms for the same data, after a logarithm is calculated for each datum. Following this transformation, the histograms show the datasets as being normally distributed. Again, the results shown in the histograms are representative of all the Datastream datasets.

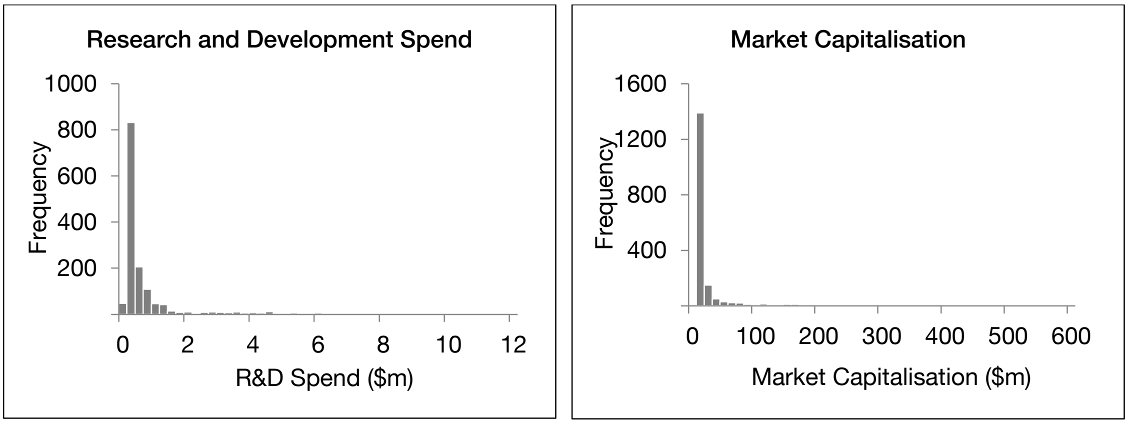

Figure 15 · Histograms of raw values for two sample datasets

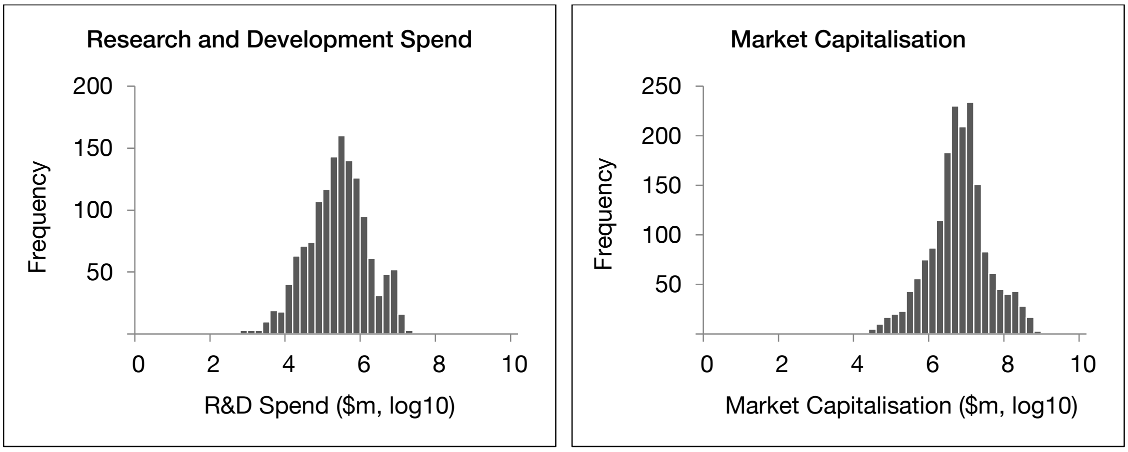

Figure 16 · Histograms of logarithm values for two sample datasets

This transformation was applied to all the Datastream datasets. Similar results were produced in all seven cases.

The Venture Capital dataset also contained wide-ranging values of similar magnitudes. The same process was applied to these values, with the same result. In all cases, the original data was left untouched, and logarithms were calculated within the views.

5.7 Analytical Techniques and Application to the Research Question

At this point, it is helpful to remind ourselves of the overarching research question:

> *This research aims to identify, quantitatively, the degree to which companies involved with the web sector have exhibited a repeat of the bubble-like state that was observed during 1999-2001.*
>
> *Excerpt from Section 1.1.*

Thus far, company and sector data has been sourced from 1995, through the dot-com bubble, until the present day. That data has been concerted such that it may be viewed and analysed as one coherent set. This provides a foundation of data covering all the necessary periods.

The research question asks whether the circumstances observed during 1999-2001 have been observed in any subsequent periods. To that end, *years* in the dataset were coded according to thematic periods. This is a form of binning, and allowed the data to be analysed based on periods. Comparing analyses of pre- and post-bubble bins will show similarities or differences between them, and the question may be answered.

Various statistical analysis techniques were checked for suitability, by evaluating them against the research question. Some techniques were applied to the dataset, to test their suitability further. These included discriminant analysis, k-means cluster analysis, and regression analysis. Factor Analysis emerged as the technique best suited to answer the research question, and to handle the volume of data effectively.

Factor Analysis was performed upon the concerted datasets, to identify variables that relate to one another. This outputs a number of derived factors, showing which columns contribute to each factor (negatively or positively), and by how much. This is a form of dimension reduction, which may be used for exploring relationships between columns of data, or simplifying wide datasets into a smaller number of columns.

The objective of this technique was to test the dataset to identify columns of that related to one another. Where columns did relate, data indicating the nature of relationship is desired.

The common primary indicator of the dot-com bubble is the excessive valuation of the companies involved, and this is frequently shown in graphs that illustrate the situation. The Factor Analysis showed the level to which this metric related to other metrics in the dataset. It showed which metrics contributed positively alongside valuations, and which contribute negatively (i.e. in an opposed manner). Relationships between other metrics were observed, and these were explored as potential contributors to the research.

5.8 Summary

This chapter has described, in detail, the technique employed to gather and analyse data to answer the research question.

Datasets were identified and retrieved from three major sources. Those sets comprised detailed online population statistics, quarterly national US venture capital reports, and annual company reports.

After sanitisation and verification, the datasets were combined in a single MySQL database, adjusted to a normal period of one year, and joined into a single dataset using *year* as the key. This was achieved using a SQL View, which enables non-destructive transformation of the input data, and easy rapid importing to analysis software via ODBC.

The data was loaded into analysis software, and tested for normality of distribution prior to actual analysis. Several skewed columns of data were found, and corrected by calculating logarithms of the data points. This remedial action yielded better-than-expected results, producing columns that fitted a normal distribution particularly well.

Thematic Binning and Factor Analysis were earlier identified as techniques for answering the research question. Bins were applied to code the years according to whether they are pre- or post- dot-com bubble. Analysis results based on these periods will support an answer to the research question.

The following chapter presents the results that occurred.

Chapter 6. Results

To explore the factors within the data, rows were binned according to the year they represented. Bins were defined as thematic ranges of years, so that each bin represented a different period within 1995 – 2012, and are detailed in the experimental design.

Four Factor Analyses were then run; one for each included binning period. The same parameters were used as in the previous Factor Analysis. A sample of the dataset, and a summary of the number of cases per bin period, may be seen in Appendix C. Correlation matrixes for these analyses may be found in Appendix D.

Table 18 shows the eigenvalues for each extracted component in the binned analyses. The eigenvalues show how much variance in the data is accounted for by the corresponding component.

Table 18 · Eigenvalues extracted per bin, per component

Bin	**1** **1995–1998**			**3** **2000–2001**		**5** **2003–2007**			**7** **2009–2012**			
Component	1	2	3	1	2	1	2	3	1	2	3	4
Eigenvalue	10.54	6.28	1.03	11.08	6.24	8.00	5.84	2.45	8.72	5.97	2.07	1.25

Figure 17 shows a composite scree plot of the eigenvalues in Table 18. The points are sorted in descending order, and each point represents a potential component in the Factor Analysis output. Components are deemed relevant when they produce an eigenvalue ≥ 1.00.

The sum of eigenvalues in an analysis is equal to the number of variables provided as input. Therefore, a component's eigenvalue represents the proportion of variance within the whole set that is accounted for by the component.

Nineteen components emerged from each factor analysis, but not all were relevant. Figure 17 presents this as a graph, omitting eigenvalues for components 11-19 as they were zero or close to zero. This improves the clarity of the relevant data (i.e. eigenvalues ≥ 1.00).

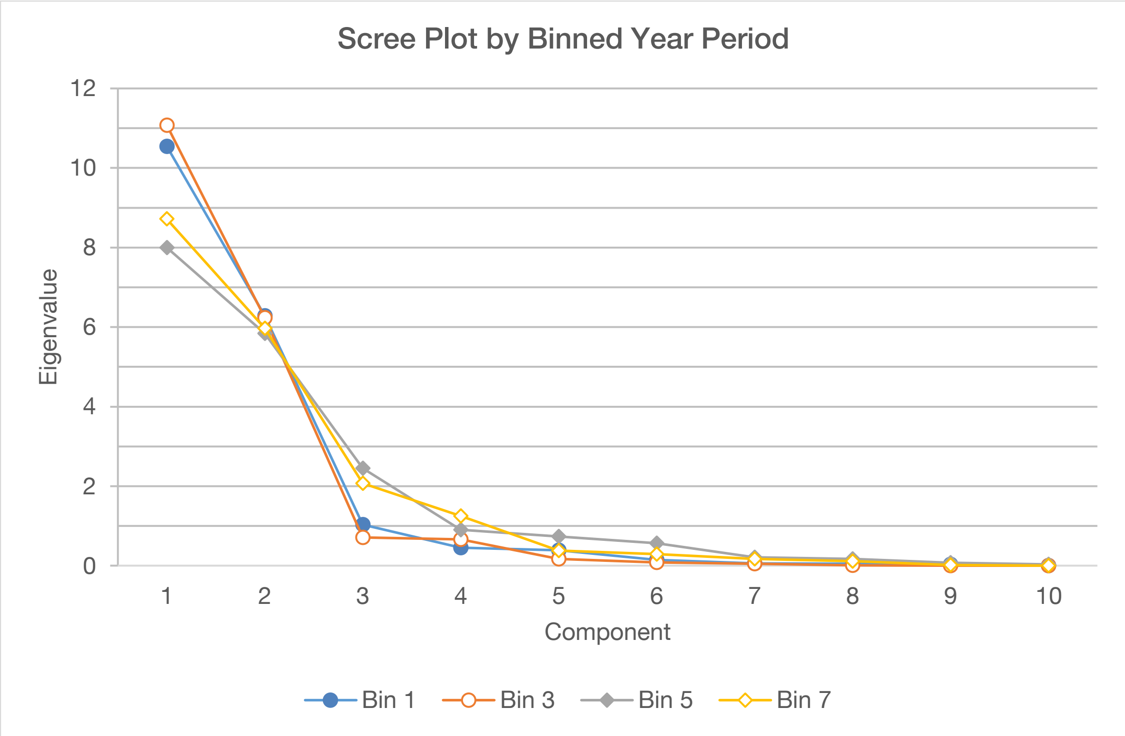

Figure 17 · Eigenvalues by Binned Year Period for a 19-component analysis
Components 11-19 had Eigenvalues at or near zero, and are omitted to improve the clarity of relevant data.

Bin 1 has two components that account for the majority of variance. One additional component (eigenvalue 1.03) is only just relevant.

Bin 3 has two components that account for the majority of variance, with similar proportions of variance explained as Bin 1. Bin 3 contains no other relevant components.

Bin 5 has three relevant components. The majority of variance is accounted for by two of them. In relation to the previous bins, this bin has produced a tighter grouping of relevant components.

Bin 7 has four relevant components. The majority of variance is accounted for by two of them. This bin exhibits similar grouping properties as Bin 5.

The same Varimax rotation was subsequently applied. Table 19 shows the resulting factors and factor loadings. To aid clarity and highlight differences between bins, the ordering of Factors I and II have been adjusted in Bins 1 and 3.

Table 19 · Binned Factor Loadings

Bin	1 1995-1998			3 2000-2001		5 2003-2007			7 2009-2012			
Factor	II	I	III	II	I	II	I	III	I	II	III	IV
sales	0.94	0.01	0.19	0.92	-0.02	0.93	0.08	0.02	0.94	0.05	0.06	0.11
price	0.46	0.07	0.83	0.67	0.18	0.51	0.19	0.01	0.15	0.05	0.07	0.98
rnd	0.94	0.02	0.08	0.92	-0.07	0.94	0.06	0.03	0.94	0.06	0.04	-0.02
volume_units	0.83	0.03	-0.51	0.81	-0.06	0.84	-0.04	-0.04	0.90	-0.03	-0.06	-0.17
volume_price	0.94	0.07	-0.09	0.90	0.12	0.91	0.10	-0.03	0.88	0.04	0.05	0.34
mktcap	0.95	-0.02	0.16	0.96	0.07	0.96	0.05	-0.01	0.90	0.01	0.03	0.36
marketshare	0.94	-0.07	0.19	0.92	-0.04	0.94	-0.03	0.01	0.94	-0.02	-0.01	0.10
numshares	0.95	0.05	0.10	0.93	-0.04	0.91	0.00	-0.01	0.93	-0.02	-0.01	-0.27
vc_phase1_value	0.01	0.89	-0.06	0.02	1.00	0.05	0.75	0.43	-0.02	-1.00	-0.05	-0.02
vc_phase2_value	0.02	0.99	0.00	0.02	1.00	0.01	0.14	0.96	0.02	0.55	0.82	0.02
vc_phase3_value	0.01	0.99	0.04	0.02	1.00	0.06	0.91	-0.24	0.02	0.67	0.70	0.01
vc_phase4_value	0.01	0.98	0.03	0.02	1.00	-0.06	-0.90	0.38	0.02	0.57	0.82	0.03
vc_phase1_deals	0.02	0.97	-0.02	0.02	1.00	0.05	0.79	0.50	0.01	-0.17	0.98	0.02
vc_phase2_deals	0.01	0.97	-0.02	0.02	1.00	-0.03	-0.24	0.91	0.02	0.78	0.61	0.02
vc_phase3_deals	0.02	0.99	0.03	0.02	1.00	0.07	0.99	-0.08	0.01	0.19	0.95	0.05
vc_phase4_deals	0.02	0.97	0.05	0.02	1.00	-0.05	-0.87	0.15	0.01	0.62	0.70	0.06
onlinepopulation_us	0.01	1.00	0.02	-0.02	-1.00	0.07	0.99	0.15	0.01	1.00	0.04	0.02
onlinepopulation_oecd	0.02	0.99	0.03	-0.02	-1.00	0.07	0.99	0.01	0.02	0.91	0.42	0.03
onlinepopulation_world	0.02	0.99	0.03	-0.02	-1.00	0.07	0.98	0.11	0.02	0.73	0.69	0.03

This analysis produced contrasting results for cases before, during, and after the dot-com bubble.

1995 – 1998 (Bin 1) has three factors; two contain uniformly strong factor loadings.

Factor I contains the venture capital (VC) and online population variable sets.

These all exhibit strong positive loadings (one 0.89 and the remainder ≥ 0.97).

Factor II contains the Company Data variable set.

Share Price exhibits the weakest loading (0.46).

The remaining relevant variables exhibit strong positive loadings (one 0.83 and the remainder ≥ 0.94).

Factor III contains the *Company Share Price* (0.83) and *Volume Traded (units)* variables (-0.51).

Volume Traded (units) exhibits the only negative factor loading in this factor.

2000 – 2001 (Bin 3) has two factors; both contain uniformly strong factor loadings.

Both factors contain the same relevant variables as their respective factors in 1995–1998.

Factor I contains the VC and Online Population variable sets.

This factor has restructured, as the Online Population variables loaded negatively in this period.

The two variable sets are now polarised:

All Venture Capital variables exhibit strong positive loadings (1.00).

All Online Population variables exhibit strong negative loadings (-1.00).

In 1995–1998, the relationship between these variable sets was mutually positive. It has become positive-negative in this period.

Factor II contains the Company Data variable set.

Share Price remains as the weakest relevant factor loading (0.67).

The remaining variables exhibit moderate to strong loadings (one 0.81 and the remainder ≥ 0.90).

Relative loadings remain similar to those of respective factors in 1995–1998; variables that were weakly positive in 1995–1998 are weakly positive in this period.

2003 – 2007 (Bin 5) has three factors.

The same two factors from 2000–2001 are represented here, plus an emergent third factor.

Factor I contains some venture capital variables, and all online population variables.

Two variables have inverted relative to their peers, and now load negatively.

VC Phase 2 Value and *Phase 2 Deals* have negligible loadings in this factor.

VC Phase 4 Value and *Phase 4 Deals* have strong negative loadings (≤ -0.87) in this factor.

The remaining relevant venture capital variables have strong positive loadings (≥ 0.75).

The Online Population variables have strong positive loadings (≥ 0.98).

Factor II contains the Company Data variable set.

All the variables load positively.

Share Price remains as the weakest relevant factor loading (0.51).

The remaining relevant variables exhibit strong positive loadings (0.84, and the remainder ≥ 0.91).

Factor III contains Venture Capital variables, for Phases 1 and 2 only.

This is a newly emergent factor.

All the factor loadings are positive.

VC Phase 1 Value and *Phase 1 Deals* have weak loadings (0.43, 0.50).

VC Phase 2 Value and *Phase 2 Deals* have strong loadings (0.91, 0.96).

This factor is connected to Factor I by two variables: *VC Phase 1 Value* and *Phase 1 Deals*.

2009 – 2012 (Bin 7) has four factors.

Factor I contains the Company Data variables.

All the variables load positively.

Share Price now has a negligible loading in this factor.

The remaining variables exhibit strong positive loadings (≥ 0.88).

Factor II contains some VC variables, and all online population variables.

One variables loads negatively. The remaining eight load positively.

VC Phase 1 Value has a strong negative loading (-1.00).

VC Phase 1 Deals and *Phase 3 Deals* now have negligible loadings in this factor.

The remaining VC variables exhibit moderate positive loadings (0.55 – 0.78).

The *World Online Population* variable exhibits moderate positive loading (0.73).

The *US* and *OECD-Countries Online Population* factors exhibit strong positive loadings (0.91, 1.00).

Factor III contains some VC variables and some Online Population variables.

This is a continuation of the emergent Factor III in 2003–2007.

All the variables load positively.

VC Phase 1 Deals and *Phase 3 Deals* have strong positive loadings. These variables were not relevant in Factor II, but have the strongest loadings in this factor.

The remaining relevant VC variables exhibit moderate positive loadings (0.61 – 0.82).

The relevant Online Population variables exhibit weak to moderate loadings (0.42, 0.69).

The *US Online Population* variable has a negligible factor loading.

Factor IV contains one factor: *Share Price*.

This is a newly emergent factor.

Share Price has a strong positive loading (0.98).

The following chapter examines these results in more depth, relating them to literature and the research question.

6.1 Summary

This chapter presented the results gained from the experiment, described in the experimental design.

The following chapter interprets these results, relates them to the literature, and provides insights to answer the research question.

Chapter 7. Discussion

The input data comprised several individual datasets, concerted together for the purposes of analysis. In many cases, the Factor Analysis identified that each set exhibited similar correlative properties, and generated factors from each set.

This chapter explores how the factors changed over the explored periods. It is divided into three major sections.

1. **Direct Interpretation.**
   Derives direct meaning and notable outcomes from the Results by comparing periods (bins), factors and variables, and analysing changes in relationships.
2. **Interpretation in Relation to Literature.**
   Identifies observations in the *Direct Interpretation* that are accounted for or described by conventional literature.
3. **Interpretation in Relation to Research Question.**
   Utilises the Results and Interpretations to answer the Research Question.

7.1 Interpretation of Binned Factor Analysis Results

The factors extracted from the Factor Analysis vary as the analysis is run for differing periods. The following discussion refers to the results exhibited in Table 19.

Several charts of factor loadings are presented in this section. The charts all present factor loadings for venture capital and online population variables, and are direct renderings of the Factor Analysis results shown in Table 19. To aid comparison and highlight differences between periods, some particular factor loadings are highlighted.

○ Online Population variables
△ Venture Capital Phase 4 variables
× Remaining Venture Capital variables

Throughout this discussion, reference is made to the four *VC Phases*, shown in Table 20 and in concurrence with the definitions of the VC data (PWC MoneyTree, n.d.). These refer to the stages of VC funding that may be provided to a company. A company typically receives initial funding at stage one and, if successful, continues to receive further funding and support in the subsequent phases.

Table 20 · VC Phases

Phase	Name	Age (Years)	Company Status
1	Seed	< 1.5	Product under development
2	Early	< 3	Product in testing or pilot production
3	Expansion	> 3	Product on sale - may not be profitable yet
4	Later	> 3	Product widely available - on-going revenue

1995 – 1998 (Pre-bubble, Bin 1) represents the pre-bubble years. Two strong and positive factors were extracted, plus a third weak factor.

Figure 18 visualises the loadings on the factor dominated by VC and online population variables. This is a one-dimensional plot; vertical spacing has been added between the data points to aid clarity.

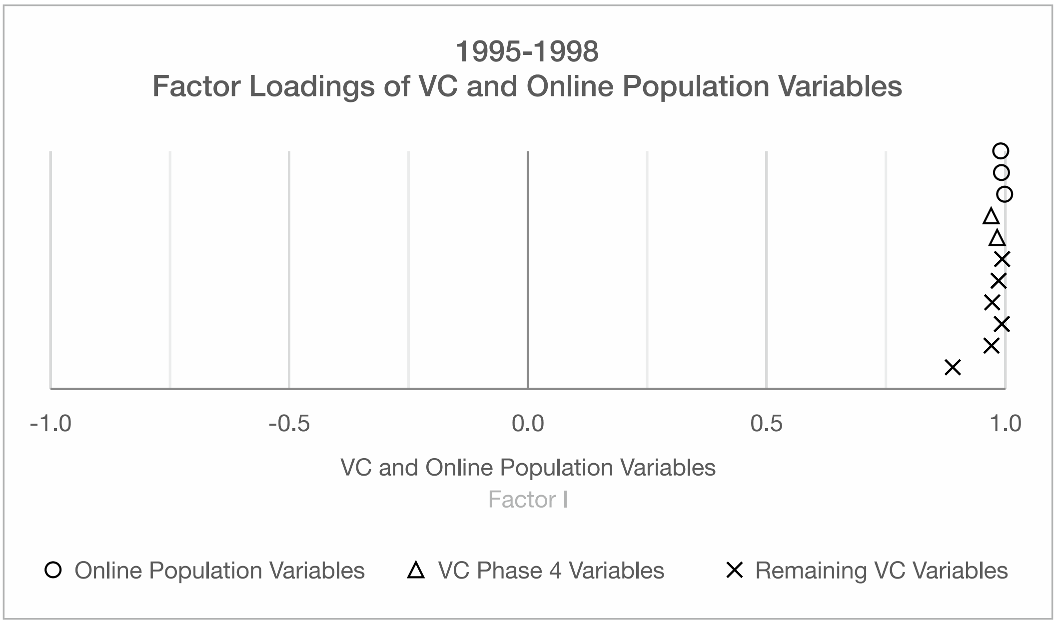

Figure 18 · 1995-1998 Factor Loadings of VC and Online Population Variables

Venture Capital and Online Population Variables (Factor I)
The positive strong factor loading of all variables can be seen in Figure 18. *Phase 1 VC Value* is notable as the weakest of the set. These results show that growth in venture capital investment and online population correlated with the factor during this period.

This suggests that, during this period, venture capital investment and online population were closely related.

Company Data Variables (Factor II)
This grouping of company data variables into one factor shows that the company data variables correlate better with their own independent factor than they do with any other factor in the analysis.

This suggests that, during this period, the observed financial performance of a company did not relate to any other data in the analysis.

Share Price has a weak factor loading in this component, indicating that it does not correlate with its peers as well as the peers do with each other.

This suggests that changes in Share Price have little in common with changes in other Company Data variables.

Share Price and Volume Traded Variables (Factor III)
This factor was sparsely populated, containing only *Share Price* and *Volume Traded (units). Share Price* is a strong positive contributor (0.83), while *Volume Traded* is weaker and negative (-0.51).

> *This suggests that as Share Price increases, the Volume Traded (units) decreases. In effect, as share prices increased, people traded fewer units of shares.*

2000 – 2001 (Mid-bubble, Bin 3) represents the during-bubble and collapse years.

Venture Capital and Online Population Variables (Factor I)
This factor contains the same sets as the corresponding factor in 1995–1998. Both variable sets exhibited strong loadings.

However, compared to 1995–1998, the online population variable set switched from positive weightings (≥ 0.99) to negative weightings (-1.00).

> *This suggests that the venture capital and online population datasets now oppose one another. Online population was increasing rapidly during this period, suggesting that venture capital activity/investment may have been decreasing at a similar rate.*

Figure 19 visualises the loadings on this factor. This is a one-dimensional plot; vertical spacing has been added between the data points to aid clarity.

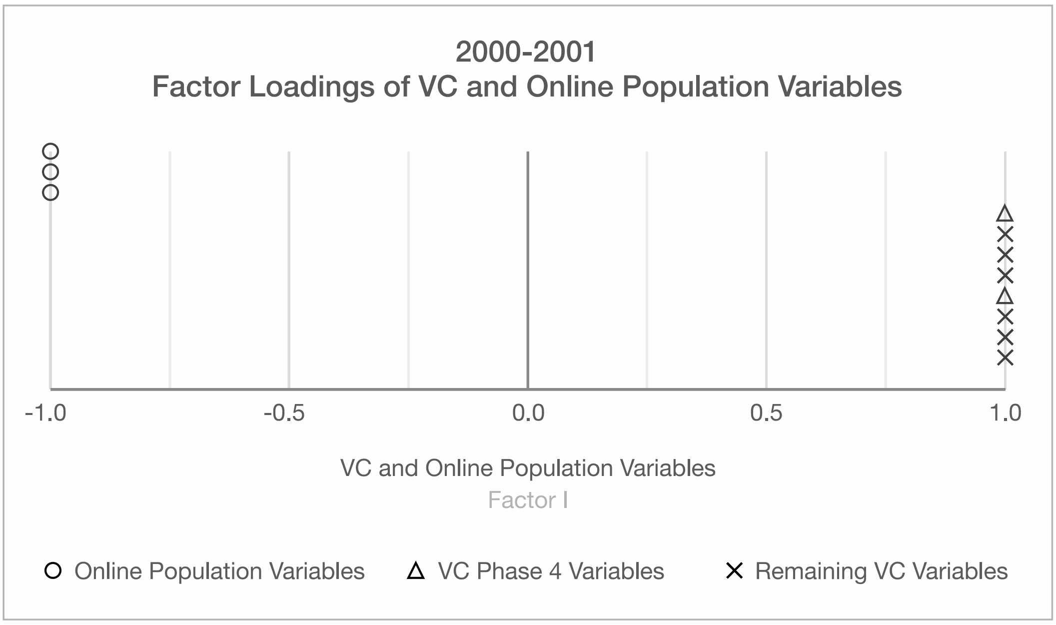

Figure 19 · 2000-2001 Factor Loadings of VC and Online Population Variables

Company Data Variables (Factor II)
This factor contains the same sets as the corresponding factor in 1995–1998.

Share Price was noted as a weak contributor to this factor in 1995–1998; it has recovered in this period. This variable and *Volume Traded (units)* remain the weakest variables in this period.

This period does not contain a third factor for *Share Price* and *Volume Traded (units)*, as it did previously. In 1995–1998, the third factor exposed a fall in *Volume Traded*, as *Share Price* increased. No such factor is present in this period.

> *This indicates an end to the previously observed inverse relationship between Share Price and Volume Traded.*

2003 – 2007 (Post-bubble, Bin 5) represents the years immediately following the bubble. It is formed of three factors, explaining 42.1%, 30.7% and 12.9% of variance respectively.

Figure 20 visualises the loadings on factors for Online Population and Venture Capital variables (Factors I and III respectively). The venture capital variables have formed an emergent factor, due to the restructuring of Factor I. It is included as a secondary dimension on this plot.

Figure 20 · 2003-2007 Factor Loadings of VC and Online Population Variables

Online Population Variables (Factor I)
This factor contained all the online population variables, and some venture capital variables. It is dominated by the online population variables.

The online population variables persist with strong positive loadings.

Some venture capital variables that were previously relevant in this weighting now have negligible loadings:

VC Phase 2 variables have changed from a strong positive weighting in 2000–2001 to negligible weightings during this period.

> *This suggests that there is little or no correlation between the VC Phase 2 variables and the factor. Given that online population was growing during this period, and has a strong positive loading, VC Phase 2 funding likely did not grow or shrink similarly. This concurs with findings in the literature review, as funding was scarce and venture capital investors took more care to perform due diligence in the years following the bubble. This is discussed further in section 7.2.*

VC Phase 4 variables have inverted to a strong negative loading (see △ markers in Figure 20).

> *Against a growing online population, this suggests that Phase 4 VC funding decreased during this period. Again, this concurs with findings in the literature review (discussed further in section 7.2). The collapse of so many companies immediately following this period may simply have eliminated many that would have been eligible for Phase 4 funding.*

Company Data Variables (Factor II)
Share Price and *Volume Traded (units)* were noted as having the weakest loadings of this factor in the previous period. In this period, the loading of *Share Price* relative to its peers has degraded.

> *This indicates that Share Price has a diminishing relationship with its peers, as its correlation with the factor degrades.*

Emergent Venture Capital Variables (Factor III)
This factor has emerged from Factor I, due to the restructuring of loadings therein. It contains a restructured alternative of the venture capital variable set.

Phase 1 variables have weaker weightings compared to those of *Phase 2*. This factor appears to have formed due to the poor correlation of *Phase 2* variables with the other venture capital variables in factor I. The *Phase 2* variables correlate best with this factor, which also contains the *Phase 1* variables.

> *This change indicates an emergent discontinuity in the venture capital variables. Factors I and III contain differing VC variables, suggesting that the behaviour of some venture capital phases is becoming different to others.*

2009 – 2012 (Post-bubble, Bin 7) represents more recent years, and was included as a contrast for 2003–2007, which was closer to the bubble.

Figure 21 visualises the loadings on factors for Online Population and Venture Capital variables (Factors II and III respectively). These are the development of Factors I and III in 2003–2007 respectively.

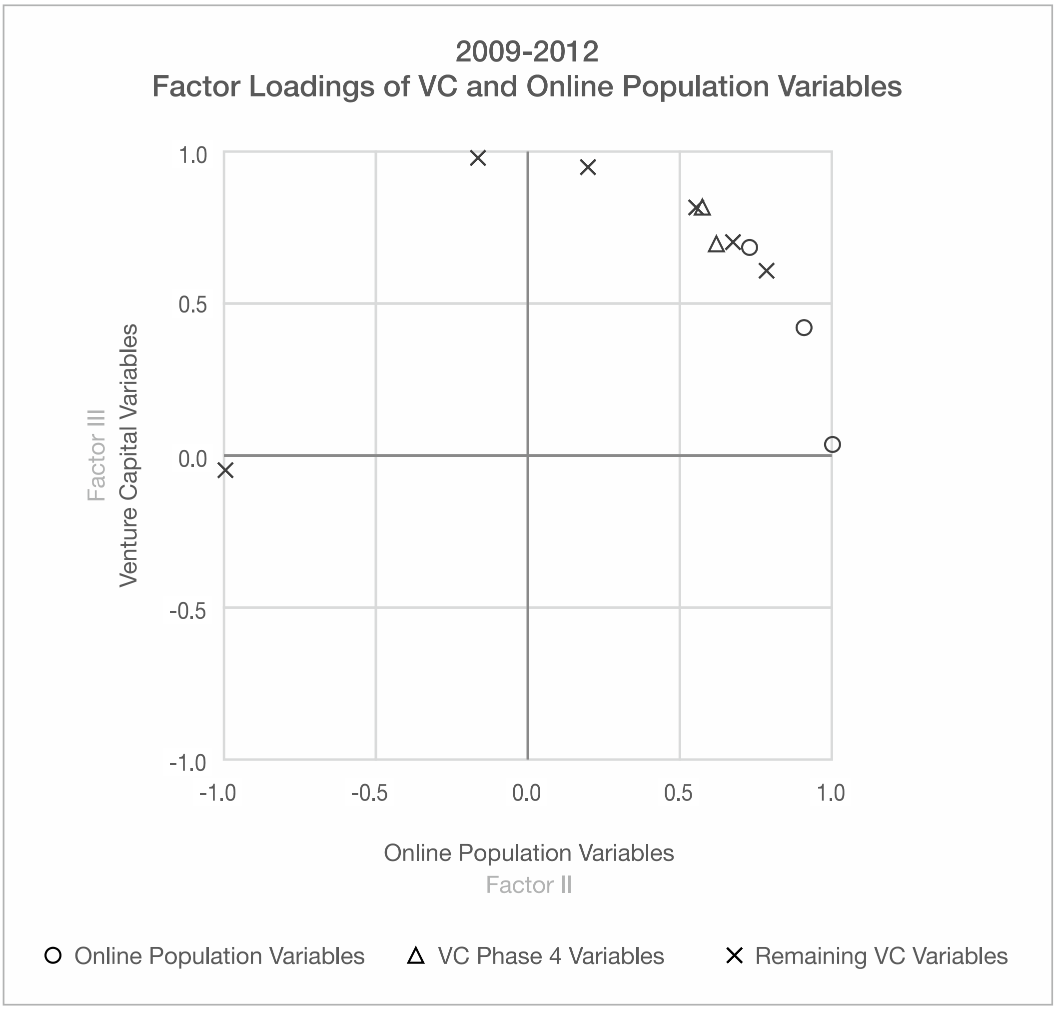

Figure 21 · 2009–2012 Factor Loadings of VC and Online Population Variables

Company Data Variables (Factor I)
For the first time, *Share Price* is sufficiently different from its peers that it has a negligible loading in this factor. It no longer correlates sufficiently with this factor to be relevant.

The range of loadings in this factor has been greater in previous periods, even when *Share Price* is disregarded (0.83 – 0.95, 0.81 – 0.96, 0.84 – 0.96). In this factor, the range is 0.88 – 0.94.

This indicates a marked change in how Share Prices relate to other company data variables within the dataset. It is discussed further below.

Venture Capital and Online Population Variables (Factor II)
The VC variables have restructured once again.

VC Phase 1 variables changed from 0.75 and 0.79 (in the comparative factor of 2003–2007) to -1.00 and -0.17 here, indicating a strong negative and negligible correlation with the factor respectively.

> *This change indicates that Phase 1 VC activity is not behaving in the same way as other venture capital phases, and almost in an opposing manner to the online population.*
>
> *In previous comparable factors, loadings for the VC Phase variable pairs have often closely matched one another. In this period, the difference between the Phase 1 variable pair is 0.83.*
>
> *This indicates a change in the relationship of Phase 1 activity variables; the value of each deal is likely decreasing (-1.00 correlation with the factor), while the number of deals made is remaining relatively static (-0.17 correlation with the factor).*

VC Phase 2 variables loaded negligibly to the comparable factor in the previous period. Here, they load positively.

> *This indicates a partial recovery in Phase 2 activities. The Value and Deals variables for this still exhibit an unusually high difference (0.23), but less than those observed in Phase 1 (discussed above). This indicates a change in the relationship between the number of deals and the value of those deals.*

VC Phase 3 variables loaded strongly positively to the comparable factor in the previous period. Here, they load moderately and negligibly.

> *This indicates a degradation of Phase 3 activities, in relation to the factor. Again, the marked difference between the loadings of these variables indicates a change in the relationship between them.*

VC Phase 4 variables loaded strongly negatively in the previous comparable factor (see △ markers in Figure 20). Here, they load moderately positively (0.57, 0.62).

> *This change indicates that VC Phase 4 activities have recovered to growth in relation to their peers.*

The online population variables exhibit strong positive correlation with the factor. However, the correlation is distinctly weaker than previous periods. Online population variables have always loaded ≥ 0.98 or -1.00 in previous periods. In this period, they have spread to 1.00, 0.91 and 0.73.

> *This change indicates that some differentiation is occurring between the three online population variables. The US variable is behaving similarly to previous periods, while the other two are quite different.*

Emergent Venture Capital Variables (Factor III)
This is a continuation of the emergent Factor III in 2003–2007. It previously contained only venture capital variables, and now additionally contains online population variables.

This factor has formed around the venture capital variables that had the weakest loadings in their usual factor (Factor II). Thus, *Phase 1 Deals* and *Phase 3 Deals* have the strongest factor loadings in this new factor.

> *This factor illustrates growth of the previously emergent Factor III. It contains more and stronger loaded variables than its counterpart in 2003–2007. This reinforces the notion that per-Phase VC activity is becoming increasingly disparate, and diverging from VC activity during the burst of the dot-com bubble in 2000–2001.*

Share Price Variable (Factor IV)
The loading of share price within the typical Company Data factor (Factor I) has degraded to the point that the variable has restructured into an emergent factor.

> *This factor demonstrates that the correlation of share price with its peers in the company data variable set has degraded to such a level that they are no longer related.*

7.2 Interpretation in Relation to Literature

This section presents an interpretation of the results in connection with the narrative of the dot-com bubble, and identifies observed changes that are indicative of the situation.

1995 – 1998 (Bin 1)

Prior to the bubble bursting, online population and venture capital investment correlated strongly with a single factor. All these variables were growing, similarly, during this period. Notably, all phases of venture capital investment correlated similarly strongly with the factor.

The revelations above apply to the *whole* period. So, as early as the 1995-96 analysis period, the interaction between online population and venture capital investment could have been observed.

Valliere and Peterson's research (2004) found that venture capitalists were willing to leave their area of expertise in order to pursue technology companies during this period.

> *While the analysis performed herein cannot conclusively judge the 'quality' of venture capital funding on offer, it does show a mass influx of venture capital funding into the technology sector during this period. This concurs with Valliere and Peterson's remarks that many venture capital companies were willing to enter an unfamiliar sector at the time.*

Sahlman and Stevenson (1985) introduced the concept of Capital Market Myopia, albeit in relation to another technology sector bubble. In this phenomenon, the collective actions of a large number of investors delivers an overall detrimental effect to the sector involved.

> *A form of Capital Market Myopia may have been in effect during this period. The result was similar to that observed by Sahlman and Stevenson (i.e. a funding collapse leading to insolvent companies). The evidence presented by data in this period shows a clear and consistent growth in venture capital investment, seemingly in concert with Online Population. Given this, and that the bubble burst in the following period, a Myopia-like situation seems likely.*

2000 – 2001 (Bin 3)

During and after the bubble, the relationship between online population and venture capital inverted. Online population continued to grow, while venture capital investment collapsed. Thus far, this concurs with findings in the literature review. This can be seen in the switching of online population variables to a strong negative loading, in 2000–2001.

Sahlman (1997) was cited stating "*from whom you raise capital is often more important than the terms*", emphasising the quality of venture capital funding as a key component to the success of a start-up. The advice and services provisioned are almost as necessary as the capital.

Hsu (2005) supported this, and demonstrated that the price of venture capital (in terms of capital received per percentage share) fluctuates with the quality of support on offer.

> *The literature discussed in this period describes a broad spectrum of VC funding/support, offering various levels of quality and competency.*
>
> *Venture capital funding was plentiful during 1995–1998.*
>
> *Putting these points together shows that in 1995–1998, poor-quality venture capital was abundantly available at low prices. That is, venture capitalists with poor technology sector experience were eager to support young technology companies, many of which were equally inexperienced in acquiring VC funding.*
>
> *The reversal of VC funding seen in this period indicates a sudden mass exodus of venture capital funding provision from the sector. This reflects the notion that much of the VC funding on offer was of poor quality.*
>
> *This concurs with assertions in the literature.*

2003 – 2007 (Bin 5)

Contemporary literature has a tendency to focus upon the drivers that led to a bubble, and present little detail of what might happen *after* a collapse.

However, the analysis results for these “post-collapse” periods may be analysed in terms of how they differ from the literature’s description of pre-bubble states.

In 2003–2007, immediately following the bubble, online population continues to grow. When combined in a factor with venture capital investment, some variables of venture capital investment do not correlate at all with the overall factor. Others correlate positively and strongly with the factor, and some correlate negatively and strongly. This is a marked change from the observed behaviour of these variables in 1995–1998 and 2000–2001.

To summarise, the four VC Phases behaved coherently before the bubble, as has been explained in the preceding sections. Immediately following the bubble, they no longer behave coherently, and appear fragmented.

This presents a market state quite unlike any described in the bubble-related literature. Where venture capital investment previously behaved as one unit, the separate phases now behave quite differently from one another. This may be indicative of increased due diligence, and a nature of caution.

2009 – 2012 (Bin 7)

In the later years, the VC Phases remain incoherent and fragmented.

The loadings of venture capital investment variables in this period are different to those in 2003–2007. Thus, the nature of venture capital is changing in some way. This observation is explored further in section 8.3.2.

7.2.1 Relation to the Conceptual Model

This section relates the findings to the conceptual model that was created prior to the experimental design (see Figure 10).

A notable outcome of the factor analysis is the turbulent nature of factor loadings for certain variables, post-bubble. This suggests that links in the conceptual model would adapt depending upon conditions in the sector. The model was, by design, created using literature on behaviours and observations during bubble periods. Thus, the illustrated model would be relevant during certain scenarios only.

While this analysis was not intended to test the conceptual model directly, some of the links made in the model can be related to findings in the analysis.

Following is a list of assertions made in the model, which may be tested using results from the factor analysis. *Relevant Variables* represent log10 variants where log10 variants were used in the analysis.

Investment Volume* correlates positively with *Valuation

Relevant Variables volume_units, volume_price, mktcap

It was expected that the volume of securities publically traded would be reflected in the valuation of a company.

In all periods, the relevant variables loaded positively into the same factor. In all cases, loadings for these variables were ≥ 0.81. This suggests that they do covary.

VC Investment* correlates positively with *Investment Volume

Relevant Variables vc_phase{1-4}_value, vc_phase{1-4}_deals, volume_units, volume_price

It was expected that the level of venture capital investment would be reflected somehow in the volume of securities publically traded.

In all periods, relevant loadings for the venture capital variables and Volume Traded variables were in separate factors. This suggests that the two groups of variables do not covary.

Product Adoption* correlates positively with *Market Share

Relevant Variables sales, marketshare

Assumption That the *sales* variable represents product adoption.

It was expected that product adoption (mapped here to company sales) would relate to the market share of the company.

In all periods, the relevant variables loaded positively into the same factor. In all cases, loadings for these variables were ≥ 0.92. This suggests that they do covary.

Three relationships in the model were identified for scrutiny at this stage:

1. **Technological Development positively influences Investment Volume**
   "Technological development" is an abstract concept, and its closest matching variables in the dataset are those of online population. The analysis found that this relationship was upheld during growth of the dot-com bubble – venture capital investment related strongly and positively to the online population variables.

2. **Investment positively influences Valuations**
   The analysis showed that venture capital investment was never linked to the valuations of companies in the technology sector. As such, this relationship has been disproved.

3. **Market Share of a company positively influences Investment in the company**
   The analysis showed that, in all periods, the market share of a company strongly related to *public* investment in the company.

Critically, the model has illustrated the link between technological development and venture capital investment. The analysis has demonstrated that this link exists, and that it was strongest during the growth stage of the dot-com bubble. Interpretation in Relation to the Research Question

7.3 Interpretation in Relation to the Research Question

At this point, it is helpful to remind ourselves of the overarching research question:

> *This research aims to identify, quantitatively, the degree to which companies involved with the web sector have exhibited a repeat of the bubble-like state that was observed during 1999-2001.*
>
> *Excerpt from Section 1.1.*

The predicate of *"companies involved with the web sector"* has been met by:

1. Selecting a source for venture capital data that offers per-sector data grouping. This allowed the analysis to look only at venture capital data for the relevant sector.

2. Selecting only companies involved with the "technology/software" sector, for inclusion in the *company data* variable set.

Following is a discussion of how each period supports and/or answers the research question.

1995 – 1998 (Bin 1)

Identified characteristics of the sector that indicated the build-up of a bubble.

This period necessarily established a profile of the conditions that existed prior to the dot-com bubble, so that future years could be checked for "*a repeat of the bubble-like state that was observed during 1999-2001".*

This period found that as early as 1995-96, venture capital investment was flowing into the technology sector, precipitating the bubble. Given that the bubble burst in 2000, this establishes a lead-time of at least four years. Further analysis with older data could determine when this pattern actually started, and is discussed further in section 8.3.1.

2000 – 2001 (Bin 3)

Showed the immediate effects of the bubble bursting.

As a measure of extreme circumstances, this period offers little insight into how future bubbles may be characterised or identified.

However, it does reflect findings in the literature, lending credibility to the findings.

2003 – 2007 and 2009 – 2012 (Bins 5 and 7)

Showed the state of the sector following the collapse of the bubble.

These periods must be analysed for "*a repeat of the bubble-like state that was observed during 1999-2001".*

Given that 1995–1998 exhibited evidence of bubble formation in venture capital investment as early as 1995-96, it would be prudent to check these periods for similar circumstances (i.e. circumstances that *precipitated* the bubble-like state in 2000–2001).

This discussion has already explained that 2003–2007 and 2009–2012 exhibit a fragmented view of VC Funding, and that this view is quite different from the coherent view observed in 1995–1998.

Thus, the bubble-like state has not been re-observed, as venture capital investment has not behaved as coherently as it did prior to the bubble.

7.4 The Value of Company Data

When the idea of an economic bubble is discussed, it often evokes questions surrounding the valuation and share price of involved companies. This is reflected in contemporary media (for examples, refer to Braithwaite, 2011; Dembosky & Demos, 2011; Gelles, 2011a, 2011b).

However, the results of the analysis showed that company/stock performance data correlated only within its own factor, and *never* with *venture capital* or *online population* data.

Given the focus of literature upon *venture capital*, it was anticipated that the performance of companies within the technology sector would be linked somehow to the funding entering it, and to the number of people using the Internet.

This raises important questions about how analysts judge whether we are ‘approaching’ or ‘in’ a bubble.

It may also explain some of the reckless activity by venture capital investors in earlier periods. These investors may have observed growing valuations on established technology companies, and mistakenly assumed that the start-ups they sponsor could enjoy similar growth. This implies that the state of the sector at large is a poor choice of metric when evaluating start-ups investment options.

Despite consistently forming separate factors, the company data variables have experienced some turbulence. Specifically, the *Share Price* variable…

- Formed its own factor in 1995-1998.
- Re-joined other *company data* variables in a single factor in 2000-2001 and 2003-2007 (Bins 3 and 5).
- Formed its own factor in 2009-2012.

This suggests that *Share Price* failed to correlate well alongside other *company data* in 1995–1998 and 2009–2012. This was unexpected. The fact that this behaviour first occurred prior to the bubble, and has just re-occurred, warrants further exploration.

7.5 Summary

Prior to the dot-com bubble, the rise in online population and venture capital investment appear closely linked. After the bubble, this link begins to degrade. Furthermore, the nature of venture capital investment changes markedly immediately after the bubble. Prior to the bubble, the four venture capital phases behave in a similar fashion. After the bubble, they behave very differently to each other.

The link between online population and venture capital investment appears to be a key driver of the formation of the dot-com bubble. Given the weakened nature of the link in more contemporary times, the same circumstances have not re-occurred.

Three distinct stages of bubble formation were noted in the literature review, and related to the dot-com bubble (see Table 3, on page 23). These periods can be linked directly to observations in each period.

- **Initial Development**
  Not seen in these results (discussed further in the following chapter).

- **Growth and Performance**
  Observed in 1995-1998.

- **Collapse**
  Observed in 2000-2001.

The *company data* variables were notable in their lack of relation to anything else in the analysis. At no analysed point has any *company data* variables relate to the driving force of venture capital. This suggests that analysing such metrics as share prices and company reports for signs of an economic bubble (as often postulated in mainstream media) is not a productive endeavour.

The following chapter is the end of this work. It presents key findings, contributions, and items of work that may be explored further in the future.

Chapter 8. Conclusion

This thesis aimed to investigate the nature of economic bubbles, particularly within the technology sector, and identify whether circumstances surrounding the dot-com bubble of 1999-2001 have re-occurred.

By analysing historic bubbles, it has been found that a lack of knowledge about an investment causes speculative investment. This serves to increase valuations of securities by investment companies, and contributes to the creation of a bubble. Of the economic bubbles that were analysed, it appears that legislation has been adapted in each case to prevent the situation re-occurring. However, further economic bubbles have occurred, sometimes for reasons that are later prohibited in legislation, and sometimes by purely legitimate means.

By analysing more recent technology sector bubbles, finer details about the formation of these bubbles have been found. For instance, the valuation rises of companies in these bubbles have been found to represent a positive feedback loop in Valliere and Peterson's model. In another scenario, the theory of Capital Market Myopia was proposed, which arises in sectors that lack a dominant company.

Prevalent throughout all the scenarios analysed is the notion that many investors lacked sufficient knowledge about the investment sector, which inhibited their ability to perform due diligence upon the investment opportunity.

Cloud service offerings by traditional (and newcomer) technology companies have enabled web start-ups to offer products beyond their traditional capability, and scale costs according to demand. The interdependence supplier and customer in this B2B relationship is key – large companies now rely upon many smaller, relatively young companies.

To answer the question of whether a bubble has re-occurred in the technology sector, a conceptual model was first developed, based upon findings from the literature. This presented some hypothetical links between certain datasets. It was hypothesised that relationships between certain metrics in the datasets would reveal themselves at the point of the dot-com bubble. The same data for subsequent periods could then be analysed in the same way to identify whether the same relationships occur. An experiment was designed to test this.

Data to support the experiment was collected and combined from various sources, in a format that allowed rapid, agile adjustments and analysis. The collected data was used as the input of a statistical analysis, which revealed correlative relationships between certain metrics at key points in time.

8.1 Key Findings

This section presents key findings from the discussion, to answer and supplement the research question.

8.1.1 The emergent nature of venture capital prevents a new bubble

The discussion revealed that reckless venture capital investment caused the dot-com bubble. The results show that in the years since, the incoherent nature of venture capital investment has prohibited the creation of a financial ecosystem, which could cause another bubble by the same means.

The results showed that venture capital investment during the dot-com bubble had similar correlative properties to the online population. This styles the growth in online population as a driver for venture capital investment during the dot-com bubble.

8.1.2 The dot-com bubble was forming as early as 1995

The relationship between venture capital investment and the online population, described above, was observed as early as 1995. This suggests that the ecosystem of venture capital investment was contributing to an economic bubble, at least five years before the bubble burst.

This work analysed data starting from 1995, and so cannot conclusively state when this behaviour in venture capital activity began. The results for the 1995-1998 period show venture capital investment and online population correlating very strongly with the same factor. This prevents the creation of a reliable backward regression.

8.1.3 Other influences may drive future venture capital

This work has shown, quantitatively, that venture capital investment during the dot-com bubble was related to the growth in the online population.

It would be prudent to note that online population may not always be the driver for venture capital activity. The results show it to have been during the dot-com bubble, but investment may have different drivers in future bubbles. Nevertheless, the technique employed herein may be used to identify such behaviours, provided the correct variables are provided to the analysis.

8.2 Contribution

Previous works in this area have created qualitative conceptual models that represent behaviours during bubble situations.

This work has created an expanded conceptual model of economic bubble development, based upon wider literature, and used it to design a quantitative method of analysing market data.

This was developed into a unique and complimentary technique of data collection, aggregation, and quantitative analysis to compare a period in time with a state that is known to be a bubble. Interpretation of the results tells how “bubble-like” the analysed period is.

Through analysis of contemporary periods, this work has found that the bubble-state observed during 1999-2001 was developing at least as early as 1995. It has found that the behaviour of venture capital investment post-bubble has not enabled the formation of a similar bubble.

The technique of data collection and analysis employed herein is repeatable. Given suitably rich data, the same technique will be applicable to future time segments. Furthermore, the technique has been developed with the intention of automation in mind. Such an automation would utilise live data feeds, and only require human interaction to verify interpretation of the results.

8.3 Future Work

In light of this research, this section suggests some follow-up studies and enhancements to the work.

Factor Analysis appears to have been an appropriate analysis technique to answer the research question, particularly given the type and volume of data. However, alternative techniques may be more suitable for some of the follow-up work described below. Specifically, a form of binned clustering may more suited to the suggested future work on venture capital trends.

8.3.1 Did the dot-com bubble 'start' before 1995?

This research created a dataset that started in 1995. This date was chosen somewhat arbitrarily, on the following rationale:

- The web had just come into existence, so sufficient numbers would appear in the *online population* statistics.
- The assumption that prerequisites for the bubble must have appeared later than 1995.

However, this research found a strong relationship between online population and venture capital investment, from the very beginning of the data.

The research could be easily expanded by sourcing pre-1995 data and re-running the analysis to determine when this relationship began.

Online population data is available from as early as 1988; presumably statistics/reports from BBS and other pre-web ISPs.

Answering this question would quantify the lead-time of the dot-com bubble, in terms of the investment that caused it.

The factor analysis technique described and applied in this thesis may be suitable for answering this question, provided an expanded dataset (encompassing pre-1995) is used.

8.3.2 Is Venture Capital investment converging on another bubble?

This research created a dataset that ended in 2012.

During the discussion, it was noted that the nature of venture capital activity has changed in the two bins following the bubble. Further analysis may be applied to determine the nature of this change, and ascertain whether it represents a tendency to the state that precipitated the dot-com bubble.

A wider gamut of data may also be explored to help answer this question; the requisite data for 2013 is available now, and a comprehensive set for 2014 will be available before year-end 2015.

Answering this question would quantify the rate of divergence from a bubble-like situation. Alternatively, and with the application of regression, it would provide an estimated date of the next convergence upon bubble-like circumstances.

A form of regression analysis may be suitable for investigating this question. Given input data for a range of periods, this may find a convergence, divergence or no change/relationship, relative to known bubble circumstances.

8.3.3 Did venture capitalists "flock" to technology during 1995-2000?

This research was limited to analysing venture capital data from the "technology/software" sector only. Literature suggests that venture capitalists were willing to leave their areas of expertise to gain exposure to the technology sector. Thus, an analysis could be designed to test whether relative levels of venture capital investment left other sectors while arriving in the technology sector, while the dot-com bubble was forming.

This assertion may be checked and quantified by analysing *all* venture capital data on a per-sector basis, and looking for a removal of funding to the technology sector during the years when technology venture capital investment was growing.

Answering this question would quantify the effect of the dot-com bubble on the wider venture capital market, and those that depend upon it. Some index of *perceived confidence* in investors could potentially be derived by analysing relative per-sector changes in venture capital investment.

A form of binned clustering may be suitable for investigating this question. The changes in clustering between time period bins may yield an answer.

8.3.4 Automated, Continuous Analysis

While the data collection and analysis herein was performed "by hand", an algorithm could be created to automatically source, sanitise, combine, and analyse data in an ongoing manner, as it becomes available.

The present reporting periods for requisite data would enable analysis at least every quarter. The results could be automatically interpreted, statistically, but human interpretation and description would be recommended in all cases.

The analysis herein has been conducted with annual data. Automatically analysing data on a per-quarter basis, as the data affords, would reveal additional benefits and complications. To support the potential for frequent analysis and reporting, the statistical and review techniques may require adaptations to be compatible with smaller bins and less data. Alternatively, a staggered approach may be taken, where quarterly data is regularly rolled up into a per-year analysis.

References

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., … Zaharia, M. (2009). Above the clouds: A Berkeley view of cloud computing. doi:10.1.1.361.8444

Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A. D., … Rabkin, A. (2010). A view of cloud computing. *Communications of the ACM*, *53*(4), 50. doi:10.1145/1721654.1721672

Balci, O. (1998). *Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice*. (J. Banks, Ed.). New York: Wiley.

Banks, C. (2009). What is Modeling and Simulation? In J. Sokolowski & C. Banks (Eds.), *Principles of Modeling and Simulation* (pp. 3–24). Wiley Publishing.

BBC. (2002a). BBC Money Box 21.09.2002 Transcript. Retrieved January 28, 2015, from http://news.bbc.co.uk/hi/english/static/audio_video/programmes/money_box/transcripts/02_09_21.txt

BBC. (2002b). BBC News | Business | Nectar launch leaves a bitter taste. Retrieved January 28, 2015, from http://news.bbc.co.uk/1/hi/business/2268797.stm

BBC. (2002c). BBC News | UK | Census website a crashing success. Retrieved January 28, 2015, from http://news.bbc.co.uk/1/hi/uk/1737861.stm

BBC. (2002d). BBC News | UK | Census website still offline. Retrieved January 28, 2015, from http://news.bbc.co.uk/1/hi/uk/2136572.stm

BBC. (2011). Zynga shares fall 5% on stock market debut. Retrieved January 28, 2015, from http://www.bbc.co.uk/news/business-16214331

Berners-Lee, T. (2010). Long Live the Web: A Call for Continued Open Standards and Neutrality. Retrieved from http://www.scientificamerican.com/article/long-live-the-web/

Braithwaite, T. (2011). Goldman eyes high return on Facebook. *Financial Times*, 1.

Buyya, R., Yeo, C. S., & Venugopal, S. (2008). Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. *10th IEEE International Conference on High Performance Computing and Communications*, 5–13. doi:10.1109/HPCC.2008.172

Cellier, F. (1991). *Continuous System Modeling*. Springer-Verlag.

Cochrane, J. H. (2005). The risk and return of venture capital. *Journal of Financial Economics*, *75*(1), 3–52. doi:10.1016/j.jfineco.2004.03.006

Coyle, G. (1999). The practice of system dynamics: milestones, lessons and ideas from 30 years experience. *System Dynamics Review*, *14*(4), 343–365.

Crain, M. (2014). Financial markets and online advertising: reevaluating the dotcom investment bubble. *Information, Communication & Society*, *17*(3), 371–384. doi:10.1080/1369118X.2013.869615

Das, A. (2011a). J.P. Morgan Plans New-Media Fund. Retrieved from http://online.wsj.com/articles/SB10001424052748703515504576142410532498754

Das, A. (2011b, February 14). J.P. Morgan Plans New-Media Fund. *The Wall Street Journal*.

Dembosky, A., & Demos, T. (2011). Zynga set to rival networking peers with 2bn IPO. *Financial Times*, 19. Retrieved from http://www.ft.com/cms/s/0/432c481a-a1dc-11e0-b485-00144feabdc0.html#axzz3KaWF7Ycw

Dhar, S., & Varshney, U. (2011). Challenges and business models for mobile location-based services and advertising. *Communications of the ACM*, *54*(5), 121. doi:10.1145/1941487.1941515

Eatwell, J., Milgate, M., & Newman, P. (1987). *The New Palgrave: a Dictionary of Economics*. New York: Palgrave.

EVCA. (2014). Annual activity statistics. Retrieved March 24, 2015, from http://www.evca.eu/research/activity-data/annual-activity-statistics/

Facebook. (2012, February). Form S-1 Registration Statement. Washington D.C. 20549. Retrieved from http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm

Falch, M., Henten, A., Tadayoni, R., & Windekilde, I. M. (2009). Business Models in Social Networking. In *CMI International Conference (Social Networking and Communities)*. Retrieved from http://vbn.aau.dk/files/19150157/Falch_3.pdf

Fenn, J. (2010). *Hype Cycle for Emerging Technologies, 2010*. Retrieved from https://www.gartner.com/doc/1417913/gartners-hype-cycle-special-report#a932235129

Fontaine, M., Cook, D., Combs, D., Sokolowski, J., & Banks, C. (2009). Modeling and Simulation: Real-World Examples. In J. Sokolowski & C. Banks (Eds.), *Principles of Modeling and Simulation* (pp. 181–245). New York: Wiley Publishing. doi:10.1002/9780470403563.ch8

Forrester, J. (1961). *Industrial Dynamics*. MIT Press.

Forrester, J. (1968). *Principles of Systems*. Wright-Allen Press.

Forro, Z., Cauwels, P., & Sornette, D. (2011). *Valuation of Zynga*. Retrieved from http://ideas.repec.org/p/arx/papers/1112.6024.html

Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). Cloud Computing and Grid Computing 360-Degree Compared. In *Grid Computing Environments Workshop* (pp. 1–10). IEEE. doi:10.1109/GCE.2008.4738445

Garber, P. (1989). Tulipmania. *Journal of Political Economy*, *97*(3), 535–560. Retrieved from http://econpapers.repec.org/RePEc:ucp:jpolec:v:97:y:1989:i:3:p:535-60

Garber, P. (1990). Famous First Bubbles. *The Journal of Economic Perspectives*, *4*(2), 35–54. Retrieved from http://grizzly.la.psu.edu/~bickes/garber.pdf

Gartner. (2013). Gartner Press Release: Worldwide Public Cloud Services Market to Total $131 Billion. Retrieved March 24, 2015, from http://www.gartner.com/newsroom/id/2352816

Gelles, D. (2011a). Twitter is valued at 4.5bn as JPMorgan fund targets 10% stake. *Financial Times*, 1.

Gelles, D. (2011b, August). Facebook's "value" soars as investors seek pre-IPO stake. Retrieved from http://www.ft.com/cms/s/0/054fc0ee-afa7-11df-b45b-00144feabdc0.html#axzz20tV7alcc

Google. (2011). Zynga Inc: NASDAQ:ZNGA - Google Finance. Retrieved January 28, 2015, from http://www.google.com/finance?cid=481720736332929

Google. (2014). Privacy Policy. Retrieved January 28, 2015, from https://www.google.com/intl/en-GB/policies/privacy/

Guha, S., Cheng, B., & Francis, P. (2011). Privad: practical privacy in online advertising. In *Proceedings of the 8th USENIX conference on Networked systems design and implementation (NSDI'11)* (pp. 169–182). Berkeley: USENIX Association. Retrieved from http://dl.acm.org/citation.cfm?id=1972457.1972475

Harris, L., & Dennis, C. (2011). Engaging customers on Facebook: Challenges for e-retailers. *Journal of Consumer Behaviour*, *10*(6), 338–346. doi:10.1002/cb.375

Hawkins, R. (2004). Looking beyond the dot com bubble: exploring the form and function of business models in the electronic marketplace. In B. Preissl, H. Bouwman, & C. W. Steinfield (Eds.), *E-Life After the Dot Com Bust* (pp. 65–79). TNO, Institute for Strategy, Technology and Policy, Delft, The Netherlands: Springer.

Hellman, T., & Puri, M. (2002). Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence. *The Journal of Finance*, *57*(1), 169–197. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1540-6261.00419/abstract

Hoofnagle, C. J., Urban, J. M., & Li, S. (2012). Privacy and Modern Advertising: Most US Internet Users Want "Do Not Track" to Stop Collection of Data about their Online Activities. Retrieved from http://papers.ssrn.com/abstract=2152135

Hsu, D. (2005). What do Entrepreneurs Pay for Venture Capital Affiliation? *The Journal of Finance*, *59*(4), 1805–1844.

ITU. (2014). 2005-2014 World Aggregate Data. Retrieved February 17, 2015, from http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2014/ITU_Key_2005-2014_ICT_data.xls

Ivanov, V., & Xie, F. (2010). Do Corporate Venture Capitalists Add Value to Start-Up Firms? Evidence from IPOs and Acquisitions of VC-backed Companies. *Financial Management*, *39*(1), 129–152. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1755-053X.2009.01068.x/abstract

Johnson, M. W. (2010). *Seizing the White Space: Business Model Innovation for Growth and Renewal*. Harvard Business Press.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. Retrieved from http://www.jstor.org/discover/10.2307/1914185?uid=3738032&uid=2&uid=4&sid=21105330502383

Kuittinen, J., Kultima, A., Niemelä, J., & Paavilainen, J. (2007). Casual games discussion. In *Proceedings of the 2007 conference on Future Play* (pp. 105–112). ACM Press. doi:10.1145/1328202.1328221

Law, J. (1705). *Money and Trade Considered With a Proposal for Supplying the Nation with Money*. Heirs and Successors of Andrew Anderson. Retrieved from https://archive.org/details/moneytradeconsid00lawj

Lenk, A., Klems, M., Nimis, J., Tai, S., & Sandholm, T. (2009). What's inside the Cloud? An architectural map of the Cloud landscape. In *2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing* (pp. 23–31). IEEE. doi:10.1109/CLOUD.2009.5071529

Linden, A., & Fenn, J. (2003). *Understanding Gartner's Hype Cycles*. Retrieved from http://www.ask-force.org/web/Discourse/Linden-HypeCycle-2003.pdf

McCartney, S., & Arnold, A. J. (2003). The railway mania of 1845-1847: Market irrationality or collusive swindle based on accounting distortions? *Accounting, Auditing & Accountability Journal*, *16*(5), 821–852. doi:10.1108/09513570310505970

McGrath, R. G. (2010). Business Models: A Discovery Driven Approach. *Long Range Planning*, *43*(2-3), 247–261. doi:10.1016/j.lrp.2009.07.005

Microsoft. (2014). World Development Indicators | Microsoft Azure Marketplace. Retrieved February 17, 2015, from https://datamarket.azure.com/dataset/worldbank/worlddevelopmentindicators

Murugesan, S. (2007). Understanding Web 2.0. *It Professional*, *9*(4), 34–41. doi:10.1109/MITP.2007.78

Musser, J., & O'Reilly, T. (2006). *Web 2.0 Principles and Best Practices*. *O Reilly Media Inc* (Vol. 2007). doi:10.1109/MITP.2007.4374050

O'Hara, M. (2008). Bubbles: Some Perspectives (and Loose Talk) from History. *Review of Financial Studies*, *21*(1), 11–17. doi:10.1093/rfs/hhn001

O'Reilly, T. (2005, September). What Is Web 2.0? *Pattern Recognition*. San Francisco: Retrieved from http://oreilly.com/web2/archive/what-is-web-20.html

O'Reilly, T. (2010). Healthcare and the Internet Operating System. *Health 2.0 Conference*. San Francisco. Retrieved from http://www.health2con.com/tv/keynote-with-tim-oreilly/

O'Reilly, T., & Battelle, J. (2009). *Web Squared: Web 2.0 Five Years On*. Retrieved from http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf

Odlyzko, A. (2010). Collective Hallucinations and Inefficient Markets: The British Railway Mania of the 1840s. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1537338

Oliva, R., Sterman, J. D., & Giese, M. (2003). Limits to growth in the new economy: exploring the "get big fast" strategy in e-commerce. *System Dynamics Review*, *19*(2), 83–117. doi:10.1002/sdr.271

Pack, H. (2002). Due Diligence. In G. Picot (Ed.), *Handbook of International Mergers and Acquisitions: Preparation, Implementation and Integration* (pp. 153–183). Palgrave Macmillan.

Pierrakis, Y. (2010, July). Venture Capital Now and After the Dotcom Crash. Retrieved from http://www.nesta.org.uk/publications/venture-capital-now-and-after-dotcom-crash

PriceWaterhouseCoopers, & National Venture Capital Assciation. (2014, January). Historical Trend Data. Retrieved from https://www.pwcmoneytree.com

PWC MoneyTree. (n.d.). PWC MoneyTree Definitions. Retrieved June 5, 2015, from https://www.pwcmoneytree.com/Definitions/Definitions#Stage

Radoff, J. (2011). *Game On: Energize Your Business with Social Media Games*. Wiley Publishing.

Sahlman, W. (1997). How to Write a Great Business Plan. *Harvard Business Review*, *75*, 98–108. Retrieved from https://hbr.org/1997/07/how-to-write-a-great-business-plan

Sahlman, W. A., & Stevenson, H. H. (1985). Capital market myopia. *Journal of Business Venturing*, *1*(1), 7–30. doi:10.1016/0883-9026(85)90004-7

Sfetcu, N. (2014). *Internet Marketing, SEO & Advertising* (2nd ed.). Nicolae Sfetcu.

Sharespost. (2011). Twitter Valuation - Financings. Retrieved June 9, 2011, from https://www.sharespost.com/companies/twitter/valuation

Shea, G. S. (2004). *South Sea Company Subscription Shares and Warrant Values in 1720*. University of St. Andrews. Centre for Research into Industry, Enterprise, Finance and the Firm. Retrieved from http://econpapers.repec.org/paper/sancrieff/0411.htm

Shea, G. S. (2007). Understanding financial derivatives during the South Sea Bubble: the case of the South Sea subscription shares. *Oxford Economic Papers*, *59*(suppliment 1), i73–i104. doi:10.1093/oep/gpm031

Siegel, J. (2003). What Is an Asset Price Bubble? An Operational Definition. *European Financial Management*, *9*(1), 11–24. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1468-036X.00206/full

Smant, D. (n.d.). South Sea Bubble. Retrieved March 15, 2012, from https://sites.google.com/site/davesmant/monetary-economics/famous-first-bubbles/south-sea-bubble

Smith, S. (2003). A Discussion of the Hype Cycle. Retrieved from http://www.davewsmith.com/ayewiki/scribble.cgi?read=HypeCycle

Sokolowski, J. (2009). Simulation and Data Dependency. In J. Sokolowski & C. Banks (Eds.), *Principles of Modeling and Simulation* (pp. 91–119). New York: Wiley Publishing.

Sterman, J. (2001). System Dynamics Modeling: Tools for Learning in a Complex World. *California Management Review*, *43*(4), 8–25. Retrieved from http://www.systemdynamics.org/conferences/2002/proceed/papers/CAREVIEW/C2STERMA.PDF

Stone, B. (2008). Tracking [US Super Tuesday] Candidates on Twitter. Twitter. Retrieved from http://blog.twitter.com/2008/02/tracking-candidates-on-twitter.html

Sudek, R. (2006). Angel Investment Criteria. *Journal of Small Business Strategy*, *17*(2), 89–103. Retrieved from http://www.angelcapitalassociation.org/data/Documents/Resources/AngelGroupResarch/1d - Resources - Research/5 Sudek_JSBS_Article_-_Investment_Criteria.pdf

Taylor, J. (2006). *Creating capitalism: joint-stock enterprise in British politics and culture, 1800-1870*. Boydell & Brewer.

Telegeography. (2009). *Global Internet Geography - Executive Summary*. Retrieved from https://www.telegeography.com/page_attachments/products/website/research-services/global-internet-geography/0005/1382/GIG_Executive_Summary.pdf

Temin, P., & Voth, H.-J. (2004). Riding the South Sea Bubble. *American Economic Review*, *94*(5), 1654–1668. doi:10.1257/0002828043052268

Twitter. (2011, March). #numbers (on Twitter's Blog). Retrieved from http://blog.twitter.com/2011/03/numbers.html

United Nations Statistics. (2013). Standard Country and Area Codes Classifications (M49). Retrieved February 17, 2015, from http://unstats.un.org/unsd/methods/m49/m49.htm

Valliere, D., & Peterson, R. (2004). Inflating the Bubble: examining dot-com investor behaviour. *Venture Capital*, *6*(1), 1–22.

Valliere, D., & Peterson, R. (2005). Venture Capitalist Behaviours: Frameworks for Future Research. *Venture Capital*, *7*(2), 167–183. Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/13691060500088076

Vascellaro, J. (2008). Gambling on Facebook Games, Zynga Cashes In. *The Wall Street Journal*. Retrieved from http://online.wsj.com/articles/SB121677837315476021

Wheale, P., & Amin, L. (2003). Bursting the dot.com "Bubble': A Case Study in Investor Behaviour. *Technology Analysis & Strategic Management*, *15*(1), 117–136. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/0953732032000046097

Witek, D., & Grettano, T. (2012). Information literacy on Facebook: an analysis. *Reference Services Review*, *40*(2), 242–257. doi:10.1108/00907321211228309

World Bank. (2015). Internet users (per 100 people) Data Table. Retrieved February 17, 2015, from http://data.worldbank.org/indicator/IT.NET.USER.P2?page=3

Appendix A List of Technology Companies

Following is the list of companies for which data was extracted from Datastream. The list is generated and curated by Datastream.

Name	*DS Mnemonic*		
3D Systems	U:DDD	Harris	U:HRS
Adobe Systems	@ADBE	Hewlett-Packard	U:HPQ
Akamai Techs.	@AKAM	IAC/Interactivecorp	@IACI
Altera	@ALTR	Informatica	@INFA
Amdocs	@DOX	Ingram Micro 'A'	U:IM
Analog Devices	@ADI	Intel	@INTC
Ansys	@ANSS	International Bus.Mchs.	U:IBM
AOL	U:AOL	Intuit	@INTU
Apple	@AAPL	Juniper Networks	U:JNPR
Applied Mats.	@AMAT	KLA Tencor	@KLAC
Arris Group	@ARRS	Lam Research	@LRCX
Aspen Technology	@AZPN	Leidos Holdings	U:LDOS
Athenahealth	@ATHN	Linear Tech.	@LLTC
Atmel	@ATML	LSI	@LSI
Autodesk	@ADSK	Marvell Tech.Group	@MRVL
Avago Technologies	@AVGO	Maxim Integrated Prds.	@MXIM
Broadcom 'A'	@BRCM	Microchip Tech.	@MCHP
Brocade Comms.Sys.	@BRCD	Micron Technology	@MU
CA	@CA	Micros Systems	@MCRS
Cadence Design Sys.	@CDNS	Microsoft	@MSFT
CDW	@CDW	Motorola Solutions	U:MSI
Cerner	@CERN	NCR	U:NCR
Check Point Sftw.Techs.	@CHKP	Netapp	@NTAP
Cisco Systems	@CSCO	Netsuite	U:N
Citrix Sys.	@CTXS	Nuance Comms.	@NUAN
Cognizant Tech.Sltn.'A'	@CTSH	nVidia	@NVDA
Commscope Holding Co.	@COMM	NXP Semiconductors	@NXPI
Commvault Systems	@CVLT	On Semicon.	@ONNN
Computer Scis.	U:CSC	Oracle	U:ORCL
Concur Techs.	@CNQR	Palo Alto Networks	U:PANW
Corning	U:GLW	Pitney-Bowes	U:PBI
Cree	@CREE	PTC	@PTC
DST Sys.	U:DST	Qualcomm	@QCOM
eBay	@EBAY	Rackspace Hosting	U:RAX
EMC	U:EMC	Red Hat	U:RHT
Equinix	@EQIX	Salesforce.Com	U:CRM
F5 Networks	@FFIV	Sandisk	@SNDK
Facebook Class A	@FB	Seagate Tech.	@STX
Fireeye	@FEYE	Servicenow	U:NOW
Freescale Semicon.	U:FSL	Sina	@SINA
Garmin	@GRMN	Skyworks Sltn.	@SWKS
Gartner 'A'	U:IT	Solera Holdings	U:SLH
Google 'A'	@GOOG	Splunk	@SPLK
Guidewire Software	U:GWRE	SS&C Technologies Hdg.	@SSNC
		Stratasys	@SSYS

Sunedison	U:SUNE
Symantec	@SYMC
Synopsys	@SNPS
Syntel	@SYNT
Teradata	U:TDC
Teradyne	U:TER
Texas Insts.	@TXN
Tibco Software	@TIBX
Twitter	U:TWTR
Tyler Techs.	U:TYL
Ubiquiti Networks	@UBNT
Ultimate Software GP.	@ULTI
Veeva Systems Cl.A	U:VEEV
Verisign	@VRSN
Vmware	U:VMW
Webmd	@WBMD
Western Digital	@WDC
Workday Class A	U:WDAY
Xilinx	@XLNX
Yahoo	@YHOO
Yandex	@YNDX

Appendix B SQL View Create Query

The following query was used to create a SQL View, of the combined data. This view was the principal input data for statistical analysis in SPSS.

```
CREATE VIEW spss AS

SELECT

    ar.id,
    ar.company_id,
    ar.year_id AS `year`,
    y.bin AS year_bin,
    ar.sales/1000 AS `sales/1000`,
    ar.price,
    ar.rnd,
    ar.volume_units,
    ar.volume_price,
    ar.mktcap,
    ar.marketshare,
    ar.numshares,
    LOG10(sales) AS log10_sales,
    LOG10(price) AS log10_price,
    LOG10(rnd) AS log10_rnd,
    LOG10(volume_units) AS log10_volume_units,
    LOG10(volume_price) AS log10_volume_price,
    LOG10(mktcap) AS log10_mktcap,
    LOG10(marketshare) AS log10_marketshare,
    LOG10(numshares) AS log10_numshares,
    ar.price-high,
    ar.price-low,
    ar.price-average,
    ar.unadjusted,
    ar.numshareschange,

    SUM(vc.value) AS vc_allphases_value,

    (SELECT SUM(vc.value)
        FROM vc
        WHERE vc.year_id = ar.year_id
            AND vc.vcphase_id = 1
    ) AS vc_phase1_value,

    (SELECT SUM(vc.value)
        FROM vc
        WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 2
    ) AS vc_phase2_value,

    (SELECT SUM(vc.value)
        FROM vc
        WHERE vc.year_id = ar.year_id
            AND vc.vcphase_id = 3
    ) AS vc_phase3_value,
```

```
(SELECT SUM(vc.value)
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 4
) AS vc_phase4_value,

LOG10(SUM(vc.value)) AS log10_vc_allphases_value,

(SELECT LOG10(SUM(vc.value))
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 1
) AS log10_vc_phase1_value,

(SELECT LOG10(SUM(vc.value))
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 2
) AS log10_vc_phase2_value,

(SELECT LOG10(SUM(vc.value))
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 3
) AS log10_vc_phase3_value,

(SELECT LOG10(SUM(vc.value))
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 4
) AS log10_vc_phase4_value,

SUM(vc.deals) AS vc_allphases_deals,

(SELECT SUM(vc.deals)
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 1
) AS vc_phase1_deals,

(SELECT SUM(vc.deals)
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 2
) AS vc_phase2_deals,

(SELECT SUM(vc.deals)
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 3
) AS vc_phase3_deals,

(SELECT SUM(vc.deals)
    FROM vc
    WHERE vc.year_id = ar.year_id
        AND vc.vcphase_id = 4
) AS vc_phase4_deals,
```

```
    (SELECT LOG10(SUM(p.online))
        FROM population p
        WHERE p.year_id = ar.year_id
            AND region_id = 241
    ) AS log10_onlinepopulation_us,

    (SELECT percentOnline
        FROM population p
        WHERE p.year_id = ar.year_id
            AND region_id = 241
    ) AS onlinepopulation_us_percent,

    (SELECT LOG10(SUM(p.online))
        FROM population p
        JOIN region r ON (p.region_id = r.id)
        WHERE p.year_id = ar.year_id
            AND r.oecd = 1
            AND r.incomeGroup_id = 1
    ) AS log10_onlinepopulation_oecd,

    (SELECT SUM(p.online)/SUM(p.`value`)*100
        FROM population p
        JOIN region r ON (p.region_id = r.id)
        WHERE p.year_id = ar.year_id
            AND r.oecd = 1
            AND r.incomeGroup_id = 1
    ) AS onlinepopulation_oecd_percent,

    (SELECT LOG10(SUM(p.online))
        FROM population p
        WHERE p.year_id = ar.year_id
    ) AS log10_onlinepopulation_world,

    (SELECT SUM(p.online)/SUM(p.`value`)*100
        FROM population p
        WHERE p.year_id = ar.year_id
    ) AS onlinepopulation_world_percent

FROM annualReport ar

JOIN vc ON (vc.year_id = ar.year_id)

JOIN `year` y ON (y.id = ar.year_id)

WHERE sales IS NOT NULL
    AND rnd IS NOT NULL

GROUP BY ar.year_id,
    ar.company_id;
```

Appendix C Sample of Dataset Contents

The following table shows the number of cases present in the dataset for each bin period. The dataset contains 1205 cases in total.

Bin	1	2	3	4	5	6	7
Cases	203	60	126	63	351	77	325

The following table contains an excerpt of the dataset, with a sample row from each year.

It has been transposed so that it may fit on the page. As such, the italicised headings represent columns in the dataset.

year	bin	sales	price	rnd	volume_units	volume_price	mktcap	marketshare	numshares	vc_phase1_value	vc_phase2_value	vc_phase3_value	vc_phase4_value	vc_phase1_deals	vc_phase2_deals	vc_phase3_deals	vc_phase4_deals	onlinepopulation_us	onlinepopulation_oecd	onlinepopulation_world
1995	1	4.8	0.83	3.79	2.95	3.67	5.36	-3.52	4.05	8.3	8.55	8.16	8.69	107	135	50	144	7.39	7.57	8.36
1996	1	5.9	0.73	5.19	5.43	6.1	6.5	-2.54	4.86	8.57	8.82	8.47	9.01	140	233	79	236	7.65	7.83	8.62
1997	1	5.8	1.11	4.74	5.52	6.6	6.66	-2.68	4.95	8.51	8.87	8.76	9.26	138	229	99	360	7.77	8.03	8.83
1998	1	6.59	1.18	5.26	4.14	5.38	6.48	-1.92	4.9	8.6	9.07	8.93	9.37	162	283	119	422	7.92	8.21	9.02
1999	2	6.72	1.81	5.63	4.6	6.3	7.55	-1.84	5.73	8.8	9.38	9.37	9.72	194	431	177	616	8	8.37	9.21
2000	3	5.43	1.4	4.71	3.75	5.3	5.87	-3.22	4.47	8.91	9.78	9.63	9.93	195	759	226	986	8.08	8.5	9.36
2001	3	5.92	1.35	5.23	5.14	6.55	6.94	-2.68	5.59	8.3	9.29	9.39	9.79	80	346	180	696	8.15	8.58	9.46
2002	4	5.16	0.33	4.33	4.58	4.96	5.4	-3.4	5.07	7.82	9.05	9.12	9.48	57	265	183	500	8.23	8.66	9.58
2003	5	5.21	1.13	4.11	4.37	5.01	6.21	-3.4	5.08	8.02	9.03	9.1	9.4	59	278	176	451	8.25	8.7	9.65
2004	5	6.9	1.25	6	5.73	7.01	7.47	-1.8	6.23	7.87	9.01	9.33	9.36	45	262	263	380	8.28	8.75	9.72
2005	5	5.45	1.32	4.26	4.8	5.96	6.51	-3.3	5.18	7.98	8.93	9.38	9.28	63	237	316	335	8.3	8.79	9.77
2006	5	5.42	1.37	4.69	3.84	5.21	6.25	-3.4	4.59	8.35	8.97	9.36	9.31	70	233	331	392	8.31	8.81	9.81
2007	5	6	1.02	4.85	4.58	5.73	6.06	-2.85	5.04	8.46	9.08	9.44	9.29	98	277	382	319	8.35	8.85	9.89
2008	6	6.91	0.95	6.04	5.47	6.71	7.07	-1.98	6.12	8.5	8.99	9.39	9.36	119	273	381	340	8.35	8.86	9.94
2009	7	5.05	0.59	4.05	3.53	3.94	5.42	-3.7	4.35	8.37	9.05	9.14	9.14	52	280	227	272	8.34	8.87	9.99
2010	7	6.58	1.45	5.83	5.01	6.5	7.16	-2.36	5.71	8.36	9.15	9.18	9.33	81	369	251	357	8.35	8.88	10.04
2011	7	6.62	1.43	5.87	5.18	6.67	7.12	-2.37	5.69	8.37	9.32	9.32	9.53	127	500	261	356	8.34	8.89	10.08
2012	7	6.64	1.54	5.87	5.03	6.54	7.23	-2.39	5.69	8.28	9.34	9.37	9.58	81	645	253	390	8.4	8.92	10.13

Appendix D Factor Analyses Correlation Matrixes

The following pages show correlation matrixes for the binned factor analyses.

Correlation Matrix for Factor Analysis of 1995-1998 (Bin 1)

	sales	price	rnd	volume_units	volume_price	mktcap	marketshare	numshares	vc_phase1_value	vc_phase2_value	vc_phase3_value	vc_phase4_value	vc_phase1_deals	vc_phase2_deals	vc_phase3_deals	vc_phase4_deals	onlinepopulation_us	onlinepopulation_oecd	onlinepopulation_world
sales	1.000	0.525	0.949	0.634	0.806	0.882	0.997	0.893	0.001	0.025	0.029	0.025	0.022	0.018	0.030	0.033	0.027	0.030	0.031
price	0.525	1.000	0.445	0.018	0.452	0.591	0.518	0.497	0.045	0.078	0.088	0.085	0.071	0.068	0.087	0.092	0.084	0.088	0.089
rnd	0.949	0.445	1.000	0.688	0.821	0.864	0.944	0.884	0.010	0.032	0.039	0.036	0.028	0.025	0.038	0.042	0.035	0.039	0.039
volume_units	0.634	0.018	0.688	1.000	0.893	0.720	0.630	0.715	0.056	0.045	0.026	0.025	0.052	0.051	0.031	0.021	0.035	0.030	0.030
volume_price	0.806	0.452	0.821	0.893	1.000	0.886	0.798	0.859	0.085	0.090	0.077	0.075	0.093	0.091	0.082	0.074	0.083	0.081	0.081
mktcap	0.882	0.591	0.864	0.720	0.886	1.000	0.881	0.948	-0.021	-0.005	0.002	0.000	-0.008	-0.011	0.001	0.005	-0.002	0.002	0.002
marketshare	0.997	0.518	0.944	0.630	0.798	0.881	1.000	0.888	-0.061	-0.048	-0.046	-0.050	-0.048	-0.053	-0.045	-0.042	-0.048	-0.045	-0.044
numshares	0.893	0.497	0.884	0.715	0.859	0.948	0.888	1.000	0.035	0.063	0.070	0.066	0.057	0.054	0.070	0.073	0.066	0.070	0.071
vc_phase1_value	0.001	0.045	0.010	0.056	0.085	-0.021	-0.061	0.035	1.000	0.911	0.813	0.830	0.938	0.962	0.836	0.763	0.872	0.830	0.822
vc_phase2_value	0.025	0.078	0.032	0.045	0.090	-0.005	-0.048	0.063	0.911	1.000	0.963	0.955	0.992	0.988	0.980	0.943	0.989	0.977	0.976
vc_phase3_value	0.029	0.088	0.039	0.026	0.077	0.002	-0.046	0.070	0.813	0.963	1.000	0.997	0.921	0.923	0.997	0.997	0.992	0.998	0.997
vc_phase4_value	0.025	0.085	0.036	0.025	0.075	0.000	-0.050	0.066	0.830	0.955	0.997	1.000	0.912	0.922	0.989	0.990	0.988	0.991	0.989
vc_phase1_deals	0.022	0.071	0.028	0.052	0.093	-0.008	-0.048	0.057	0.938	0.992	0.921	0.912	1.000	0.995	0.947	0.893	0.962	0.942	0.940
vc_phase2_deals	0.018	0.068	0.025	0.051	0.091	-0.011	-0.053	0.054	0.962	0.988	0.923	0.922	0.995	1.000	0.944	0.891	0.963	0.940	0.936
vc_phase3_deals	0.030	0.087	0.038	0.031	0.082	0.001	-0.045	0.070	0.836	0.980	0.997	0.989	0.947	0.944	1.000	0.990	0.998	1.000	1.000
vc_phase4_deals	0.033	0.092	0.042	0.021	0.074	0.005	-0.042	0.073	0.763	0.943	0.997	0.990	0.893	0.891	0.990	1.000	0.980	0.992	0.993
onlinepopulation_us	0.027	0.084	0.035	0.035	0.083	-0.002	-0.048	0.066	0.872	0.989	0.992	0.988	0.962	0.963	0.998	0.980	1.000	0.997	0.995
onlinepopulation_oecd	0.030	0.088	0.039	0.030	0.081	0.002	-0.045	0.070	0.830	0.977	0.998	0.991	0.942	0.940	1.000	0.992	0.997	1.000	1.000
onlinepopulation_world	0.031	0.089	0.039	0.030	0.081	0.002	-0.044	0.071	0.822	0.976	0.997	0.989	0.940	0.936	1.000	0.993	0.995	1.000	1.000

Correlation Matrix for Factor Analysis of 2000-2001 (Bin 3)

	sales	price	rnd	volume_units	volume_price	mktcap	marketshare	numshares	vc_phase1_value	vc_phase2_value	vc_phase3_value	vc_phase4_value	vc_phase1_deals	vc_phase2_deals	vc_phase3_deals	vc_phase4_deals	onlinepopulation_us	onlinepopulation_oecd	onlinepopulation_world
sales	1.000	0.543	0.916	0.627	0.699	0.830	0.999	0.837	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	-0.003	-0.003	-0.003
price	0.543	1.000	0.477	0.335	0.682	0.776	0.539	0.523	0.181	0.181	0.181	0.181	0.181	0.181	0.181	0.181	-0.181	-0.181	-0.181
rnd	0.916	0.477	1.000	0.698	0.726	0.825	0.917	0.883	-0.047	-0.047	-0.047	-0.047	-0.047	-0.047	-0.047	-0.047	0.047	0.047	0.047
volume_units	0.627	0.335	0.698	1.000	0.889	0.747	0.628	0.758	-0.038	-0.038	-0.038	-0.038	-0.038	-0.038	-0.038	-0.038	0.038	0.038	0.038
volume_price	0.699	0.682	0.726	0.889	1.000	0.881	0.696	0.807	0.140	0.140	0.140	0.140	0.140	0.140	0.140	0.140	-0.140	-0.140	-0.140
mktcap	0.830	0.776	0.825	0.747	0.881	1.000	0.828	0.893	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085	-0.085	-0.085	-0.085
marketshare	0.999	0.539	0.917	0.628	0.696	0.828	1.000	0.836	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015	0.015	0.015	0.015
numshares	0.837	0.523	0.883	0.758	0.807	0.893	0.836	1.000	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	0.020	0.020	0.020
vc_phase1_value	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase2_value	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase3_value	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase4_value	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase1_deals	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase2_deals	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase3_deals	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
vc_phase4_deals	0.003	0.181	-0.047	-0.038	0.140	0.085	-0.015	-0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
onlinepopulation_us	-0.003	-0.181	0.047	0.038	-0.140	-0.085	0.015	0.020	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
onlinepopulation_oecd	-0.003	-0.181	0.047	0.038	-0.140	-0.085	0.015	0.020	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
onlinepopulation_world	-0.003	-0.181	0.047	0.038	-0.140	-0.085	0.015	0.020	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000

Correlation Matrix for Factor Analysis of 2003-2007 (Bin 5)

	sales	price	rnd	volume_units	volume_price	mktcap	marketshare	numshares	vc_phase1_value	vc_phase2_value	vc_phase3_value	vc_phase4_value	vc_phase1_deals	vc_phase2_deals	vc_phase3_deals	vc_phase4_deals	onlinepopulation_us	onlinepopulation_oecd	onlinepopulation_world
sales	1.000	0.461	0.923	0.662	0.748	0.873	0.994	0.827	0.121	0.030	0.129	-0.120	0.117	-0.034	0.142	-0.112	0.143	0.145	0.145
price	0.461	1.000	0.367	0.215	0.594	0.619	0.444	0.242	0.173	0.026	0.177	-0.180	0.173	-0.057	0.201	-0.160	0.203	0.205	0.205
rnd	0.923	0.367	1.000	0.735	0.770	0.865	0.917	0.879	0.122	0.036	0.103	-0.096	0.112	-0.024	0.121	-0.084	0.126	0.127	0.129
volume_units	0.662	0.215	0.735	1.000	0.905	0.770	0.667	0.789	-0.018	-0.022	0.036	-0.033	-0.008	-0.023	0.025	-0.040	0.015	0.018	0.014
volume_price	0.748	0.594	0.770	0.905	1.000	0.890	0.738	0.757	0.097	0.005	0.162	-0.148	0.100	-0.059	0.162	-0.147	0.152	0.158	0.154
mktcap	0.873	0.619	0.865	0.770	0.890	1.000	0.869	0.870	0.092	0.001	0.098	-0.102	0.087	-0.046	0.110	-0.085	0.108	0.111	0.110
marketshare	0.994	0.444	0.917	0.667	0.738	0.869	1.000	0.826	0.033	0.008	0.031	-0.028	0.029	-0.009	0.035	-0.025	0.035	0.036	0.036
numshares	0.827	0.242	0.879	0.789	0.757	0.870	0.826	1.000	0.054	-0.002	0.059	-0.060	0.049	-0.031	0.065	-0.049	0.063	0.066	0.065
vc_phase1_value	0.121	0.173	0.122	-0.018	0.097	0.092	0.033	0.054	1.000	0.421	0.464	-0.477	0.903	0.032	0.692	-0.349	0.804	0.776	0.825
vc_phase2_value	0.030	0.026	0.036	-0.022	0.005	0.001	0.008	-0.002	0.421	1.000	-0.022	0.262	0.516	0.885	0.083	-0.048	0.286	0.155	0.249
vc_phase3_value	0.129	0.177	0.103	0.036	0.162	0.098	0.031	0.059	0.464	-0.022	1.000	-0.894	0.508	-0.378	0.956	-0.922	0.871	0.912	0.878
vc_phase4_value	-0.120	-0.180	-0.096	-0.033	-0.148	-0.102	-0.028	-0.060	-0.477	0.262	-0.894	1.000	-0.564	0.535	-0.910	0.887	-0.831	-0.879	-0.825
vc_phase1_deals	0.117	0.173	0.112	-0.008	0.100	0.087	0.029	0.049	0.903	0.516	0.508	-0.564	1.000	0.221	0.719	-0.557	0.863	0.794	0.835
vc_phase2_deals	-0.034	-0.057	-0.024	-0.023	-0.059	-0.046	-0.009	-0.031	0.032	0.885	-0.378	0.535	0.221	1.000	-0.313	0.186	-0.105	-0.252	-0.169
vc_phase3_deals	0.142	0.201	0.121	0.025	0.162	0.110	0.035	0.065	0.692	0.083	0.956	-0.910	0.719	-0.313	1.000	-0.879	0.969	0.992	0.975
vc_phase4_deals	-0.112	-0.160	-0.084	-0.040	-0.147	-0.085	-0.025	-0.049	-0.349	-0.048	-0.922	0.887	-0.557	0.186	-0.879	1.000	-0.834	-0.836	-0.797
onlinepopulation_us	0.143	0.203	0.126	0.015	0.152	0.108	0.035	0.063	0.804	0.286	0.871	-0.831	0.863	-0.105	0.969	-0.834	1.000	0.988	0.994
onlinepopulation_oecd	0.145	0.205	0.127	0.018	0.158	0.111	0.036	0.066	0.776	0.155	0.912	-0.879	0.794	-0.252	0.992	-0.836	0.988	1.000	0.994
onlinepopulation_world	0.145	0.205	0.129	0.014	0.154	0.110	0.036	0.065	0.825	0.249	0.878	-0.825	0.835	-0.169	0.975	-0.797	0.994	0.994	1.000

Correlation Matrix for Factor Analysis of 2009-2012 (Bin 7)

	sales	price	rnd	volume_units	volume_price	mktcap	marketshare	numshares	vc_phase1_value	vc_phase2_value	vc_phase3_value	vc_phase4_value	vc_phase1_deals	vc_phase2_deals	vc_phase3_deals	vc_phase4_deals	onlinepopulation_us	onlinepopulation_oecd	onlinepopulation_world
sales	1.000	0.247	0.879	0.748	0.803	0.855	0.995	0.825	-0.067	0.097	0.096	0.099	0.064	0.096	0.084	0.092	0.066	0.090	0.098
price	0.247	1.000	0.138	-0.038	0.464	0.490	0.235	-0.115	-0.066	0.106	0.096	0.114	0.084	0.100	0.123	0.130	0.065	0.099	0.114
rnd	0.879	0.138	1.000	0.802	0.789	0.840	0.874	0.860	-0.072	0.085	0.088	0.085	0.043	0.089	0.061	0.075	0.071	0.085	0.087
volume_units	0.748	-0.038	0.802	1.000	0.860	0.739	0.757	0.845	0.018	-0.046	-0.040	-0.050	-0.046	-0.040	-0.061	-0.058	-0.018	-0.037	-0.048
volume_price	0.803	0.464	0.789	0.860	1.000	0.895	0.798	0.692	-0.055	0.090	0.089	0.089	0.063	0.086	0.075	0.077	0.054	0.077	0.087
mktcap	0.855	0.490	0.840	0.739	0.895	1.000	0.851	0.784	-0.035	0.055	0.050	0.060	0.044	0.053	0.067	0.071	0.035	0.053	0.061
marketshare	0.995	0.235	0.874	0.757	0.798	0.851	1.000	0.829	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.001	0.000	0.000	0.000
numshares	0.825	-0.115	0.860	0.845	0.692	0.784	0.829	1.000	0.006	-0.010	-0.010	-0.010	-0.008	-0.010	-0.010	-0.009	-0.006	-0.009	-0.010
vc_phase1_value	-0.067	-0.066	-0.072	0.018	-0.055	-0.035	0.000	0.006	1.000	-0.587	-0.700	-0.609	0.121	-0.808	-0.244	-0.657	-1.000	-0.926	-0.759
vc_phase2_value	0.097	0.106	0.085	-0.046	0.090	0.055	0.000	-0.010	-0.587	1.000	0.983	0.994	0.726	0.951	0.850	0.855	0.574	0.837	0.961
vc_phase3_value	0.096	0.096	0.088	-0.040	0.089	0.050	0.000	-0.010	-0.700	0.983	1.000	0.971	0.599	0.984	0.746	0.822	0.688	0.895	0.967
vc_phase4_value	0.099	0.114	0.085	-0.050	0.089	0.060	0.000	-0.010	-0.609	0.994	0.971	1.000	0.714	0.953	0.880	0.908	0.598	0.861	0.977
vc_phase1_deals	0.064	0.084	0.043	-0.046	0.063	0.044	0.001	-0.008	0.121	0.726	0.599	0.714	1.000	0.479	0.884	0.554	-0.135	0.259	0.552
vc_phase2_deals	0.096	0.100	0.089	-0.040	0.086	0.053	0.000	-0.010	-0.808	0.951	0.984	0.953	0.479	1.000	0.702	0.863	0.799	0.960	0.985
vc_phase3_deals	0.084	0.123	0.061	-0.061	0.075	0.067	0.001	-0.010	-0.244	0.850	0.746	0.880	0.884	0.702	1.000	0.871	0.235	0.587	0.799
vc_phase4_deals	0.092	0.130	0.075	-0.058	0.077	0.071	0.001	-0.009	-0.657	0.855	0.822	0.908	0.554	0.863	0.871	1.000	0.653	0.869	0.934
onlinepopulation_us	0.066	0.065	0.071	-0.018	0.054	0.035	0.000	-0.006	-1.000	0.574	0.688	0.598	-0.135	0.799	0.235	0.653	1.000	0.921	0.750
onlinepopulation_oecd	1.000	0.247	0.879	0.748	0.803	0.855	0.995	0.825	-0.067	0.097	0.096	0.099	0.064	0.096	0.084	0.092	0.066	0.090	0.098
onlinepopulation_world	0.247	1.000	0.138	-0.038	0.464	0.490	0.235	-0.115	-0.066	0.106	0.096	0.114	0.084	0.100	0.123	0.130	0.065	0.099	0.114