Predicting Video Game Sales in the European Market

Walter Steven Beaujon

Research Paper Business Analytics

vrije Universiteit   amsterdam

Predicting Video Game Sales in the European Market

Walter Steven Beaujon

Research Paper Business Analytics

Faculty of Sciences
VU University Amsterdam
De Boelelaan 1081
1081 HV Amsterdam
Nederland

March 2012

Preface

This report is the BMI-paper for the final year of the BMI-master. It is a report that contains a real-world problem and a solution that is found using business-knowledge as well as mathematics or informatics.

The main purpose of this report is to come up with an effective way to predict the sales of video games in the European market. These predictions will be for software sales, as well as hardware sales and will be based on actual European sales figures from the past 7 years.

It has been my pleasure to work with Rob van der Mei during the creation of this report. He was very helpful during the planning stages and has given essential advice that shaped what the actual subject would be about. During the writing stage of the report, he was instrumental in giving feedback and coming up with new ideas to implement in order to further the report's progress.

Summary

The video game industry is one of the fastest growing industries in the world. It is expected to be worth $68 billion in 2012. With hundreds of games being released each year, where some can sell upwards of 5 million units and others barely manage 10 thousand units, it is very important to know what factors will lead to a successful game. This report provides a thorough analysis of the European game industry, using a database of thousands of games and their European sales figures, that serves as the foundation for predictions that can be made about future sales. In it you will find an examination of the hardware sales from 2005 to 2011, which will lead to predictions about the expected market share for each of the console manufacturers in 2012 and on. The section on monthly sales will show how sales are distributed in a year, and how this is expected to be in the future. The individual games in the database will be supported by 8 different fields of information that each contribute to the game's success or failure. These fields will be used as the basis for the predictions of a game's total sales. The weekly sales decline for individual games will also be observed in order to draw conclusions about the expected continued decline. These predictions can be made based on the first few weeks of actual sales, and there will be a section that shows the change in the reliability of using fewer or more points to make these predictions. This report includes a program that can be used by potential game creators to provide sales predictions in different stages of the game's development. It can be used to predict the coming weeks of sales for a game using only the first few weeks of sales data. It can also make predictions about the expected total sales and the allows game makers to compare different scenarios in which they could release their game and compare the impact these decisions would have on their sales.

Table of Contents

1. Introduction

The video game industry is the economic sector involved with the worldwide development, marketing and sale of video games. In 2007, the game industry earned about $41.9 billion in revenue. This number is expected to grow with 9.1% annually to $48.9 billion in 2011 and $68 billion in 2012, making it the fastest-growing component of the international media sector. In the US, the game industry is already worth more than the Music and Movie[1] industries combined. The UK industry has surpassed the Music industry as early as 2008, and is assumed to have already passed DVD sales[2]. Interactive software sales in nine of the major European markets[3] totaled approximately €7.3 billion in 2007. Software sales represent the games alone and not the actual hardware. The largest European markets are the UK, which earned €2.3 billion in software revenue, followed by France (€1.6B), Germany (€1.4B), Spain (€0.7B) and Italy (€0.6B)[1].

Despite how successful the industry has become, it is not guaranteed that all games will be a commercial success. Certain games are plagued by such high development costs that they have to sell over a million units just to break even. Oftentimes developers will have deadlines or budgetary constraints that they need to meet, which lead to many games being released before they are completely finished, which could greatly affect sales. This report will explore the possibility of predicting how well a new game will sell. These predictions will be entirely based on historical data that has been collected from different sources. The data collected will be of the European market and therefore the predictions will only be meaningful in this region, as broadening the scope will require a thorough analysis of cultural differences.

The report will start with a chapter on the collection of the data along with the meaning behind the data, followed by a chapter about initial observations and the next chapter will be about the actual analysis. The last chapter will contain the conclusion and the appendices will have a user guide for the actual prediction program as well as charts, graphs and the literature used.

---

[1] Box office only, not counting DVD sales & rentals.
[2] Total revenue from 2011 is still pending.
[3] UK, France, Germany, Spain, Italy, Netherlands, Switzerland, Sweden & Finland.

2. Data preparation

This chapter contains the methods behind the collection of data, an explanation of the different fields and the ways in which the data was manipulated. Lastly, the chapter lists the obstacles faced in this stage.

2.1. Data collection

This section outlines the websites where the data was collected from.

2.1.1. Sales data

The website vgchartz.com[2] has up-to-date weekly sales figures for all game consoles, individual games as well as total software sales for every major region. Through this site I was able to collect sales figures for Europe[4] from March 12, 2005[5] to December 31, 2011. The console sales consisted of the name of the console with its weekly hardware sales figures. The software sales also showed the consoles and its weekly software sales figures. The individual game sales listed the 40 highest selling games per week by their name, console, publisher, genre, current sales and total sales as well as the number of weeks since its release. Lastly the site also contains the total sales for all 2450 games released in Europe in the past seven years.

This data was missing some fields that I felt were relevant, such as developer, some measure of quality, whether it was a sequel and whether it was licensed. These were collected from different sites.

2.1.2. Developer data

Developer data was collected from Wikipedia.org, where each console had large tables containing the games and their developers.

2.1.3. MetaScore data

Metacritic.com[3] is a website that uses a weighted average to distill many video game critics' voices into one single MetaScore. MetaScores range from 0-100, with higher scores indicating better overall reviews. Using Metacritic I was able find the MetaScores for most of the games found on vgchartz. This was the primary indication of quality for individual games.

---

[4] vgchartz.com has accurate sales figures for the UK, France & Germany, which make up over 70% of the European game market. The complete European sales figures are extrapolated from these figures.
[5] 12/03/2005 was the release of the Nintendo DS, which is the oldest console that still consistently sells units today.

2.2. Data definitions

This section contains definitions of all the fields found in the data.

2.2.1. Game title

This is simply the title of the game, many games are released on multiple platforms, but for the sake of this report these titles will be treated as separate entries. Game title is not used in any form of the analysis and serves only to find whether the game is a sequel or licensed. Sequels can be found by whether the title has a number or year behind it (i.e. The Sims 3, Pro Evolution Soccer 2011). Licensed property is identified by a well-known (non-game) property in the title (i.e. Spider-Man: Edge of Time, FIFA soccer 2010).

2.2.2. Game console

Game console refers to the hardware the game must be played on. The seven consoles are the Nintendo DS, Nintendo Wii, Nintendo 3DS, Sony PlayStation 3, Sony PSP, Microsoft XBOX 360 and PC. Earlier consoles were not taken into account, as their sales figures are more difficult to come by.

2.2.3. Publisher

The publisher is the company that green-lights a new game, assigns a developer to make it, creates a budget to market it and ultimately publishes it under their name. Which publisher is attached to a title has a large influence on its success or failure.

2.2.4. Developer

The developer is the studio that actually makes the game, made up of artists and programmers. When a studio is well-known for making great games, it can often influence many people to buy their games even if the game is a new IP.

2.2.5. Genre

This shows what genre the game belongs to. Some of the most popular genres in gaming are: Action, FPS (First Person Shooters), Racer and Sports.

2.2.6. Release date

The release date shows the day the product was released. The release date has a lot of influence on how well it sells. Game sales often spike around the holidays (November and December) and competition is very high around this period. Other games might need to be released around a certain event to maximize sales, such as a World Cup game, or a tie-in game for a movie.

2.2.7. Sequel

This binary value shows whether or not a game is a sequel or a new IP. Sequels represent a much smaller risk to publishers, since the property is already known to gamers and can be counted on to sell similarly to their predecessors.

2.2.8. License

This value shows whether or not a game is attached to a (non-gaming) license. Licensed games can count on their popularity in other media to attract buyers to pick up this title even if they don't play many games overall. They also serve as a way for parents to buy games for their children, based on the knowledge that their children enjoy that IP. Licensed games about new movies are often released around the same time as the movies and serve as quick way to earn money, while simultaneously marketing the movie to a wider audience.

There are four categories in the data:
1. *Unlicensed*: These are games that do not have any significant non-gaming licenses attached to them.
2. *Promotional license*: This is any game that uses a license from a movie , TV show, toy line, or any other brand to attract potential buyers.
3. *Sports License*: This license is generally attached to sports games and allows the games to use real athletes, real teams, or real brands.
4. *Music License*: Games that require a music license are games where music plays a very important role in the gameplay, and different songs need to be licensed in order to be included in the game. The main difference between this license and the other ones is that here the license owners don't pursue developers to make a game about their intellectual property.

2.2.9. MetaScore

The MetaScore is the weighted average score (from 0 to 100) for a game based on the reviews of a wide range of critics. MetaScore is often used as a marketing tool as high scores can greatly influence people to buy this game and the score reflects very well on the publisher, developer and IP.

2.2.10. Total sales

Total sales are the measure of a games success and will be the measurement used in the predictions.

2.3.   Data manipulation

The first thing I did was combine all the game data from the three different sources. This was done by comparing the titles and consoles from one source to that of a different source. Some slight changes to the titles were made in order to successfully combine them. All the sequel and license data was filled in by hand.

Since the weekly sales data included only the 40 best-selling titles a way was devised to increase this number. All titles included their weekly sales as well as their total sales up to that date. In some cases titles would be missing from the charts for a single week at a time. Using the total sales from before the missing week and after the missing week, I was able to calculate how many units were sold during the missing week. This method was also used when the title first charted during its second week of release. When a title is missing for more than one week at a time, I chose to not make any assumptions about the missing weeks, since I was not yet ready to make estimations on sales depreciation.

2.4.   Limitations

One of the first limitations I was faced with was the fact that after combining the data from all three sources there were still several titles without a MetaScore, developer or exact release date. These missing values could negatively impact the analysis and therefore the predictions. Titles that do not have MetaScores could not have their scores replaced with a score from another score aggregating website, since these sites could use completely different algorithms. Finding the developer and European release date for titles where this was missing proved to be an arduous task as each of these values would have to be individually searched for. Furthermore, the data does not include values for age ratings, marketing, pre-orders, included online game modes, pricing, budget or whether the game was also available for download. Finding these values for 2450 games would take weeks, if not months.
Age ratings are systems used to label games by age according to their content. In Europe this is done by the British Board of Film Classification (BBFC)[7] and the Pan European Game Information (PEGI)[8]. A lower age rating means the game is available for a wider audience.
The type of marketing campaign used before the launch of a game can also greatly influence the total sales of said game.
Pre-order are made when consumers reserve the right to purchase the game upon release. The amount of pre-orders a game has before release can also gauge its success.
If the game has a compelling online experience it attracts more customers and users will be less likely to resell their games (used game sales are not measured in the total sales).
The actual price the game goes on sale for can also greatly influence consumers, as does the game's development budget.

Lastly, many games are also available for download on their respective consoles. Since downloads are not counted in the total sales, downloadable availability might lead a drop in retail sales.

The second limitation comes in the form of the weekly sales being limited by the 40 best titles. As game sales tend to depreciate over time, it is unfortunate that the weekly sales would not show how games would sell in a later stage of its lifecycle. As the total weekly software sales per console were also known, I was able to calculate how many games were sold below the top 40 mark. For the final week of the data I found that, given the sales figure for the fortieth game and the number of remaining sales per console, there had to be at least 175 other games that were counted in the software totals, but could not make it to the top 40. This staggering figure could greatly affect the accuracy of my predictions for weekly sales. For this reason, I have decided to focus the predictions on total sales per game, rather than weekly sales per game, as this data was up-to-date for all 2450 games released in Europe before 2012.

3. Primary observations

This chapter contains observations of the data that were found before continuing with an in-depth analysis.

3.1. Holidays

One of the first observations is the fact that games tend to sell much better during the holidays (November and December). Over the past six years 38.4% of all hardware sales and 43.1% of all software sales occurred during these months. Figures 1 and 2 show how the sales reach their highest point right before Christmas. Figure 3 shows that despite having different heights, the sales within a year tend to have the same shape.
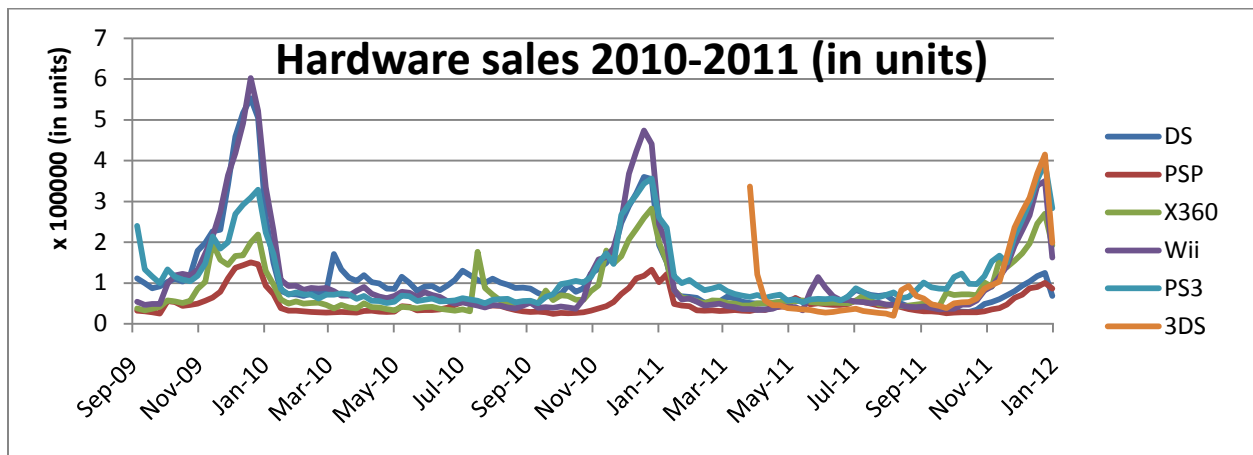
Figure 1: The total *hardware* sales per console in units for the last 28 months in the data.
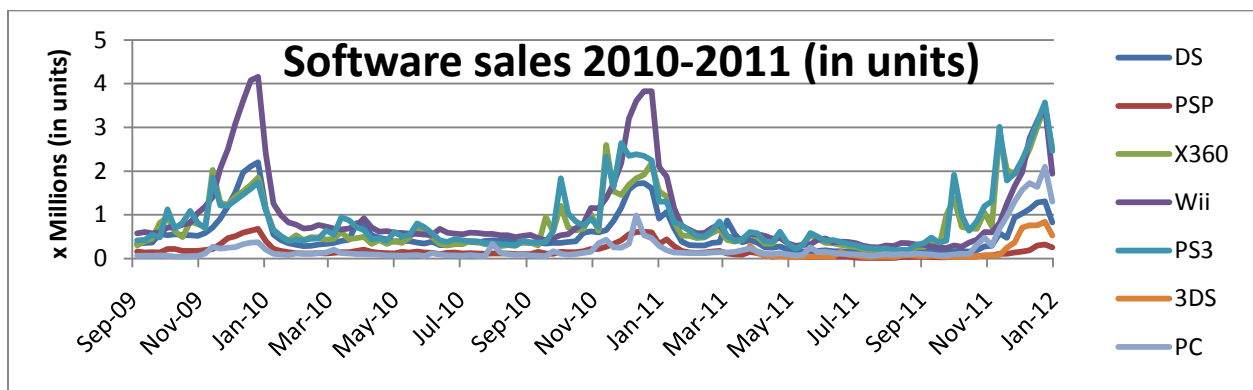
Figure 2: The total *software* sales per console in units for the last 28 months in the data.
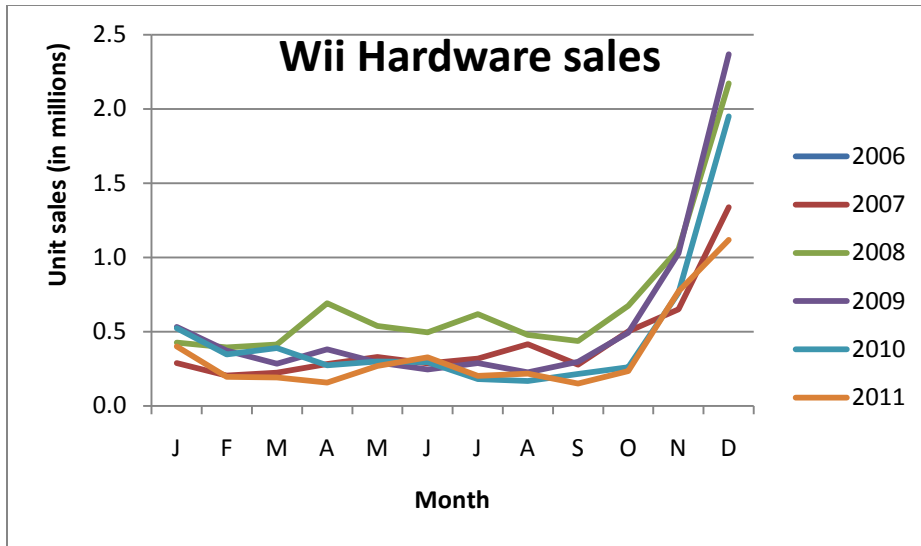
Figure 3: The Wii's hardware sales in units for every year to date.

3.2. Consoles

In order to understand the sales, it is important to understand which consoles are the most popular. These six consoles are separated into two categories: Home consoles and handheld consoles.

The three home consoles are the Xbox 360, Wii and PS3. Table 1 shows how of these three consoles, the Xbox 360 and PS3's sales growing annually, while the Wii's sales started dropping after 2008 despite experiencing very high sales in its first 3 years.

The handheld consoles are the DS, PSP and 3DS. Table 1 shows how the DS experienced the most sales, but also suffered from the greatest year over year loss for 3 years running. The PSP has sold substantially less over its lifetime, compared to the DS, and has also peaked around 2007-2008, but has managed to keep its sales steady for three years. The drop in popularity of the handheld consoles can be attributed to the rise of the smartphone and the introduction of the tablet. These new devices tend to be able to do much more than standalone game consoles and their games cost significantly less than those of the handheld consoles. Both the DS and PSP are expected to play a much smaller role in the coming years, considering that their successors have been introduced in the market. The 3DS released (in Europe) on the 25th of March 2011 and experienced a reasonably good first week of sales. By the third week, however, its sales had dropped below all other consoles and this continued for several months (as shown in Figure 1). On the 12th of August 2011, less than 5 months after its European launch, Nintendo dropped the price of the 3DS by around a third[6], which led to a holiday season so successful that it still managed to outsell both the DS and PSP. The PSP's successor, the PlayStation Vita, is scheduled to release in Europe at the end of February. As of writing it has sold 520,000 units in Japan after 6 weeks on the market. The PSP and 3DS sold 590,000 and 840,000 units in Japan, respectively, during their first 6 weeks.

Hardware	DS	PSP	X360	Wii	PS3	3DS
2005	3,514,602	2,621,803	443,074			
2006	6,014,943	3,312,992	1,841,797	697,443		
2007	10,702,518	4,134,693	2,026,633	5,111,302	3,418,547	
2008	11,357,824	3,364,467	3,755,689	8,098,919	3,995,364	
2009	9,557,223	2,654,051	3,440,803	6,753,269	4,753,758	
2010	6,443,450	2,257,076	4,019,345	5,746,361	5,280,052	
2011	3,379,576	2,509,782	4,336,196	4,473,619	6,115,449	3,984,661
Total	50,970,136	20,854,864	19,863,537	30,880,913	23,563,170	3,984,661

Table 1: The yearly *hardware* sales per console in units.

3.3. Year over year growth

In Table 1 we saw that both the Xbox 360 and the PS3 have experienced a year over year growth in console sales, while the DS, PSP and Wii have been in decline for the past 3 years. It was expected that software sales would always reflect the amount of total hardware units sold in a console's lifetime, but as Table 2 shows, this is not entirely the case. Despite the DS having much higher total sales than the other consoles, it still sold about half as many software units as the home consoles in 2011. This implies that year over year software sales are much more closely related to year over year hardware sales than total hardware sales. This holds true for all the consoles that experienced growth as well as those that have been in decline.

Software	DS	PSP	X360	Wii	PS3	3DS
2005	3,880,100	4,242,571				
2006	10,672,193	5,781,735	2,964,249	1,501,925		
2007	26,777,040	9,867,695	12,149,509	18,121,414	13,841,548	
2008	42,280,287	12,776,952	29,729,671	53,791,796	31,932,432	
2009	32,360,431	9,401,744	30,116,176	53,014,451	30,940,980	
2010	28,778,085	9,950,989	36,045,545	51,567,373	42,187,805	
2011	20,569,885	5,402,655	43,306,801	39,922,287	46,252,173	6,448,161
Total	165,318,021	57,424,341	154,311,951	217,919,246	165,154,938	6,448,161

Table 2: The yearly total *software* sales per console in units.

3.4. Attach rates

Attach rate refers to the average amount of games each console owner buys. This is calculated by dividing total software sales with total hardware sales. Figure 4 shows the attach rates for all 6 consoles. You'll notice that the handheld consoles have an attach rate of around 3 after being on the market for over 6 years. There are three major reasons for these figures: first, a great deal of PSP and DS owners will only get one or two games, second, handheld owners might be buying several units of their favorite handheld, and third, downloaded games don't count towards the software totals. The first option is very likely since both handhelds have, in the past, been marketed heavily towards a non-gaming

demographic using very niche games. DS games like Brain Age, Brain Age 2 and Nintendogs[6] are very popular among middle-aged people (especially women), who might never consider buying another game. The two best-selling games on the PSP are both from the Grand Theft Auto franchise, which is one of the most well-know and critically acclaimed franchises in gaming and would explain why many people wouldn't think to get another game. The second option is also quite believable, since both consoles have had several significant hardware changes over the years that are not counted as different hardware SKU's[7]. There have been 4 significant versions of the DS and 5 for the PSP, so if plenty of gamers decided to purchase a new unit every time it was upgraded, it would lower these results.

The third reason says that both consoles have a great deal of games available to purchase online and download, which are not counted in the total software results, as these refer to physical copies only. In fact, the PSPgo (the 4th PSP SKU) can't even play physical PSP discs and can only play downloaded titles, meaning every PSPgo sold would count as PSP hardware sales, but could never contribute to software totals.

The three home consoles all have an attach rate between 7 and 8, with the Xbox leading with a 7.77. The main reason that the 360's rate is higher than that of the PS3 is because the PS3 sells 18% more hardware units, but only 7% more software units. The Wii has a high attach rate due to the fact that their games have the highest total sales[8].
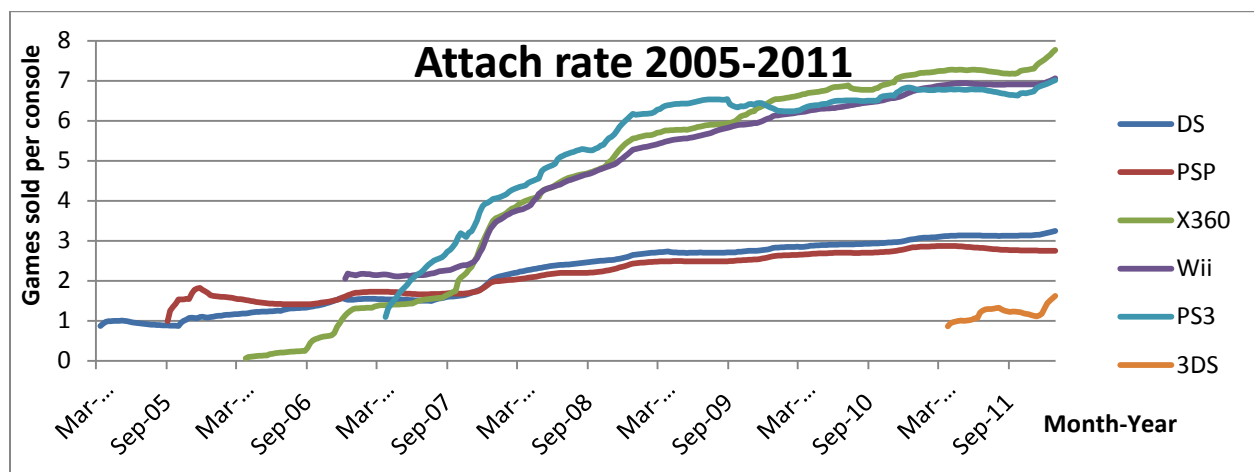
Figure 4: The growth in attach rate for each of the consoles is represented here.

3.5.   Sales decline

The data shows that most titles will launch to a high first week and start to decline every week from that point on. On average games that see decline will decrease with 36% from week 1 to week 2, and then with 33% from 2 to 3. The remaining stats can be found in Table 3.

---

[6] Nintendogs is the best-selling DS game in Europe. Brain Age & Brain Age 2 are the 2nd and 5th best-selling games, respectively.

[7] SKU = Stock Keeping Unit. DS's in different colors still belong to the same SKU, but DS lites have a different SKU from DS's.

[8] Of the 15 best-selling games available 8 are on the Wii, 7 of which have sold over 5 million units. These 8 games make up nearly 40% of the Wii's total software sales (84 million).

Often games that are made to promote new movies will be released a few weeks to a month before the release of the movie. These games will see a few weeks of decline, but then start increasing as soon as the movie launches. The same holds true for most sports games, which tend to be released a month before the start of the season. Games by Nintendo and THQ will often have a much larger second week than their first week, and to a lesser extent this is also true for Ubisoft tiles and PSP games. There are two possible reasons for this exception. First, some of these games are released right before the end of the week, meaning their first week of sales will only count a day or two, whereas the second week will count a full week. Second, the studios could be shipping very little units for the first week. Nintendo has been known to do this, in order to give consumers the impression that their products are selling out rapidly, which has led to an increased demand for their products.

As previously mentioned, the holidays have very increased sales. As a result many publishers choose to release their games right before the holidays. These releases include high-profile games and less anticipated titles. It should be noted that there is no guarantee that a random game released in November or December will sell better than if it were released in another period. This is due to the massive amount of new games being released during the holidays, which tend to overshadow each other. Due to the spike in consoles sales, popular games that have been out for years will also see significant spikes in sales. Table 4 shows the average increase in sales for games who's sales have in fact gone up within the first 5 weeks.

Weeks	2	3	4	5
Average drop	-37%	-33%	-24%	-21%
Std Dev	0.18	0.15	0.13	0.13

Table 3: The average drop in sales for the first 5 weeks.

Weeks	2	3	4	5
Average increase	63%	27%	20%	24%
Std Dev	1.28	0.29	0.17	0.16

**Table 4: The average increase in sales for the first 5 weeks,
counting only games that have increasing sales after the first week.**

3.6.   Sequels

Since sequels always revisit existing properties they tend to be more recognizable than brand new games. The data in Table 5 supports this theory: on average new IP's sell 311,559 units, while sequels sell 353,366. This increase of 13% may not seem that significant, but this data contains many outliers, so a trimmed mean would result in a more reliable result. By trimming off 2.5% of the outliers (1.25% off the top and bottom), we see that new IP sells 229,697 on average, while sequels sell 300,000. This 30% increase is considerably more significant.

	Mean	2.5% Trimmed Mean
New IP	311,559	229,697
Sequel	353,366	300,000

Table 5: The average sales (in units) of sequels and new IP with the trimmed mean (in units).

3.7. Licensed properties

On average promotional licensed property tends to sell below all other types of games. This is primarily because these games are often made on a very tight schedule with a tight budget. These factors come up very often in game reviews and the average MetaScore in Table 6 shows this. Games with the license of a major sports league sell much better than promotional games. The same counts for music games, as these groups are not as limited in their development stage as the promotional titles.

	Average Sales	MetaScore	2.5% Trimmed Mean	2.5% Trimmed MetaScore
Unlicensed	409,603	73.5	313,979	73.7
Promotional	183,024	61.9	164,330	62.0
Sports	340,376	74.2	315,367	74.3
Music	291,161	77.0	270,636	77.0

Table 6: The different forms of licensing and what impact they have on average sales (in units) and average score.

3.8. MetaScore

By observing the relation between MetaScores and total sales in Table 7 we see that the amount of scores on each interval forms a bell curve around the 70's, much like with grades on an exam. As expected, higher scores lead to higher average sales. The only exception is from games scored in the 20's, whose expectations are similar to that of games in the 70's. It should be noted however that of the 7 titles that scored in the 20's, 2 have sold over half a million. The overall expected MetaScore is 70.98.

Score	#	Mean Sales	Std Dev
20's	7	257,143	251,266
30's	32	103,438	136,281
40's	120	139,915	170,510
50's	218	146,449	185,133
60's	375	214,087	260,565
70's	561	257,093	278,043
80's	455	537,438	752,148
90's	109	991,589	832,483

Table 7: The number of games (in units) that were given a certain score, as well as their average sales (in units) and standard deviations.

4. Data analysis

The data analysis chapter contains sections that describe the different methods of prediction as well as their results. The first sections tries to predict the annual sales by finding a formula that matches the actual annual sales. The second sections has finds similar formulas for the monthly sales. The third sections uses the same formulas to predict the weekly sales of the games themselves. The last section shows different methods that were used to predict the total sales of a game using the historical data as well as their respective results.

4.1. Predicting yearly sales

This section aims to find a formula that matches the actual annual sales (hardware and software) of each console. These formulas can then be used to predict the sales in 2012.

The section contains the different formulas tested, how they were tested, and the eventual results. For the sake of validating the formulas we will ignore the actual sales from 2011 and try to predict both 2011 and 2012 using the years before.

4.1.1. Formulas

The four formulas used were: the power formula, the logarithmic formula, the exponential formula, and the polynomial formula (in this case it's a third degree polynomial)

> Power formula:
> Logarithmic formula:
> Exponential formula:
> Polynomial formula:

Here y is the expected sales for a given year and x is the year number of that year. If a console was released in 2005, then the x in 2008 would be equal to 4, for its fourth year on the market. A, B, C, and D are constants that need to be found.

4.1.2. Method

The method used to find the best formula was based on the annual registered sales before 2011. The DS was released in 2005, meaning that the six years from 2005 to 2010 were used, while the PS3 was released in 2007 and therefore only had four years of data to work with. Using the four formulas above, the values for A, B, C, and D were optimized in order to create a formula that closely resembles the actual sales for each of the years before 2011. The formula with the lowest square error in each of the years prior to 2011 was selected as the best formula. These best formulas were plotted with the actual

sales in order to see how closely the expected sales for 2011 matched the registered sales. The formulas were plotted further to show the expected sales for 2012 as well.

4.1.3. Results

It became immediately clear that the power, logarithmic, and exponential formulas were not applicable for the finding the sales of any console. These formulas could not rise in certain areas and fall in others. The third-degree polynomial formula was found to be quite effective for finding a reasonable formula with a low sum squared error, but in some cases a second-degree polynomial was also appropriate. Figure 5 shows a third-degree polynomial plotted against the actual hardware sales for the Wii. The formula itself: matches the actual sales very closely from 2005 to 2010. The expected sales from 2011 are also quite close to the actual sales from 2011. Figure 6 shows the software sales for the PS3 with a second-degree polynomial that grows very similarly to the actual growth from 2007 to 2010. The expected 2011 sales are very close to the actual sales and the software sales are expected to start declining after 2011. 3DS sales were not predicted, because the console was released in 2011. The formulas themselves can be found in the appendix along with the other graphs.
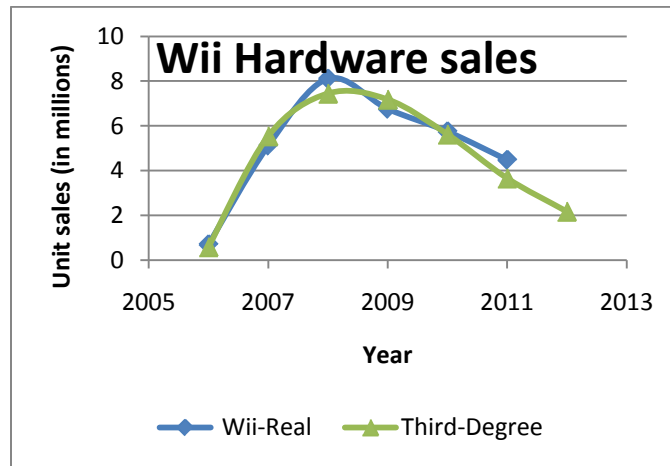
Figure 5: The actual Wii hardware sales (in units) along with the third degree polynomial that shows the expected sales for 2011 and 2012.
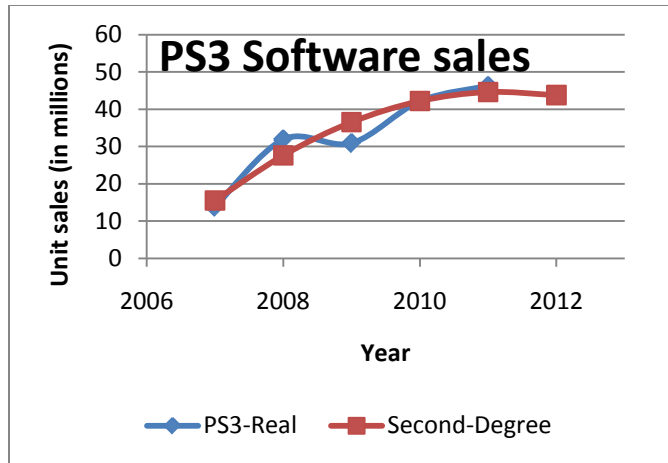
Figure 6: The actual PS3 total software sales (in units) along with the second degree polynomial that shows the expected sales for 2011 and 2012.

The 2011 and 2012 predictions, based on the polynomial formulas, can be seen below. Table 8 shows the hardware predictions and Table 9 shows the software predictions. The 2011 predictions, based on the formulas that best resembled the sales prior to 2011, came quite close to the actual results from 2011, which implies that the 2012 sales could also be fairly accurate. The 2012 predictions show which of the consoles will still have healthy sales in the future and this could greatly benefit a developer's choice when choosing a system for their next game.

Hardware	Actual 2011	2011 prediction	2012 prediction
DS	3,379,576	3,854,341	532
PSP	2,509,782	1,413,200	1,010,600
X360	4,336,196	4,081,424	3,928,439
Wii	4,473,619	3,659,000	2,173,000
PS3	6,115,449	5,852,525	6,397,283

Table 8: The *hardware* sale predictions for 2011 and 2012, along with the actual sales for 2011.

Software	Actual 2011	2011 prediction	2012 prediction
DS	20,569,885	19,118,200	133
PSP	5,402,655	6,065,700	372,100
X360	43,306,801	36,166,563	32,932,317
Wii	39,922,287	35,213,013	653
PS3	46,252,173	44,605,823	43,788,161

Table 9: The *software* sale predictions for 2011 and 2012, along with the actual sales for 2011.

4.2. Predicting monthly sales

This section shows the analyses of the monthly change in sales.

4.2.1. Methods

Looking at every console and every year of sales, I observed the spread of sales between the 12 months. The first year of sales was ignored, since these were did not include the first few months and were therefore lopsided. The sales in 2011 were also ignored in order to compare its actual spread with that of the predicted values. The observed spreads of sales were very similar within the observed years, implying that the 2011 and 2012 spread of monthly sales would not be any different.

4.2.2. Results

The predicted spread was found and compared to the actual spread from 2011. Figures 7 and 8 show the similarities between the expected spread and actual spread for the PS3's sales. Tables 10 and 11 show the actual spread of sales for hardware and software for each of the consoles as well as the average spread, based on the previous years. The tables show (in **bold**) great deviations in the handheld sales for January and December between the actual sales and the averages from the past years, where the home consoles sales were predicted much more accurately. The reason behind these results might be due to the lowering year over year sales in handheld consoles. After the year starts off well, due to the boost from the previous year's holiday sales, they descend into weaker summer sales and eventually lower holiday sales. The remaining graphs can be found in the appendix.
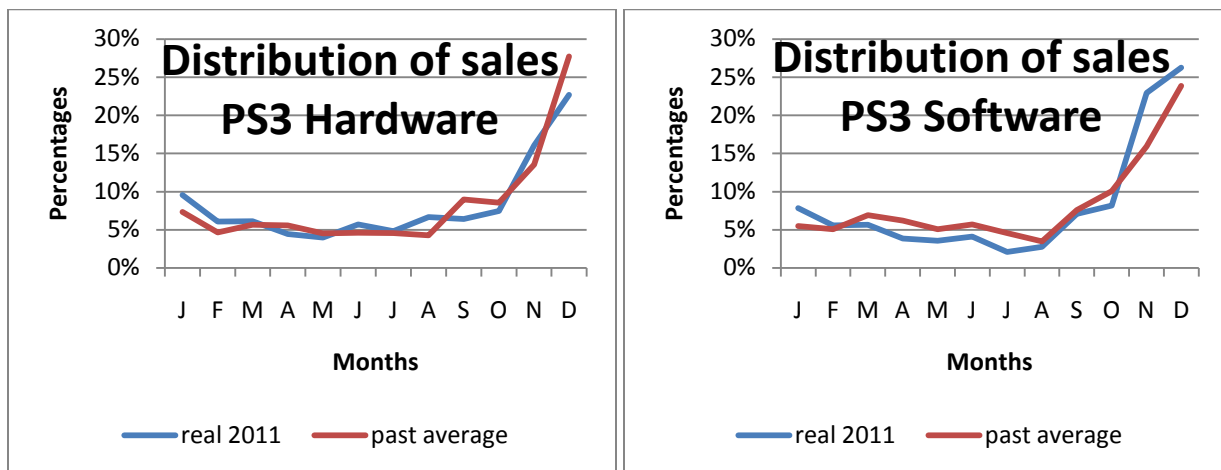
Figures 7&8: The distribution of sales (in units) in 2011 for the PS3 hardware and software with the expected distribution of sales, based on the average values from the prior 3 years.

HARDWARE	J	F	M	A	M	J	J	A	S	O	N	D
DS-real	**11%**	7%	9%	7%	7%	**10%**	9%	7%	4%	4%	11%	**13%**
DS-expected	**6%**	5%	6%	7%	6%	**6%**	8%	7%	6%	6%	13%	**24%**
PSP-real	**11%**	5%	7%	7%	9%	**10%**	8%	8%	5%	5%	11%	**15%**
PSP-expected	**8%**	6%	6%	7%	7%	**7%**	8%	6%	6%	6%	10%	**23%**
X360-real	9%	5%	6%	5%	5%	6%	5%	6%	6%	8%	17%	22%
X360-expected	8%	5%	6%	6%	5%	4%	5%	5%	6%	7%	15%	28%
Wii-real	9%	5%	5%	4%	6%	8%	5%	5%	4%	6%	18%	26%
Wii-expected	7%	5%	5%	6%	6%	5%	5%	5%	5%	7%	13%	30%
PS3-real	10%	6%	6%	4%	4%	6%	5%	7%	6%	7%	16%	23%
PS3-expected	7%	5%	6%	6%	5%	5%	5%	4%	9%	9%	14%	28%

Table 10: Spread of monthly hardware sales per console, as predicted by the average spread, compared to the actual spread.

SOFTWARE	J	F	M	A	M	J	J	A	S	O	N	D
DS-real	**12%**	7%	**13%**	5%	3%	4%	3%	5%	3%	4%	17%	23%
DS-expected	**6%**	5%	**6%**	5%	5%	5%	7%	7%	6%	8%	14%	26%
PSP-real	**20%**	**12%**	**11%**	5%	3%	4%	1%	3%	3%	4%	12%	**21%**
PSP-expected	**7%**	**5%**	**6%**	5%	5%	5%	5%	4%	6%	8%	15%	**29%**
X360-real	8%	5%	6%	4%	4%	4%	2%	2%	7%	8%	23%	28%
X360-expected	5%	4%	5%	4%	4%	4%	4%	4%	9%	10%	19%	28%
Wii-real	**12%**	7%	7%	5%	4%	5%	3%	4%	3%	4%	17%	30%
Wii-expected	**6%**	4%	5%	6%	5%	5%	6%	5%	5%	7%	14%	32%
PS3-real	8%	6%	6%	4%	4%	4%	2%	3%	7%	8%	23%	26%
PS3-expected	5%	5%	7%	6%	5%	6%	5%	3%	8%	10%	16%	24%

Table 11: Spread of monthly software sales per console, as predicted by the average spread, compared to the actual spread.

4.3. What month should a developer release a game?

When publishing a new game, deciding the month of release could greatly influence the amount of sales the game will see. We know from previous sections that the greatest volume of game sales occur during the holidays of November and December, and in this section we will see how this affects the sales of individual games. These results are based on the sales figures for 1581 games. Figure 9 shows the amount of games released in each month, and in it we can see that most games are released in September, October and November. This is done so publishers can benefit from as many weeks as possible within the holidays, a theory that is supported by the fact that December only had 60 releases, which is an 80% drop from November. March and June both had a fairly high number of releases. Most games that miss the holiday release schedule are pushed to March, so that they can still be a part of that fiscal year (The fiscal year for most game publishers ends on March 31). The high number of releases in June is due to publishers wanting to ensure that gamers can get their hands on them before the summer starts.
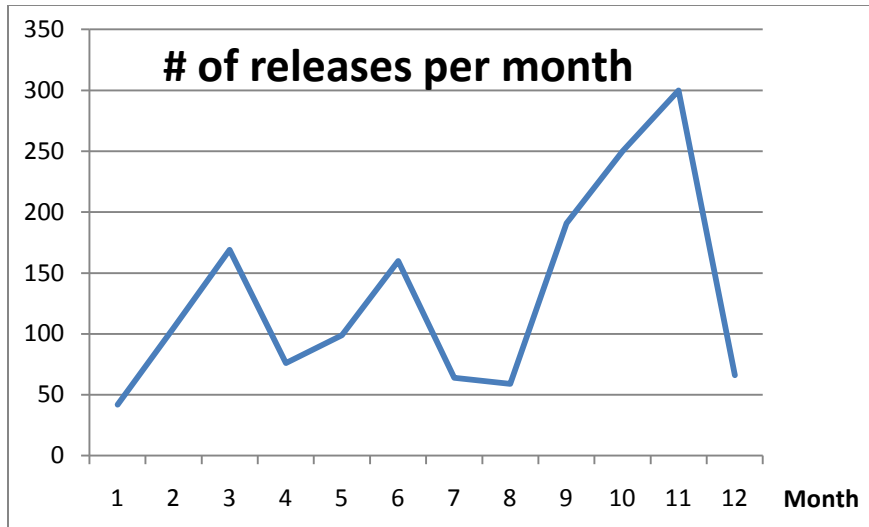
Figure 9: The amount of individual games released in each month. Results are based on 1581 games that were released from 2005 to 2011.

Next we'll look at how well games sell given their release months. Figure 10 shows the average first week sales, as well as the 10% trimmed mean in order to remove outliers. In April we see high first week sales, due to a few outliers that in turn do not show up in the trimmed graph. We also see that the games released in the summer do not have high first weeks either. This is because people are more likely to wait a few weeks before picking something up if they have other responsibilities, such as final exams. October and November had the highest number of releases and also have the highest first week sales, showing that despite all the competition, it is still worth it to release a new game in this period.
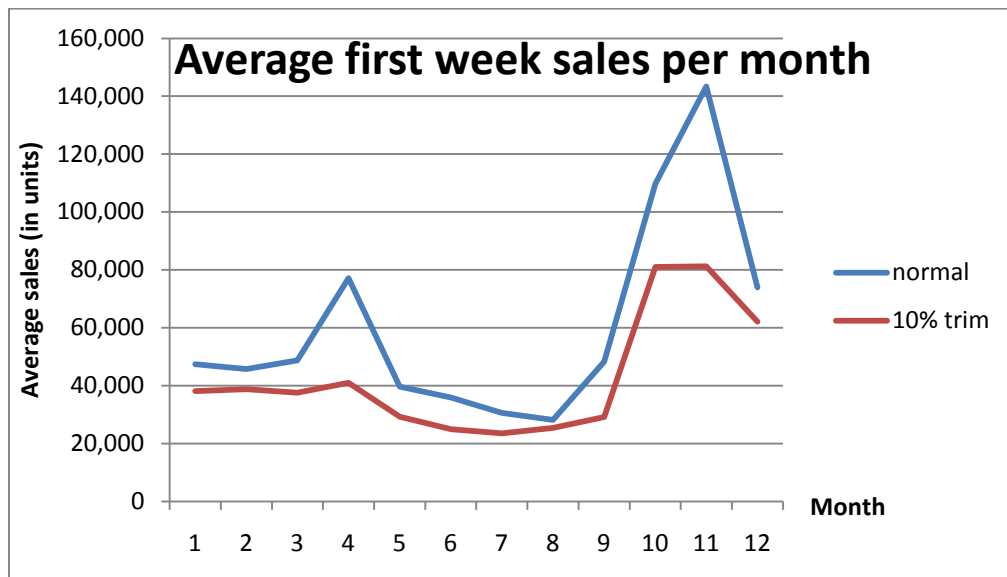
Figure 10: The average first week sales for each month.

Figure 11 and Table 12 show that the total sales do differ a bit from the first week sales. Although the October and November games sell well in their first week as well as overall, they still sell over 20% of

their total sales in the first week, which is quite high compared to the other months. Games in December also sell very well, and since so few games are released in this period, they don't have to worry as much about competition as games released in October and November.
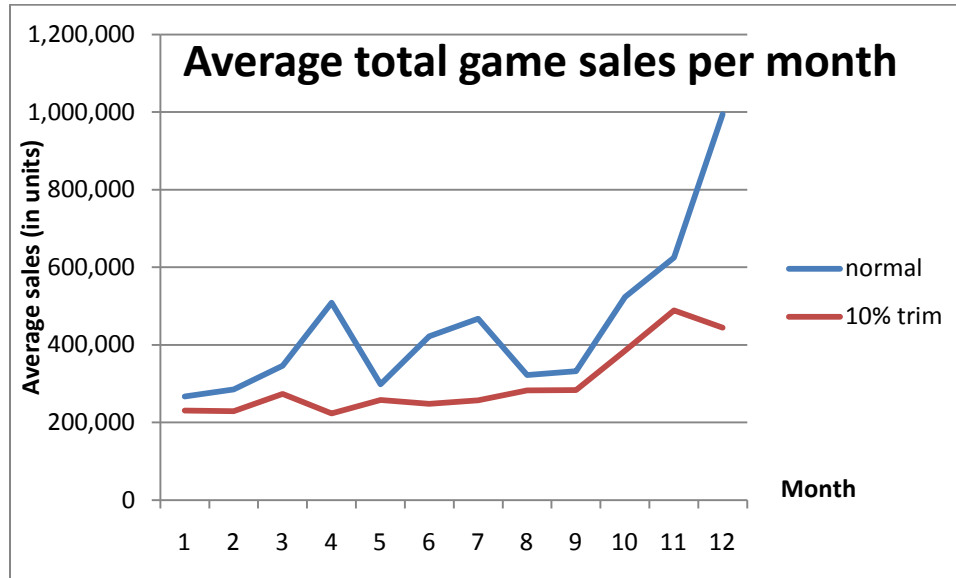
Figure 11: The average total sales for each month.

Month	#	Total	1st week	% 1st week of total
Jan	42	266,667	47,385	17.8%
Feb	105	284,952	45,754	16.1%
Mar	169	345,917	48,658	14.1%
Apr	76	508,553	77,136	15.2%
May	99	298,687	39,610	13.3%
Jun	160	421,750	35,935	8.5%
Jul	64	467,813	30,537	6.5%
Aug	59	322,203	28,104	8.7%
Sep	191	332,094	48,321	14.6%
Oct	250	523,760	109,734	21.0%
Nov	300	624,833	143,367	22.9%
Dec	66	994,697	73,918	7.4%

Table 12: The amount of games that were released in a certain month, their average total sales, average 1st week sales, and percentage of total sales sold in the first week.

4.4. Predicting the change in sales after a game's launch

This section shows the method used to predict game sales after the game's launch using the first few weeks of sales figures.

4.4.1. The first six weeks

This section deals with the weekly sales for the individual games. The database contains 586 games that have had registered sales in the first 6 weeks of their release. The four formulas from 4.1.1. were tested on these 586 games to see which comes up with the most accurate expected sales. The polynomial formula was relatively good at predicting the 6 sales, but would jump to infinity or negative infinity after a few weeks. Both the logarithmic and exponential formulas were unreliable as they would often converge to values below 0 within the first year of sales. The power formula was the best formula when considering descending weekly sales of games. Even though some games could be influenced to increase their sales from one week to the next, all games will ultimately descend to zero sales a week with enough time. Of the 586 games 483 had power formulas that showed a descending line. The power formulas encountered looks like this: , where A is the amount of first week sales, and B is the factor that determines the rate of descent. The average value for B was 0.8 (as well as the median). Given the first week (actual or predicted) sales of a particular title, one could use this formula (with the average of 0.8) and get a good estimate of the following weekly sales. Figures 12 and 13 show the actual sales for the first 6 weeks with the expected sales using the formula above.
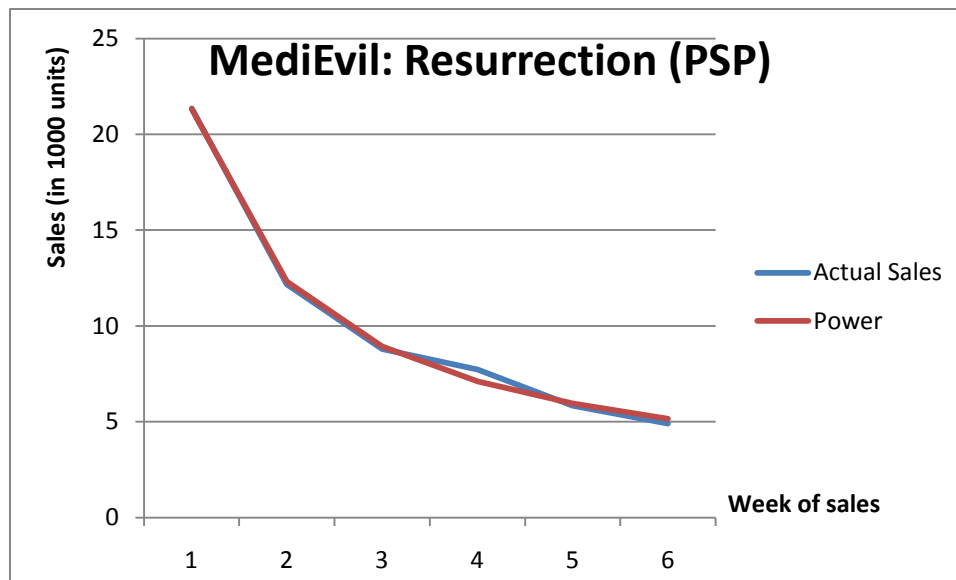
Figure 12: A prediction of weekly sales using the power formula based on the first 6 weeks of data.

26
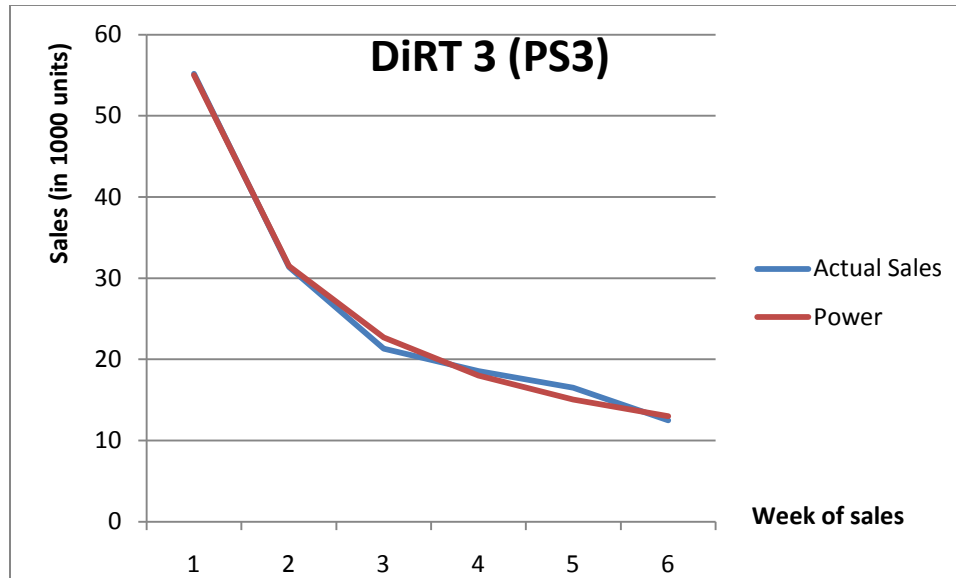
Figure 13: Another prediction of weekly sales using the power formula based on the first 6 weeks of data.

Games that experience a sudden increase in sales, be it because of the coming holidays or a new advertising campaign, cannot be predicted by the power formula, since it only moves in one direction. Graphically, the increase in sales appears as a temporary shift from the power formula to a second-degree polynomial formula, although the actual values of these polynomials cannot be easily predicted, because their height and widths are always different. It should be noted, however, that once the sales start to descend once more, one of two scenarios tend to occur. In the first, the sales return to the expected sales as predicted with the initial power formula. In the other case the initial power formula no longer fits, because the sales start descending as if a new power formula is started at the sales peak.

4.4.2. The first year

In this section we will use the formulas found in the last section and compare the predicted sales with the actual sales from the entire first year. Games that had too few registered weekly sales in their first year were ignored as well as games that had very erratic jumps in sales figures. The remaining games can be split into four possible groups.

The first possibility is that the power formula, which is based on the first six points, creates an array of expected sales that very closely matches the actual sales for every week that is registered. This is the ideal scenario, but it is very uncommon, since there are many factors that could cause a deviation from the power formula. Figure 14 shows a graph where the B of 1.44 creates predicted values that line up perfectly.
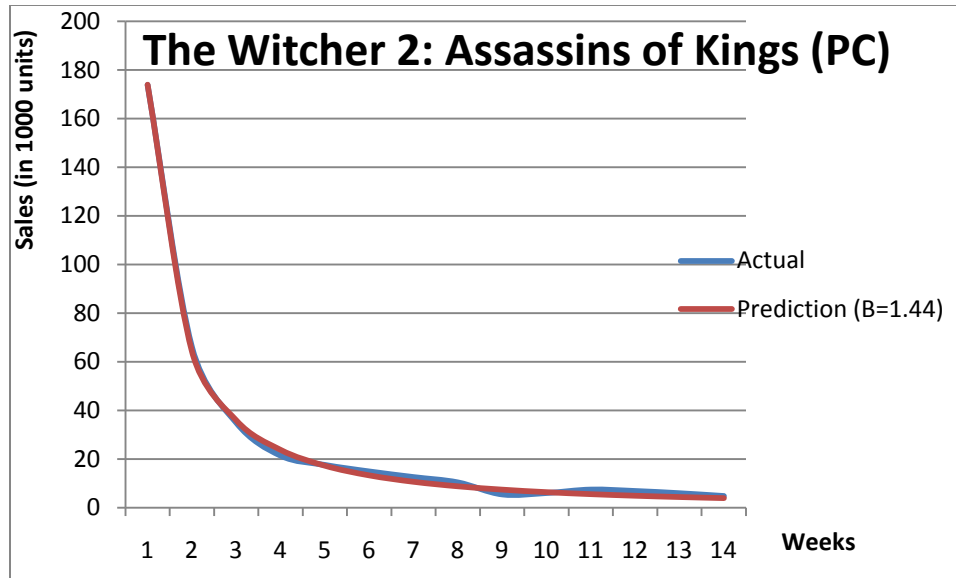
Figure 14: A game who's actual sales very closely match the predicted sales on every point.

The second possibility is much more common than the first and shows a an array of expected sales that closely matches the actual sales, except for a part where the actual sales rise to a much higher point and drop down again to the level of the expected sales as in Figure 15. These jumps always occur during the holidays for six to ten weeks, so their occurrence can be predicted, but the actual height of the holiday sales cannot be predicted. The Appendix shows many more graphs that are accurately predicted and have a holiday boost.
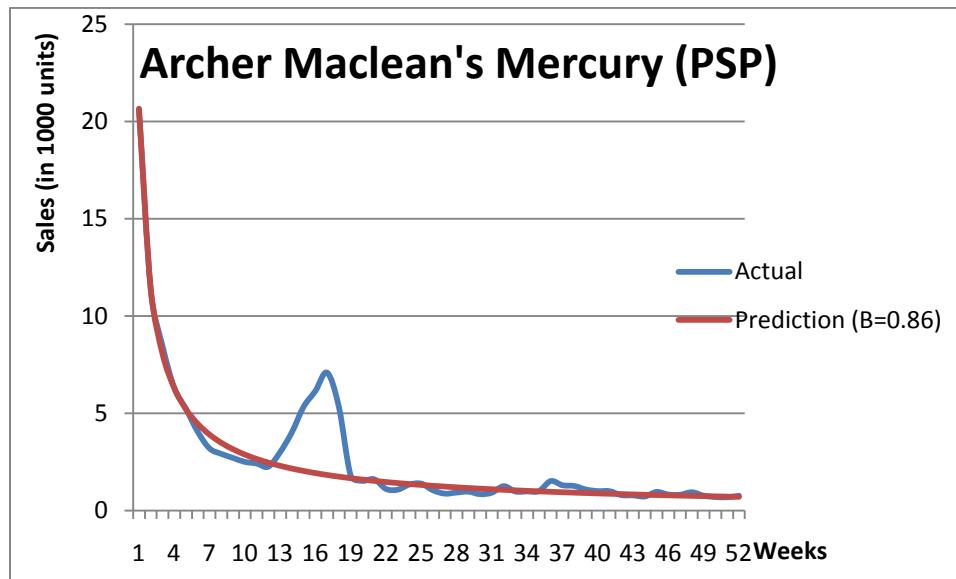
Figure 15: A game who's actual sales very closely match the predicated sales in every point, except for the holiday season.

The third possibility is when one of the six points at the start of the game's life cycle is an outlier and causes the power formula to not be properly aligned with the actual sales for the first years as in Figure 16. If the outlier is ignored during the process of finding the ideal power formula, the expected results would align with the actual sales.
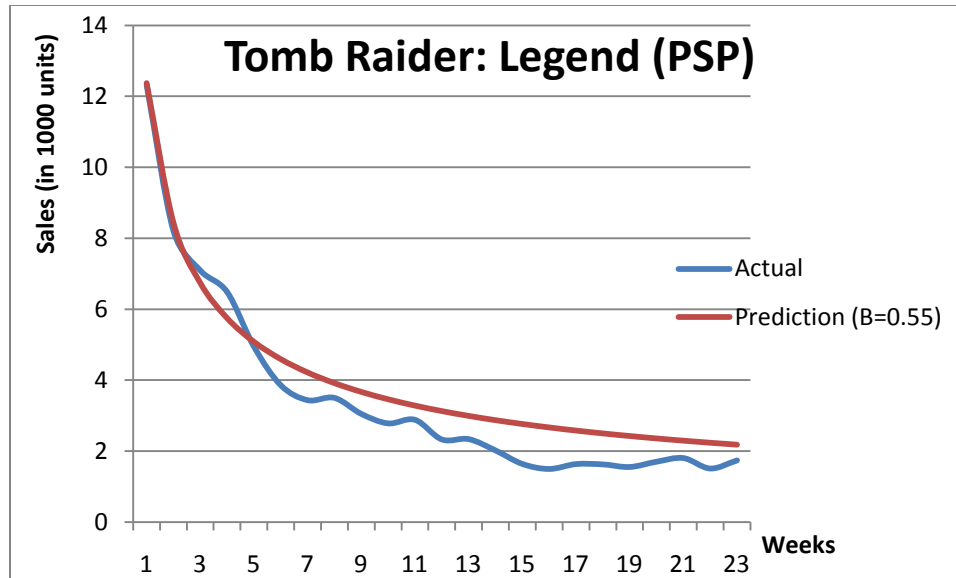
Figure 16: A game who's actual sales do not match up very well with the predicted sales, due to an outlier in the first 6 weeks.

The last possibility is when the sales rise at some point in the first six weeks and then start dropping at a more expected rate from that given high point. Figure 17 shows a plot (using B=0.52) that is much higher than the actual sales units. The reason for this deviation is due to the difference in sales between first and second week. The sales from the first week are far too low and should be ignored in order to base the prediction on the later points. The recalculated power formula matches the predicted results to the actual results perfectly, as in Figure 18, where a different starting point and a different B (.79) were used.
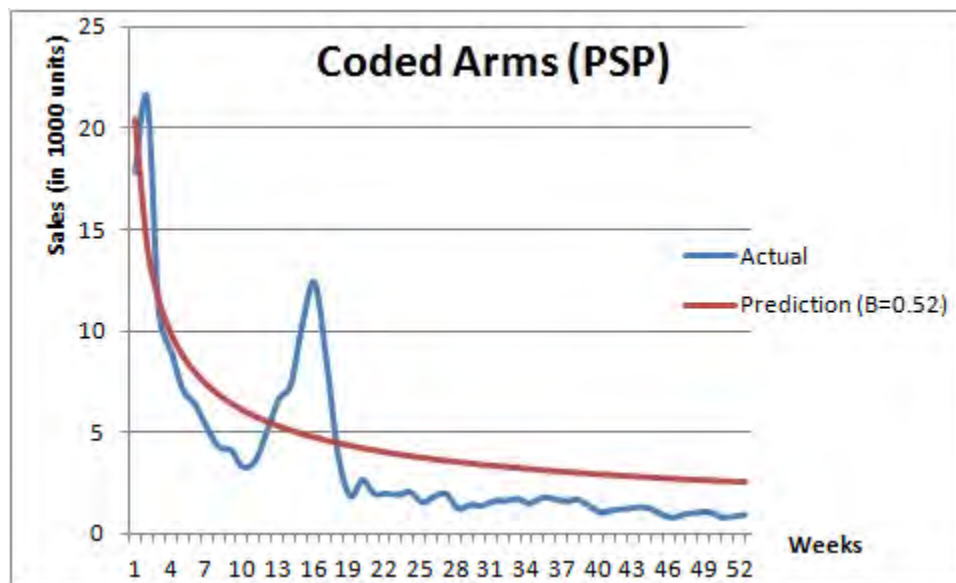
Figure 17: A game who's actual sales do not match up with the predicted sales, due to an outlier. The peak between weeks 11 and 18 is due to the holidays and does not influence the prediction.
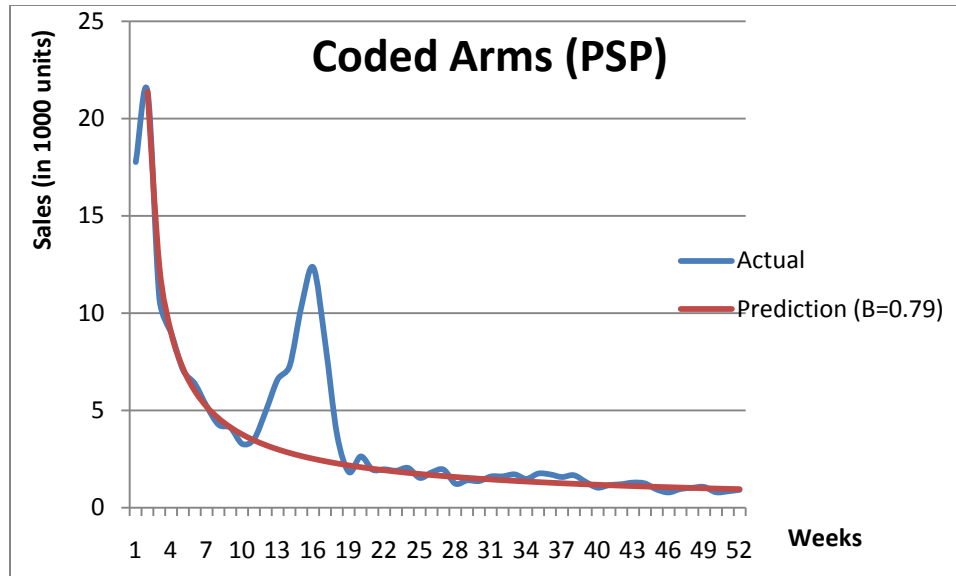
Figure 18: the same games as figure 15, but this time the power formula was based on using the outlier as a starting point.

4.4.3. Optimizing the number of points

The previous two sections deal with the prediction of sales using a formula that is derived from the first 6 weeks of actual sales. This section aims to compare the accuracy of the predictions when they are made using fewer actual points. The tests were executed using 2 to 10 of the first actual points and for each number of points a power formula was created. This power formula was used to estimate the first 10 and first 52 weeks of sales and these were compared to the actual sales in those periods. The games were separated into two groups: games that sold less in their second week than in their first week and games that showed an increase in sales during the second week. The errors were based on the percentage of the absolute difference between the real weekly sales figure and the expected one. The average of all the percentages in the array was then calculated to find the average error per game. The reason averages were used instead of a sum of errors was because certain games had 52 registered sales in a year, while others had fewer than 20. The average error could be used to view these games on equal ground. The average error was calculated for the 344 games that had descending sales in the first 5 weeks and the results can be seen in Table 13. The value of 49% in 10 weeks for 2 points means than, on average, all of the expected sales for the first 10 weeks found using only the first 2 points had a margin of error of under 50% from the actual recorder sales. It should also be noted that, even though the averages show that the error gets smaller as you add more points, the formula made from the 2 point approach is more accurate than the other three choices in 16% of the cases. In fact, the five point approach is only more accurate in 62% of the cases. The perfect amount of points to use when predicting the coming weekly sales depends on how much more accurate the user wants the results to become. Choosing 4 points over 3 means getting results that are 5.1% more accurate, but the user needs to decide whether 5% is worth waiting another week for. The results show that for an entire year the error will even increase when selecting 5 points over 4.

Points Used	10 weeks	52 weeks
2	49%	67%
3	33%	46%
4	28%	43%
5	25.9%	45%
6	25.7%	51%
7	24.6%	54%
8	23.7%	52%
9	23.33%	48%
10	23.27%	44%

Table 13: The average error in 10 weeks of time or 52 weeks of time, based on how many point were used to create the formula.

4.5. Predicting total game sales

This section shows the application of a method, called the observation based method, which predicts the total sales for a game based on 8 input fields as well as the expected difference in sales when certain fields are changed. This method was validated by predicting the total game sales for all 2450 games in the database and comparing these to their actual results. The methods can be applied using the included program.

4.5.1. Method

The observation based model finds the expected sales based a combination of the information found in chapter 3 and the assigning of values to the entries of certain fields. One of the observations from the data was that sequels sell 30% better than new IP. For this reason, the observed model finds the expected sales based the fields and multiplies this by either 1 or 1.3, depending on the value of the sequel field. This idea is repeated for four other fields: License, Genre, Month and Score (itself combined into tens). Since Console and Year vary greatly (as seen in Ch3), these fields are combined into a large field. The remaining two fields are Developer and Publisher. All the values within these fields are given numbers that represent the basic number of sales expected from these names without considering the factors that alter these sales contained in the other fields. The numbers for the Developer and Publisher are added up and then multiplied by the factors for each of the fields, resulting in the expected sales for the given game.

4.5.2. Results

When the observation based method is applied using the fields of the 2450 games in the database the results from Table 14 are found. It shows that the resulting predicted number of total sales is larger or smaller than the actual sales with a certain probability. For example: if the predicted value is 300,000,

then the probability that the actual sales will be below 200,000 will be 24% and the probability that the actual sales will be over 272,000 60% (1-40%).

Predicted value is … than actual value	#	%
More than 50% larger	581	24%
Less than 50% smaller	721	29%
In between these 2 values	1148	47%
More than 20% larger	891	36%
Less than 20% smaller	901	37%
In between these 2 values	658	27%
More than 10% larger	989	40%
Less than 10% smaller	965	39%
In between these 2 values	496	20%

Table 14: The accuracy of the predicted sales compared to the actual sales.

4.5.3. Predicting the change in total sales when changing input fields

Using the same observed method a new method can be created that predicts the difference in sales between two different scenarios, without the need to include a known publisher and developer. Using the remaining six fields, a game creator could compare how the decision to delay a game a few months would affect their total sales, or what kind of influence a high expected MetaScore can have, compared to a low one. These decisions can be combined to help a publisher decide if it's wiser to release an unfinished average game in the high season or delay it and release a more polished great game in the low season.

5. Conclusions

The main question a game developer or publisher would ask is: How can we benefit from this report? The answer is split into several parts, so that each serves a specific purpose during a certain part of the development cycle.

Pre-production phase

The comparison of console hardware and software sales and the future predictions that arose from these analyses serves to give a better idea of which consoles will still be dominant in 2012 and beyond. These results show that the DS, PSP and Wii are reaching the end of their life cycles, while the PS3 and Xbox 360 are still close to their peak. This information about market share is very important for a developer who is still in the early concept stages and choosing what console to develop for.

Production phase

When the development is well on its way, the publisher must consider the best month to release the title, in order to maximize the benefits. Even though the holidays have the highest number of releases, the holidays also have the highest number of average sales per game, making it the best period to release your games.

Post-production phase

Sales decline is an essential part in understanding in what way sales will develop. The expected sales can be calculated from the first weeks of sales and can be reevaluated every week for accuracy's sake. The supplied program contains the functionality that allows a publisher to enter the first 2 to 6 weeks of sales in order to predict the expected sales the game will see in the coming weeks. The program can further supply the user with an estimate of the total sales for any interval of time.

The included observed model can be used at any stage of development to estimate the total sales based on 8 different fields. The program includes both a function that predicts the estimated total sales and a function that can be used to compare 2 different scenarios to see how they would influence the expected total sales.

Appendix

The appendix contains graphs and tables related to the previous chapters.

Charts from chapter 3

Monthly *hardware* sales per console

The following graphs show the monthly sales in *hardware* units for every year of each of the 5 consoles.
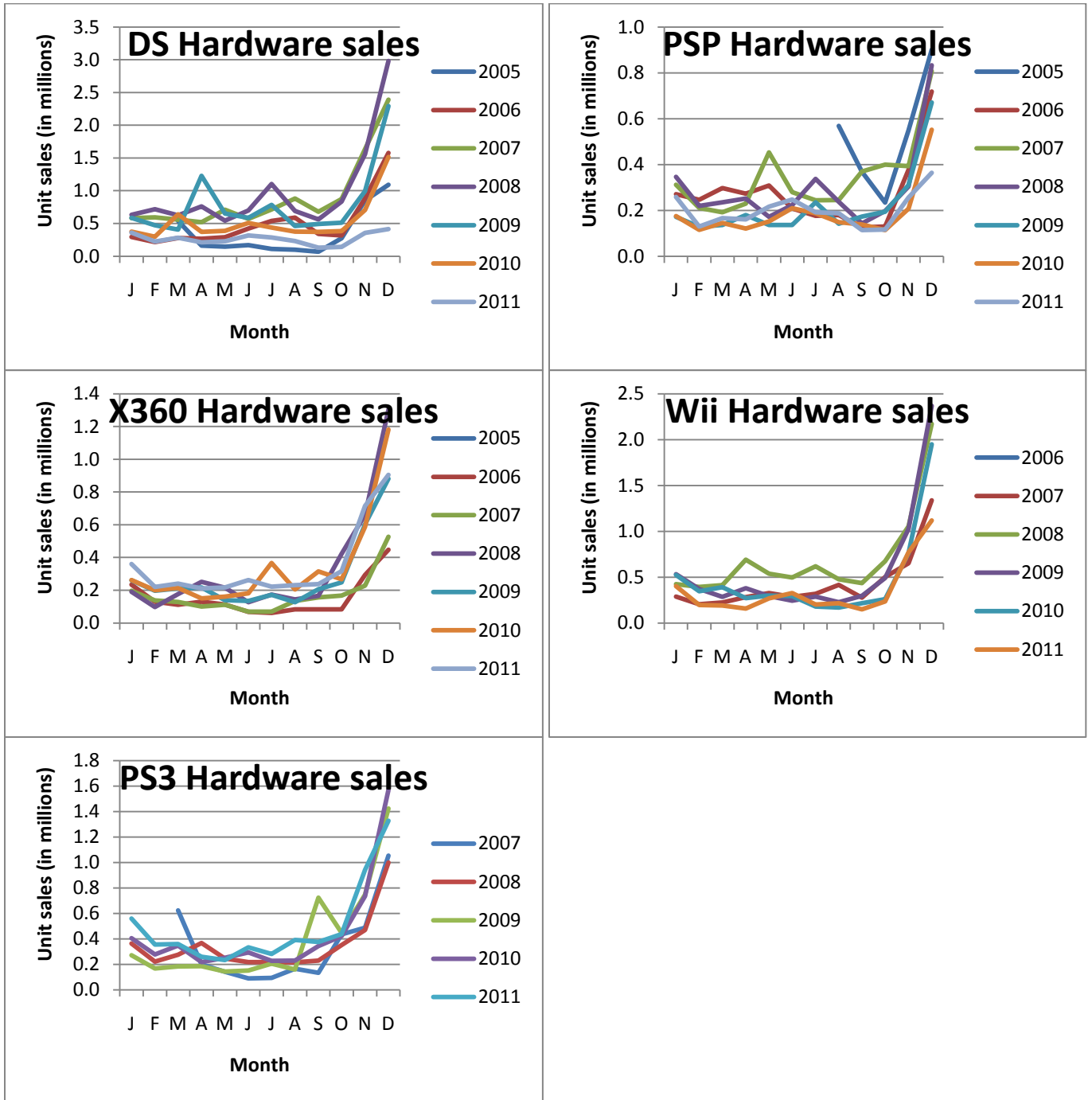
Monthly *software* sales per console

The following graphs show the monthly sales in *software* units for every year of each of the 5 consoles.

DS Software sales

PSP Software sales

X360 Software sales

Wii Software sales

PS3 Software sales
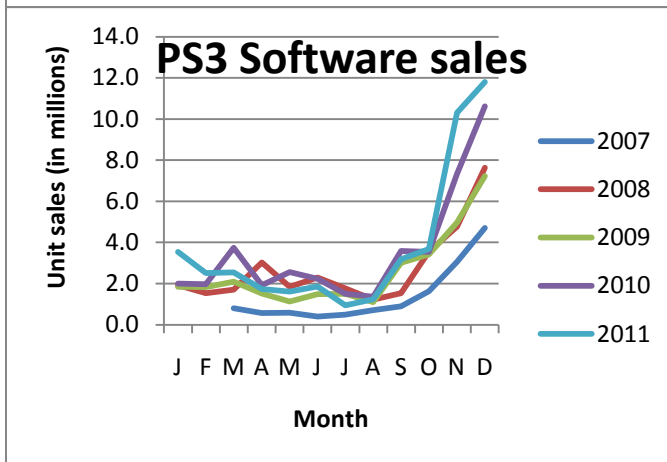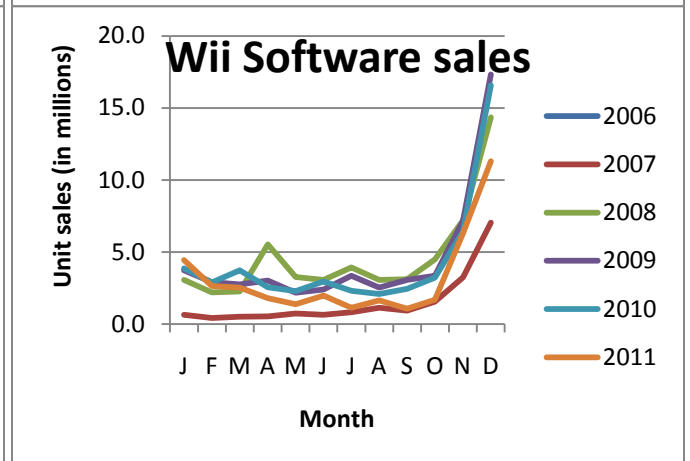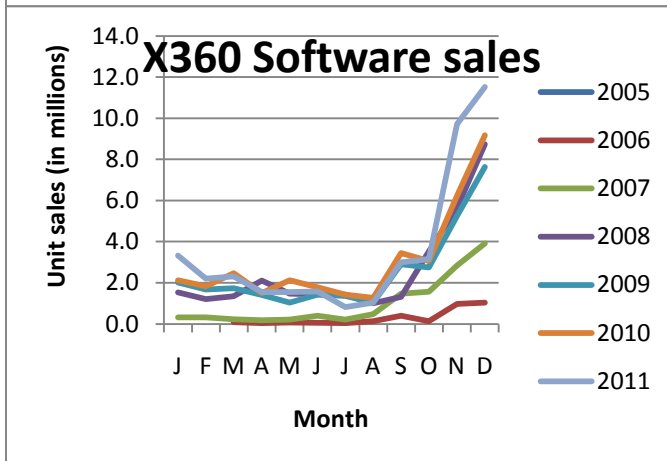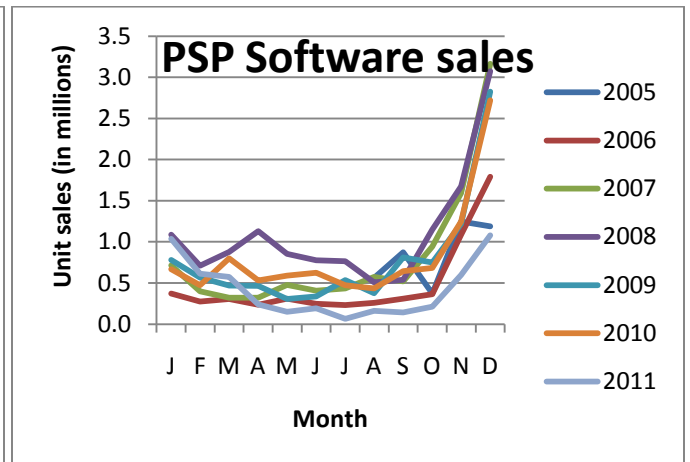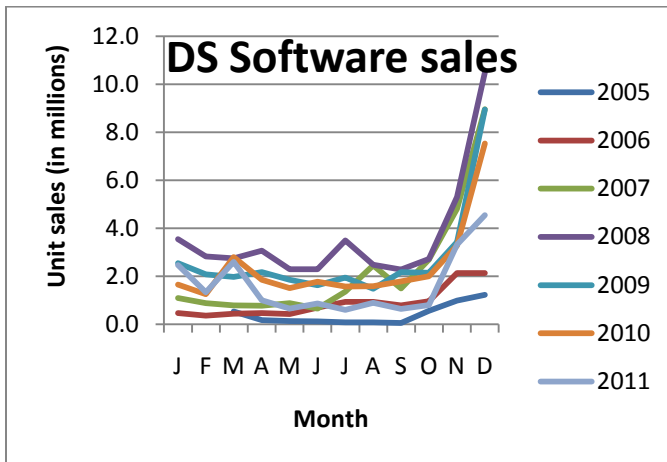
Yearly hardware sales

The table shows the constants for the polynomial formulas used for predicting the *hardware* sales for each of the consoles. The graphs show these formulas plotted with the actual sales (in units).

Hardware	A (X^3)	B (X^2)	C (X)	D
DS	57,026	-1,486,371	8,804,368	-4,503,967
PSP	33,800	-570,700	2,445,700	664,200
X360	0	-107,532	1,459,997	-869,481
Wii	147,000	-2,408,000	11,149,000	-8,299,000
PS3	0	-13,858	697,195	2,713,000
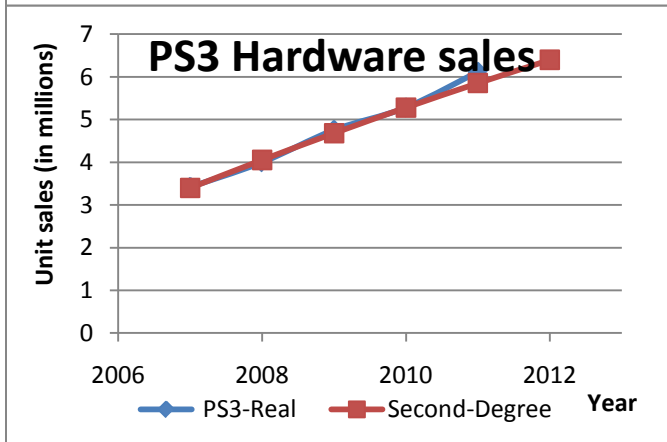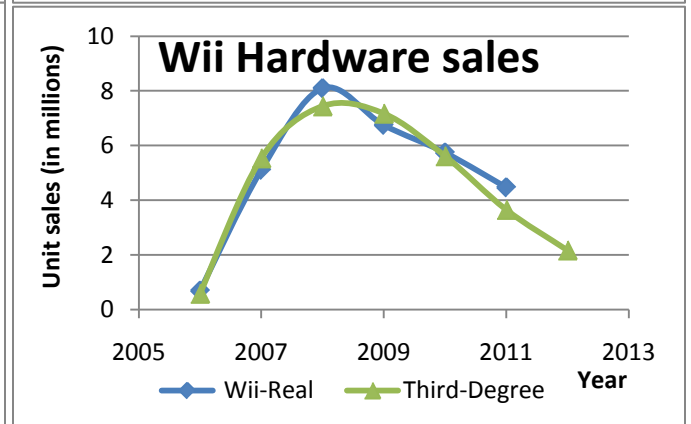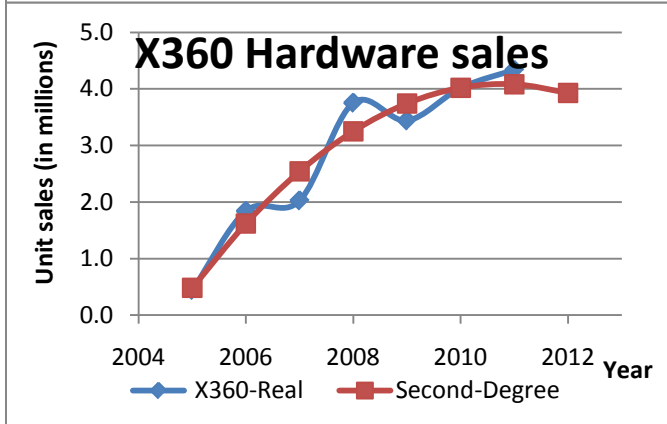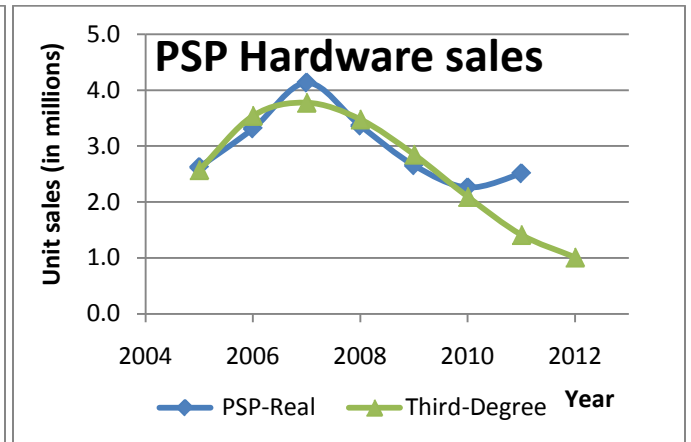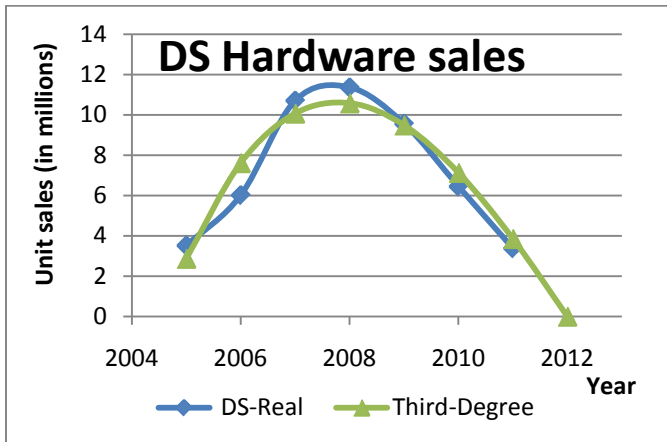
Yearly software sales

The table shows the constants for the polynomial formulas used for predicting the *software* sales for each of the consoles. The graphs show these formulas plotted with the actual sales (in units).

Software	A (X^3)	B (X^2)	C (X)	D
DS	-123,666	-1,177,239	19,440,115	-16,860,359
PSP	-47,400	-110,100	3,968,500	-60,700
X360	0	-1,677,643	18,575,111	-14,888,963
Wii	-445,169	-835,475	32,185,332	-31,665,287
PS3	0	-1,617,845	16,978,631	158,790

DS Software sales — DS-Real, Third-Degree

PSP Software sales — PSP-Real, Third-Degree

X360 Software sales — X360-Real, Second-Degree

Wii Software sales — Wii-Real, Third-Degree

PS3 Software sales — PS3-Real, Second-Degree
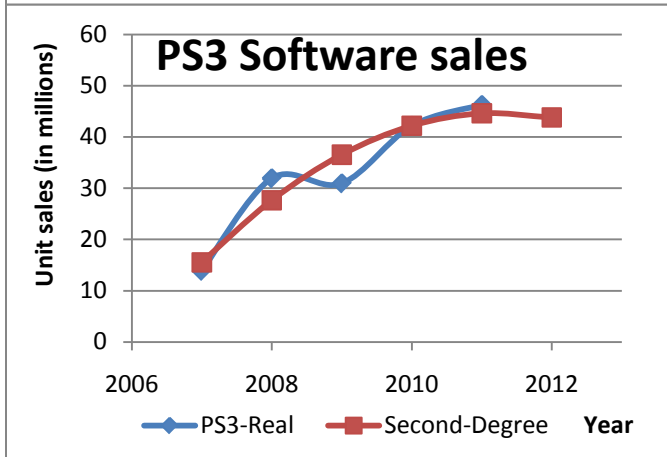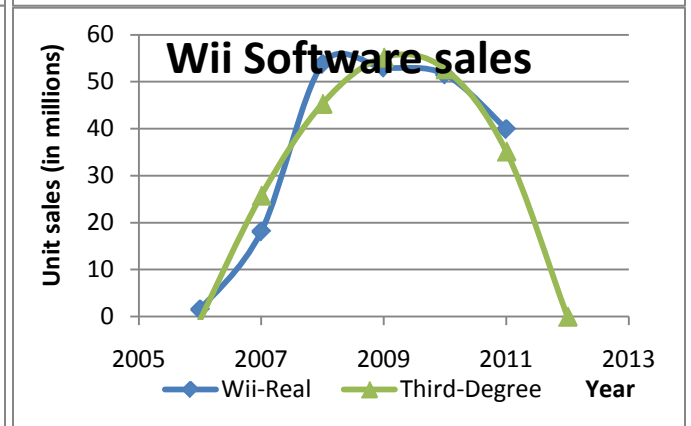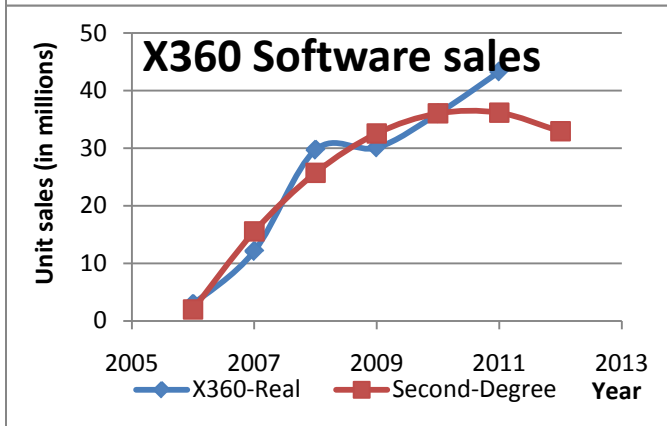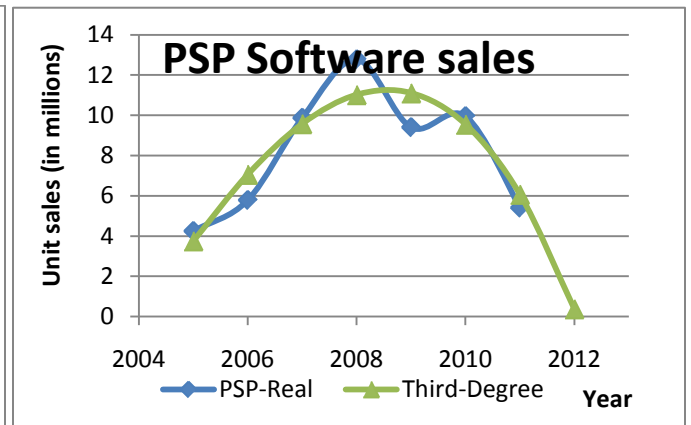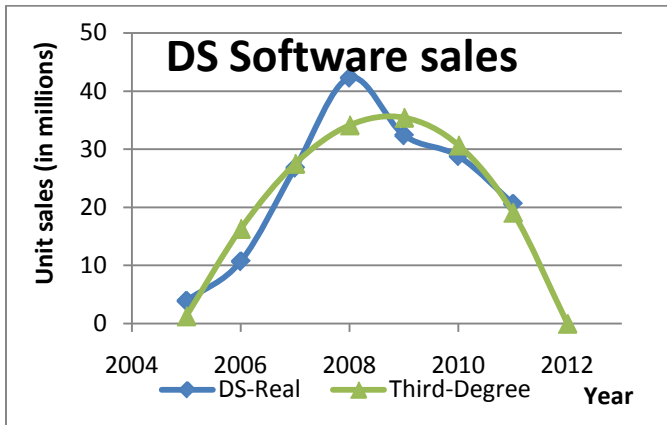
Charts for section 4.2.

Monthly sales spread: Hardware

These graphs show the actual distribution of *hardware* sales in 2011 with the expected distribution based on the average from earlier years.
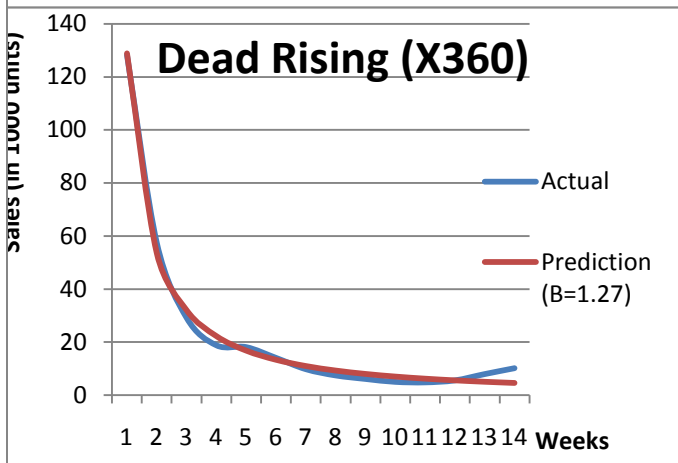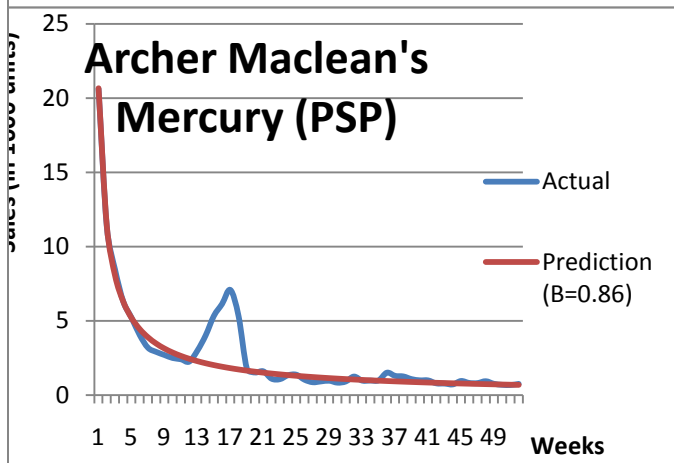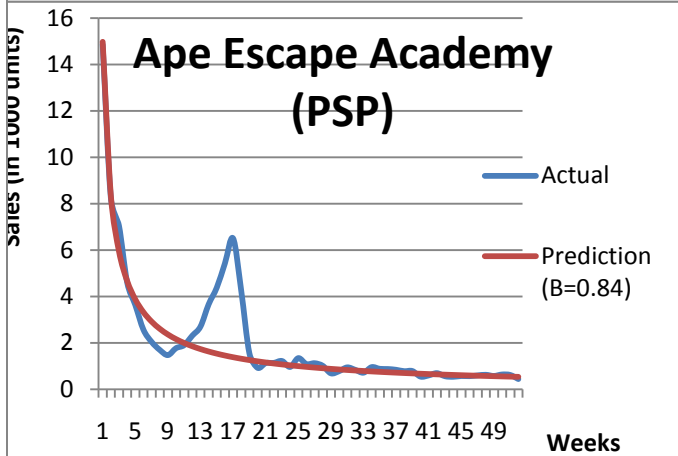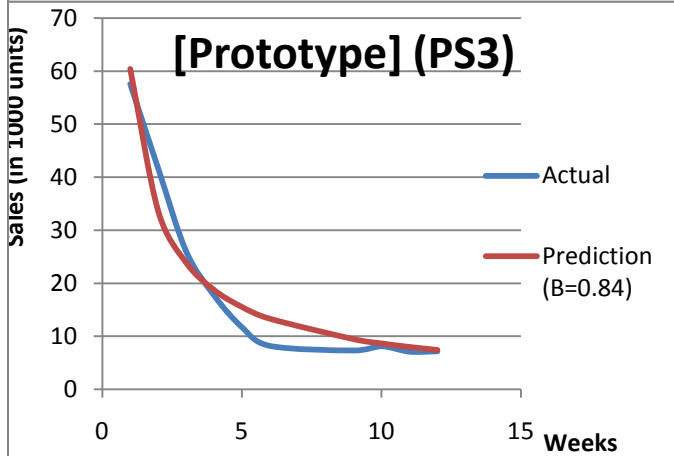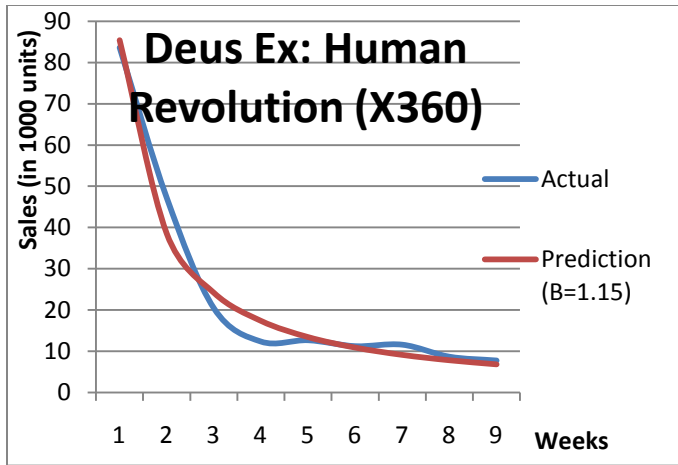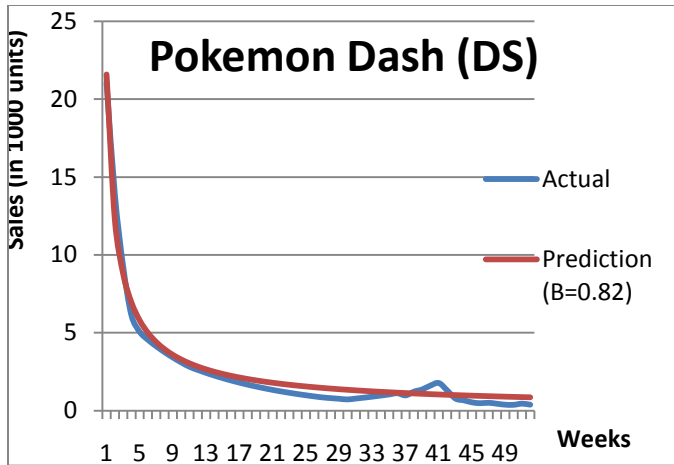

Distribution of sales DS Hardware

Distribution of sales PSP Hardware

Distribution of sales X360 Hardware

Distribution of sales Wii Hardware

Distribution of sales PS3 Hardware
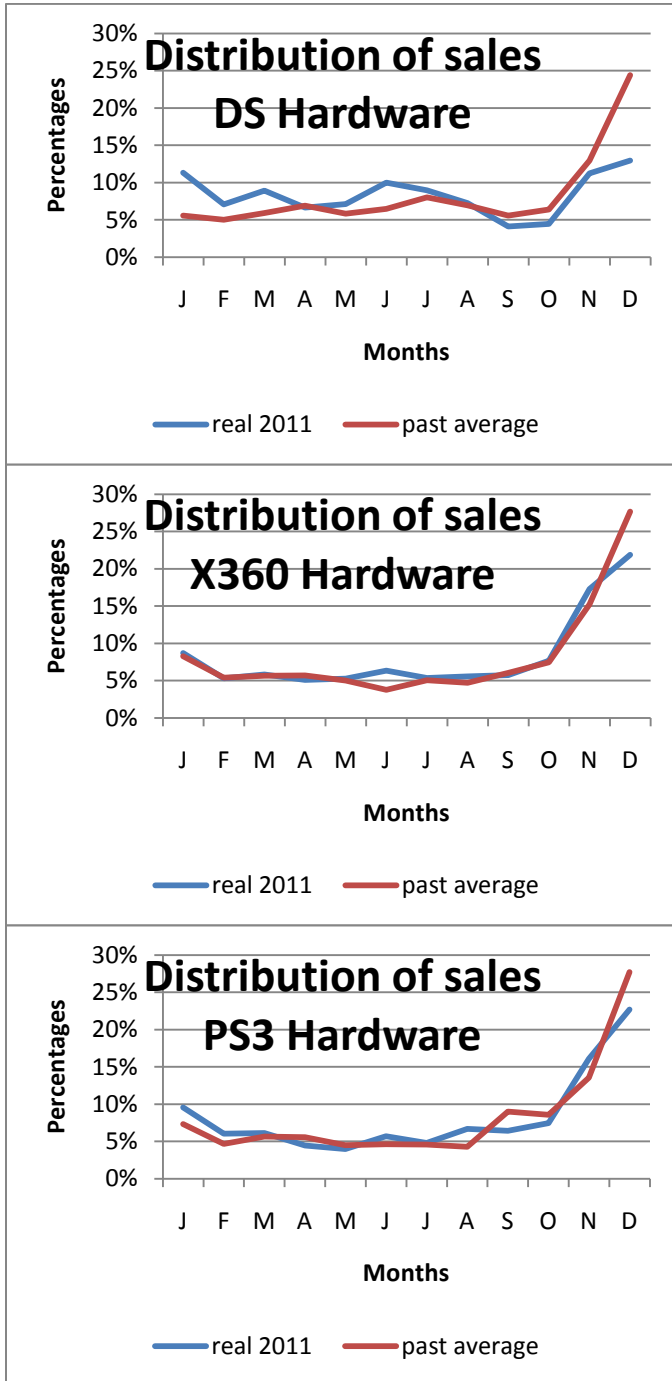
Monthly sales spread: Software

These graphs show the actual distribution of *software* sales in 2011 with the expected distribution based on the average from earlier years.
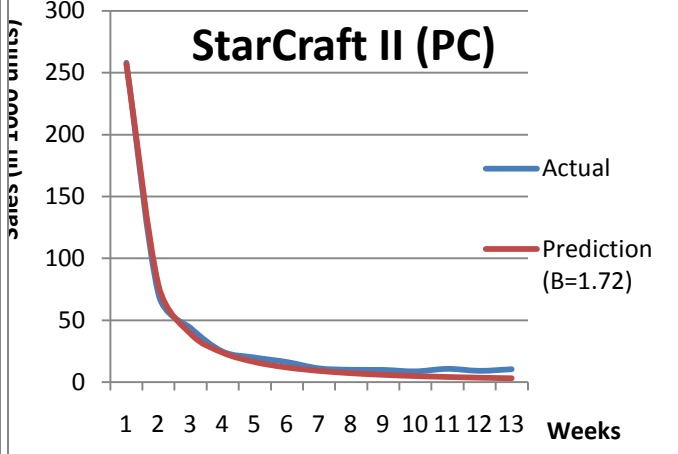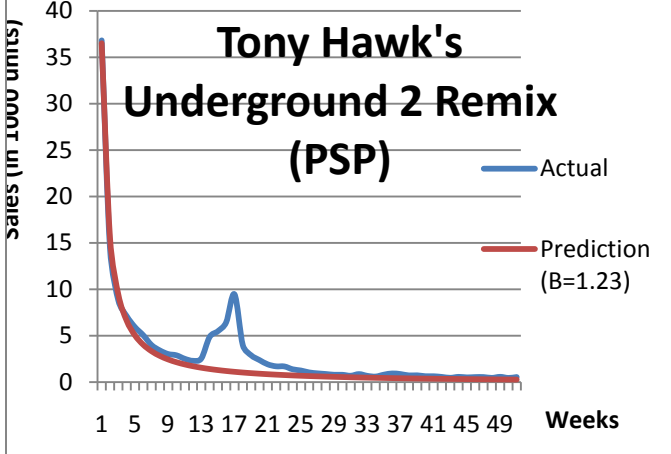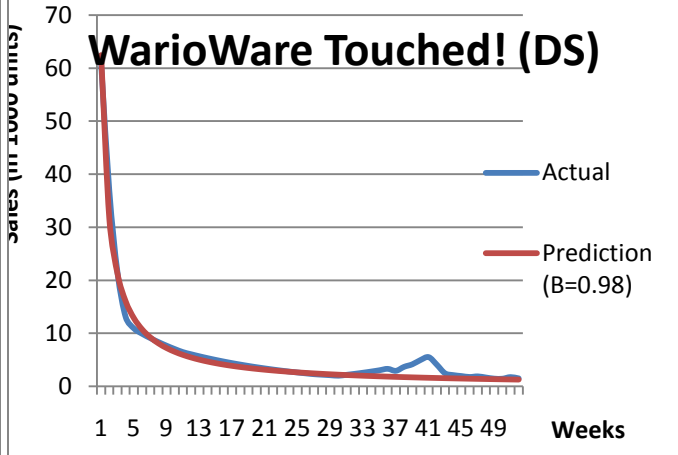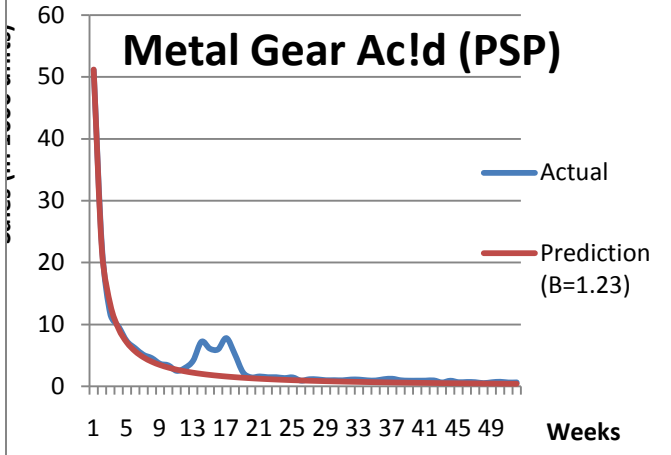

Distribution of sales DS Software

Distribution of sales PSP Software

Distribution of sales X360 Software

Distribution of sales Wii Software

Distribution of sales PS3 Software
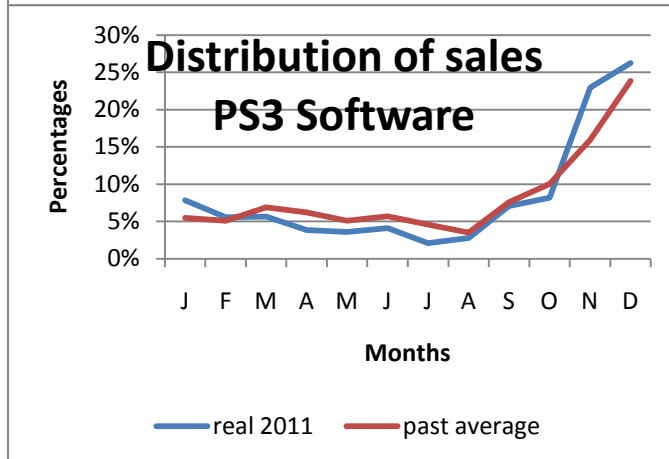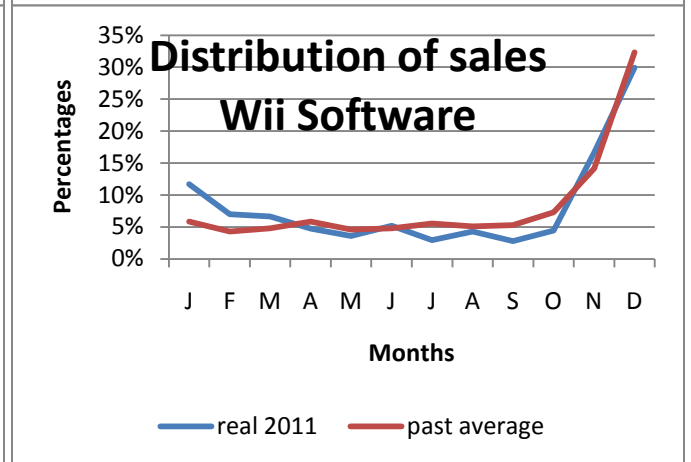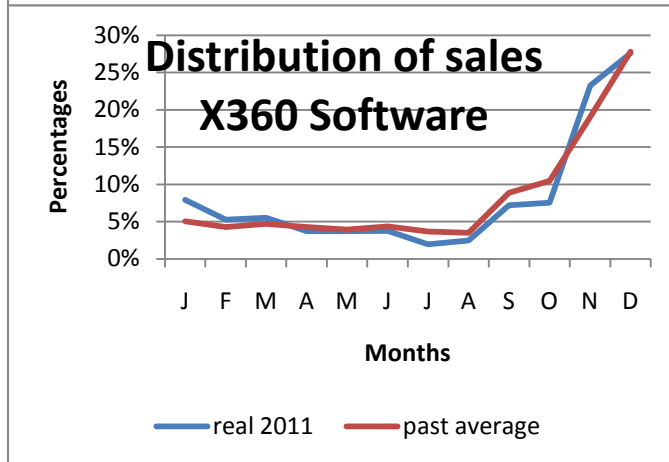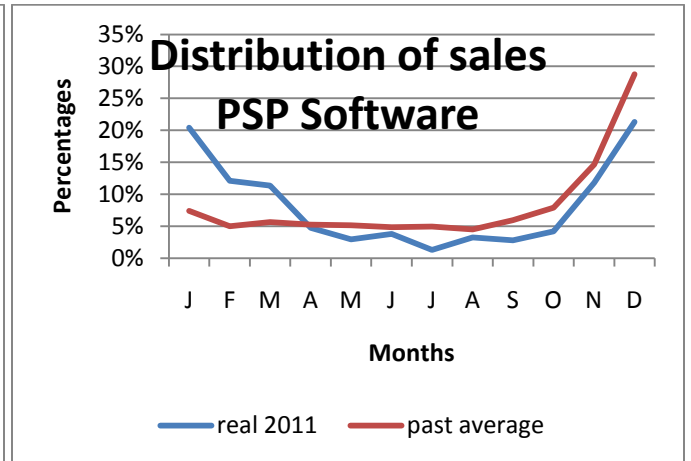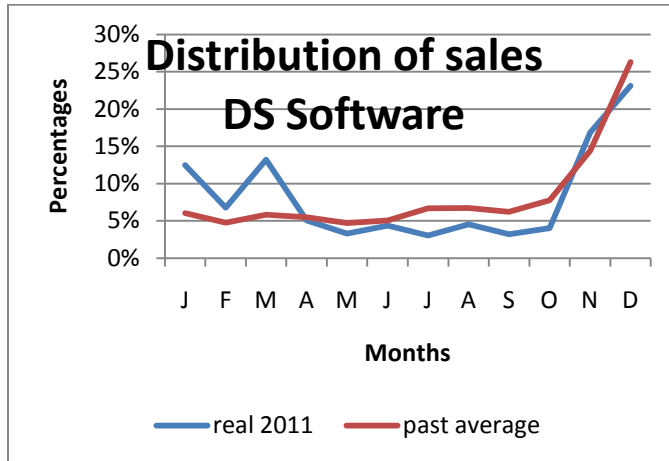
Charts for 4.4.

These graphs show the first weeks of actual sales (in units) for several different games, as well as a plot of the expected sales based the power formula using the value in the graph as the constant B.

Call of Duty 3(PS3)

Asphalt: Urban GT (DS)

The Sims 2 (PSP)

Tomb Raider: Legend (PSP)

UFC 2009 Undisputed(PS3)

The Witcher 2: Assassins of Kings (PC)

Pokemon Dash (DS)

Sales (in 1000 units)

Weeks

Actual

Prediction (B=0.82)

25 20 15 10 5 0

1 5 9 13 17 21 25 29 33 37 41 45 49

Deus Ex: Human Revolution (X360)

Sales (in 1000 units)

Weeks

Actual

Prediction (B=1.15)

90 80 70 60 50 40 30 20 10 0

1 2 3 4 5 6 7 8 9

[Prototype] (PS3)

Sales (in 1000 units)

Weeks

Actual

Prediction (B=0.84)

70 60 50 40 30 20 10 0

0 5 10 15

Ape Escape Academy (PSP)

Sales (in 1000 units)

Weeks

Actual

Prediction (B=0.84)

16 14 12 10 8 6 4 2 0

1 5 9 13 17 21 25 29 33 37 41 45 49

Archer Maclean's Mercury (PSP)

Sales (in 1000 units)

Weeks

Actual

Prediction (B=0.86)

25 20 15 10 5 0

1 5 9 13 17 21 25 29 33 37 41 45 49

Dead Rising (X360)

Sales (in 1000 units)

Weeks

Actual

Prediction (B=1.27)

140 120 100 80 60 40 20 0

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Halo Wars (X360)

Crackdown(X360)

Metal Gear Ac!d (PSP)

WarioWare Touched! (DS)

Tony Hawk's Underground 2 Remix (PSP)

StarCraft II (PC)

WipEout Pure (PSP)

Sales (in 1000 units) — Weeks

Actual / Prediction (B=1.2)

TOCA Race Driver 2: Ultimate Racing Simulator (PSP)

Sales (in 1000 units) — Weeks

Actual / Prediction (B=0.95)

MediEvil: Resurrection (PSP)

Sales (in 1000 units) — Weeks

Actual / Prediction (B=0.79)

Lumines: Puzzle Fusion (PSP)

Sales (in 1000 units) — Weeks

Actual / Prediction (B=0.7)

Need for Speed Underground Rivals (PSP)

Sales (in 1000 units) — Weeks

Actual / Prediction (B=0.89)

Metal Gear Solid 4: Guns of the Patriots (PS3)

Sales (in 1000 units) — Weeks

Actual / Prediction (B=1.35)

Hot Shots Golf: Open Tee (PSP) — Actual; Prediction (B=0.69)

Coded Arms (PSP) — Actual; Prediction (B=0.79)

Worldwide Soccer Manager 2009 (PC) — Actual; Prediction (B=1.04)

Super Smash Bros. Brawl (Wii) — Actual; Prediction (B=0.91)

FIFA Soccer 09 (PS3) — Actual; Prediction (B=1.14)

Pokemon Black & White (DS) — Actual; Prediction (B=1.05)

Bibliography and references

[1] http://vgsales.wikia.com/wiki/Video_game_industry

[2] http://www.vgchartz.com/

[3] http://www.metacritic.com/

   http://www.metacritic.com/about-metacritic

   http://www.metacritic.com/about-metascores

[4] Video Gamers in Europe – 2008 by Nielsen Games

   http://knihovnam.nkp.cz/docs/ISFE_Consumer_Research_2008_Report_final.pdf

[5] http://www.theesa.com/facts/pdfs/ESA_EF_2011.pdf

[6] http://www.eurogamer.net/articles/2011-07-28-eu-3ds-price-cut-around-a-third

[7] British Board of Film Classification

   http://en.wikipedia.org/wiki/British_Board_of_Film_Classification

   http://www.bbfc.co.uk/

[8] Pan European Game Information

   http://en.wikipedia.org/wiki/Pan_European_Game_Information

   http://www.pegi.info/

[9] A Marketing Manager's View on Successful High Tech Launch Strategies (E.J. Hultink, J.P.I. Schoormans)

[10] Marketing Management 12th ed. (P. Kotler, K.L. Keller)

[11] Launch Strategies and Product Advantage (E.J. Hultink, S. Hart) 2006

[12] Principles of Marketing 13th ed. (P. Kotler, G. Armstrong) 2010

[13] Marketing Management in New-Product Development in the Finnish Online Game Industry (P. Lehdonvirta) 2009

[14] SALES FORECASTING

   Published in The IEBM Encyclopedia of Marketing, Michael J. Baker (Ed.),

   International Thompson Business Press, 1999, p. 278-290.