**How Are You Doing?**

This Word document file contains the text for this book, How Are You Doing? The book is in PowerPoint, and at the bottom of each page there are Notes that describe what each page is about. This Word document is a compilation of these notes – all in one place, in case you find that easier to work with.

**Table of Contents**

|  |  |
| --- | --- |
|  |  |

 **Introduction**  3

**Part I Normalizing the Data**  6

 Chapter 1 Equated Day Factors (EDFs) 7

 Chapter 2 Holiday Factors 16

 Chapter 3 Normalization Factors 32

**Part II Seasonally-Adjusting the Data**  38

 Chapter 4 Initial Seasonal Factors 38

 Chapter 5 Holiday Factors 44

 Chapter 6 Normalization Factors 56

**Part III Trending Daily Data**  69

 Chapter 7 Day of the Month (DOM) Factors 69

 Chapter 8 Estimating Daily Trend 82

**Part IV Applications**  104

 Chapter 9 Equated Day Factors (EDFs) 104

 Chapter 10 Estimating Price Elasticity 107

 Chapter 11 Monthly Retention (& Attrition) 117

 Chapter 12 Forecasting 132

 **Overview**  139

**How Are You Doing?**

So, “how are you doing?”. That may be the most common question in the English language. It also gets to the heart of what much reporting and analysis aims to achieve. My name is Peter Gascoyne; I’ve been a consultant for about 30 years, working with a wide variety of industries and areas, private and public, trying to help them better understand their data and performance. That understanding can be greatly enhanced when data is seasonally-adjusted.

**ii Introduction**

In this introduction to “How Are You Doing?”, I will give a short overview of what this “book” covers, and some guidance on working with this “book”. I’ve put “book” in quotation marks as, of course, it is a PowerPoint presentation. But at over 500 pages, plus an Overview, it seems deserving of the label. I chose to present this in PowerPoint format simply because most of it involves charts & tables that best lend themselves to PowerPoint, rather than a more Standard format.

Please don’t be put off by this book’s length. As you’ll see when you look at the chapters, every page has at most 2 lines of large print text at the top. Fuller explanation of each page can be found in the “Notes” at the very bottom of the page. If you prefer, all these notes can also be found separately in the Word file called “BookText”.

**iii The book draws upon a few data sources for demonstration.**

In order to demonstrate the points in this book, there are two main examples I’ve drawn upon. The first was a history of the New York Stock Exchange (NYSE) sales volumes that I was able to download from the web. That history proved very useful for showing how to: normalize the data; develop seasonal factors; estimate trend; develop day of the month (DOM) factors, and trend data on a daily basis.

The other primary data source I completely made up – which is harder than you might think. This data was used for the hypothetical firm, XYZ, and helped in demonstrating trending, estimating retention and attrition, as well as forecasting.

I also drew upon Wisconsin home sales to quickly explain some points on developing the initial seasonal factors, in Chapter 4.

Of course, your own data will likely behave quite differently to the examples presented here. It would be virtually impossible to try to anticipate all the ways that all the different types of data may behave over time. Hopefully however, you’ll find enough range of example to understand and achieve an appropriate way of treating your own data.

**iv In this book, the location of the displayed table or chart is identified.**

This book relies heavily on models and tables and charts that are found in Excel files linked to the MakingApples website where I’ve placed this book. As appropriate, you’ll find the tables and charts reference these Excel files using the following format, as demonstrated in the example shown here. The reference will always be in the bottom left corner. The text will have two parts. The first part contains the Excel file name, then a colon “:”. The second part is the name of the specific tab where it is found. Thus, in this example, the table is found in the “OrigData” tab of the “NYSE Daily Data” file.

 **v The book is divided into four parts, with each chapter drawing upon the Excel files listed below.**

This book is divided into 4 parts. Each chapter draws upon the Excel files listed in these Table of Contents.

**vi Part I of the book describes how to normalize the data, to make the adjustments such that every month is of equal “length”.**

In Part I, we walk through normalizing the data, which is the process of adjusting monthly data so that every month is of approximately equal length. This is accomplished by developing Equated Day Factors (EDFs, described in Chapter 1) that compare the “weight” of each day of the week relative to a daily average. In Chapter 2, holiday factors are developed which capture the degree to which holidays, and the days immediately preceding & following them, differ from their “normal” level of activity for their given day of the week and time of year. EDFs and holiday factors are combined in Chapter 3 to derive normalization factors, which are used to adjust the monthly data so all the data – in past, and in future – is now presented as though the months are of equal length. Ideally, you have daily data to use to develop these normalization factors. Even if you do not collect daily data, I urge you to walk through this procedure and develop crude estimates of your EDFs & holiday factors, so that you can proceed to seasonalizing with a normalized set of data, however rough it may be.

**vii Part II walks through seasonally-adjusting the data, by developing seasonal factors and estimating trend.**

In Part II, seasonally-adjusting the data is explained, a process which takes the normalized data and adjusts for seasonality, the typical cyclicality across the 12 calendar months of the year. An initial set of seasonal factors are calculated in Chapter 4. Chapter 5 walks through the construction of the trend model that adjusts the history for growth & events, in order to arrive at a “final” set of seasonal factors. Chapter 6 walks through estimating trend, capturing your best estimate of how the data trends over time using manually input estimates of the growth rates and events.

**viii Part III describes trending daily data, which applies only to those organizations that collect and track data on a daily basis.**

Part III describes how to trend daily data. If your organization doesn’t capture daily data, you can skip this section. Chapter 7 describes how to develop Day of the Month (DOM) factors that reflect the pattern across the month. Chapter 8 then goes through the process of trending daily data across time.

**ix Part IV walks through some of the key applications to draw from the trended, seasonally-adjusted data.**

Finally, in Part IV, we walk through some of the more valuable applications of your trended data. Chapter 9 describes Reporting – not so much what it looks like, as there are so many other sources available that address that, but what information to present that succinctly captures how you’re doing. Chapter 10 shows how you can estimate your price elasticity, a valuable metric that captures how much you can expect sales to fall (or rise) given a proposed price increase (or decrease). Chapter 11 walks through estimating retention and attrition (or churn), emphasizing using a monthly measure rather than an annual. Finally, in Chapter 12, we walk through forecasting, a process we’ll find is quite simple & quick to do once you’ve trended, and learned from, your history.

**Part I: Normalizing the Data**

Pg. **1-1** In the “Overview” presentation, we looked briefly at seasonally-adjusting data as a powerful tool for understanding your performance: past, present, & future. And we looked at some of the issues with the more traditional methods of trending data and evaluating how you are doing.

We’re now going to get into the details of how you seasonally-adjust and trend your data. These details will be divided into two broad parts: in this first part we’ll walk through normalizing the data, the process of adjusting monthly data so each month is of approximately equal length. In the second part, we’ll use the normalized data and develop seasonal factors that allow us to seasonally-adjust the data and have a much easier time of trending the data over time.

Again, as I emphasized in the “Overview”, be aware that all of this describes one approach for seasonally-adjusting your data. What I like about this approach is that it enables you to seasonally-adjust your data yourself: you’ll be in the driver’s seat, using fairly basic Excel formulas to perform the data adjustments. And you’ll understand why the various steps are being taken. Even if you use software that seasonally-adjusts the data for you, I would hope this will enable you to much better understand what that software is doing, or at least, should be doing.

**1-2 The seasonal-adjustment process is laid out below. Seasonal factors are obtained by normalizing the data, then adjusting history for growth & events.**

The seasonal-adjustment process requires a number of steps that are outlined here in this diagram. You’ll start with your original data and will first “normalize” that data so every month is of approximately equal length. To do that, you will need to more accurately quantify the length of each month, which is done by calculating “equated day factors” and “holiday factors”. The equated day factors estimate how long each day of the 7-day week is, relative to the daily average. Holiday factors capture how much quieter or busier activity is on holidays and the days immediately preceding & following them. Combining these two sets of factors arrives at the “normalization factors” that are used to normalize the data.

Next, a simple initial set of seasonal factors is quickly calculated, using the normalized data. These initial seasonal factors are used to derive an initial set of seasonally-adjusted history. That history is then trended. The trend estimates describe the growth rates and events over time. The history is adjusted for this historic growth & events, and a final set of seasonal factors obtained. When the final set of seasonal factors and normalization factors are applied to the original data, a final seasonally-adjusted history will be the result.

In this part I, we walk through the process of normalizing the data.

**Chapter 1: Normalizing the Data – Equated Day Factors (EDFs)**

**1-3** So, let’s begin by estimating the equated day factors…

**1-4 Equated Day Factors (EDFs) are an index of the relative weight of activity for each day of the 7-day week.**

The Equated Day Factors (or EDFs) describe how busy each day of the week is relative to the daily average. The factors are represented as an index, where 1.00 is the daily average, and each day is characterized as being some percentage above or below that average. Why are these important? Because each month, except for February, has 30 or 31 days, where 2-3 days of the week occur 5 times during the month, while all the other days occur only four. That extra day may be quite significant, depending upon how much the pattern varies across the week. If you are especially busy on Monday, then all else being equal, having 5 Mondays one month and 4 Mondays the next, that 1st month may be 2-3% longer or more, simply due to this one aspect of the “calendar effect”.

The EDF chart here is for the NYSE daily sales volumes from 2006 thru late 2016. Activity is about 10% below average on Monday. The sum of the factors totals exactly 5.00 (yes, the numbers here add to 4.99 but that’s due to rounding). EDFs will always add to exactly the number of days of the week for which there is activity, be it 5 days, or 7 days, or any other figure.

**1-5 Development of the EDFs involves seven steps.**

The process for estimating the EDFs involves 7 steps.

1. The data is collected.
2. The data is next sorted to enable easy calculation of the EDFs.
3. Next, an initial set of factors is calculated.
4. Then we use the standard deviations to weed out those weeks containing outliers we want to remove for purposes of calculating the “normal” EDFs.
5. The factors are then charted.
6. The chart helps in tweaking the number of standard deviations to be employed for arriving at the final set of EDFs for a given time period.
7. The last step involves repeating this exercise for other time periods, then reviewing the results to determine if there may be one or more overall averages that apply to the entire history.

**1-6 The data used to develop Equated Day Factors, and Holiday Factors, will ideally have the following attributes:**

So, the 1st step in calculating the EDFs is to collect the data. This may be one of the most important steps: not just for calculating the EDFs, but for analyzing and learning from your history overall.

I would strongly urge you to go as far back in time as possible. You’ll want the data to be more or less reliable. And you won’t want values being so small that there is an inordinate amount of volatility, as can be the case when working with smaller numbers.

Even though your organization may be quite different today as compared to time past, there may still be key pieces of information you can glean from your past that could be beneficial – the impact of unusual natural or political events; estimating your price elasticity; determining what kinds of practices have been more or less successful historically, etc. Most of the analytical work is in setting up the calculations. Whether you do so for two years or for twenty will not make that much difference. Again, you may find some interesting info from your past that would have gone neglected if you took a shortcut and failed to go back very far in time. And by going back much further in time you can derive more accurate measures of your holiday factors – to be examined in the next chapter. Ideally, you can go back about 25 years. Going back that far will enable you to get at least 3 sets of data for each of the 7 days of the week that a fixed date holiday can fall. This can be quite valuable for analyzing the most difficult and generally most important holiday of the year, Christmas.

**1-7 What if you have no daily data – do you just skip this procedure? NO.**

What if you have no daily data? What if all your history is presented in monthly form only? You still should endeavor to estimate your EDFs, for you still want to express all monthly data in roughly equal-length months. To do this you will want to fill in the form here, inserting rough approximate estimates of how relatively busy each day of the week is. I would recommend sitting down with a couple of other people in your organization and walking through how busy each day is. Ideally you can gather folks from the appropriate area: sales managers for sales data, accountants for revenue data, and so on.

Sit down and come up with a number representing roughly how busy you are on the “average” Monday: maybe you express it as 100, or perhaps it’s a rough guess of the actual figure, say 12,400, or whatever. OK, if Monday is 100, or 12,400, how busy is Tuesday, then Wednesday, and so on, through the rest of the week. Then add up the total for the week and express each day as a percentage of the daily average. If you’re open all 7 days, and you have Monday at 100 and the week total is 750, then Monday’s factor would be 7 x 100 / 750 = 0.93. I would round each number to the nearest 1/100th, for these are crude estimates at best – just make sure the factors total to the given number of days that you have activity.

By the way, you will then want to repeat this exercise for holiday factors – to be described later.

**1-8 The dataset used here as example will be the 1991-2016 New York Stock Exchange (NYSE) sales volume data.**

For the purposes of demonstrating how to estimate EDFs specifically, and how to seasonally-adjust data generally, I have chosen to use the data showing sales volumes for the NYSE. I chose this data for a number of reasons: it’s publicly available, so I didn’t have to go through what can be a difficult exercise of inventing data, or using some organization’s specific data and “altering” it in order to hide information they want kept confidential. Other advantages of the NYSE data are that it’s daily data so we can use it for demonstrating the estimation of equated day factors & holiday factors. It goes back all the way to 1900 so we have plenty of history to draw upon – don’t worry, I only am using data going back to 1991. Because we’re dealing with data that is so broadly known as the NYSE, another advantage here is that one doesn’t have to be an “insider” to evaluate and understand at least some of what might be driving the data behavior we observe. Finally, I should point out that we will be using volume data, not closing index amounts. It’s doubtful that we would find much difference in prices across the week, or at holidays, etc. If we did, the whole world would soon know about it and the pattern would quickly vanish.

**1-9 To sort the data so it can be easily analyzed for developing EDFs, we want the data in columns, where each column is one of the 7 days of the week.**

Once the data is pulled, we want to sort the data into columns, with each column being a different day of the 7-day week.

**1-10 Data sorting starts with the original data: this set is downloaded from NYXData website.**

We start the data sorting process by looking at the original dataset. Three things to point out here. 1st, there is data for “group shares”, “group trades”, and “group dollar volume”. This is the format the data comes in when you download it from the NYXData website. We will focus on just the group shares, in Column B.

The second thing to note is that many days are missing. The exchange was closed on January 1st, for the holiday, and on Jan 5th & 6th for the weekend, and so on. This is how data will be: there will likely always be some days that are “absent” from the dataset. Nonetheless, we will want to fill those absent days with a “0” so that we can easily apply formulas that will sort the data properly for us.

**1-11 Note the NYSE volumes have many days “missing”, when the exchange was closed.**

The second thing to note with this data is that many days are missing. The exchange was closed on January 1st, for the holiday, and on Jan 5th & 6th for the weekend, and so on. This is how data will be: there will likely always be some days that are “absent” from the dataset. Nonetheless, we will want to fill those absent days with a “0” so that we can easily apply formulas that will sort the data properly for us.

**1-12 One starts with the original data: the NYSE volumes have many days “missing”, when the exchange was closed.**

Finally, I want to point out here how I label where the given table or chart is found amongst the files located on the “MakingApples” website. The labels will always have two parts: the 1st part will be the file name, the 2nd part will be the tab. Thus, this Original Data example is drawn from the “NYSE Daily Data” file, under the “OrigData” tab.

**1-13 The next step in sorting the data is to have placeholders for every day of the year, & to identify the day of the week, and the week & month of the year.**

The next data sorting step is performed on a different tab, to make it cleaner and easier to keep straight. Here I have listed in Column A every day for the entire period being examined (in this instance, from Dec 31, 1990 thru the end of 1995). I picked up Dec 31 because it falls on the Monday, and I want to sort all daily data into weeks, with the weeks described in terms of the date for the Monday starting their week. I’ve manually entered “Mon” thru “Sun” in Column B, cells B11-B17 to be exact. Cell B18 simply picks up the value from cell B11. That formula is then copied for the rest of the 5 year period – it’s that simple. Similarly, in Column C, cell C11 picks up the date (here it’s Dec 31, 1990). C12 thru C17 picks up that same date, for we want every day of the week to be described as falling in the week of Dec 31, 1990. Then, in Cell C18, we simply put the formula “+C11+7”, which gives us the Jan 7, 1991 date. Again, we then simply copy that formula all the way down for the rest of the 5-year period. Column D has a formula to identify the calendar month that the given date falls in; this will enable us to later easily calculate the totals by month if we want to.

Finally, Column E uses an “Index/Match” formula to pick up the values from the “OrigData” tab. This formula construction enables us to ensure we get “0” values on those days where no activity occurred.

**1-14 The final data sorting step uses Excel “array formulas” to set out all the data by week, with separate columns for each of the 7 days of the week.**

The final sorting step, performed on a different tab, is to use “array formulas” to pull in the data by week, and by Day of Week. If you’ve never used array formulas before, I highly recommend you’re doing a little research on how they work. They are very powerful, quite commonly employed, and fairly easy to master. They allow you to perform all manner of calculations in just one basic formula. Here I inserted a formula in Cell B11 that pulls in the volume data from the “DataByDay” tab, with the instructions that the “Week of” date matches the date in Col A, and the Day of the Week matches the days in Row 10. The formula is copied thru the rest of that 1st week, and then copied down thru all the remaining weeks, through 1995. A weekly total is calculated in Col I.

We’re now done with sorting the data and are ready to calculate the EDFs.

**1-15 Next, we index the data, calculating daily factors that represent how each day compares with the average for the week.**

Now that the data is all sorted properly, in a new file (called “EDF” here), we pull in the data from our “NYSE Daily Data” file. To keep the numbers smaller and easier to work with, I’ve expressed the numbers in millions, performed here by multiplying each day’s value by the “Unit” inserted in Cell I1. The “Unit” here is 1/1 millionth, or 0.000001. Note that this cell is displayed with a bright yellow fill. Anywhere that has a yellow highlight represents a place where a manual entry may or may not be necessary; everything else in the spreadsheet is just formula – no additional work required.

The 1st date in the data section (Cell A11) is highlighted – the 1st Monday’s date is entered here; formulas take care of all the other weeks. The rest of this Daily Sales Volume section simply picks up the data from the other file. Section 2, the Daily Factors section, expresses each day’s volume as a percentage of the average daily volume for that week. For the outlined week of Jan 14, the week total was 906 million, for about a 181 million daily average. Monday’s volume of 120 million is about 66% of that average. The Daily Factors for every week always total the “Total Number of Days of the Week”, “5”, which is referenced by formula that picks up the value placed in the highlighted Cell R1.

**1-16 For this entire 5-year set of data (1991-95), we also calculate the simple average factors for each day of the week.**

At the top of section 2, a set of simple average factors are calculated; they are the average for the entire 5-year period, for each day of the week. Monday comes in quite low, at 0.84.

**1-17 What do all these daily factors look like? The chart shows most are in a fairly tight range, but there are numerous outliers.**

So what do all these factors look like? The chart shows that most of the data comes within a fairly narrow range. Most of the values for Monday, for example, come in between 0.60 and 1.20, though there are numerous occasions where the values are below or above that range. But there are many “outliers”, values that are well outside the “norm”. The most obvious outliers are where the factor is “0’, those days where the NYSE was obviously closed. Many of the closed days fall on Monday, not surprising as many holidays (Memorial Day, Labor Day, Presidents Day), fall on a Monday (ML King Day was not established as a holiday until 1998).

There are also a number of outliers on the high side. We don’t want to pick up those weeks where there are outliers – not only is the value unusually low or high for the day of the outlier, but then the other days of the week will be correspondingly higher or lower than normal because of the outlier day(s). Is there an easy way to rid ourselves of these outliers?

**1-18 How do we get rid of these outliers? We will want to keep data within a range, as expressed in terms of standard deviations.**

Use of statistical formulas is kept to a minimum with the seasonal adjustment methodology that I’ve developed. I have intentionally kept formulas simple to enable one to easily understand what’s going on. The main exception to this rule here will be the use of standard deviations.

The standard deviation is a measure of how volatile the underlying data is; it captures how much the data varies from the mean, or average. The more widespread the data, the higher the standard deviation. In our dataset, Monday appears to be more diverse than Wednesday for example, so we will expect it to have a higher standard deviation.

How do we use standard deviations to whittle down our data and remove the outlier weeks? By evaluating each data point to determine if it falls within a given number of standard deviations from the mean.

**1-19 We let Excel calculate the standard deviations for each day, over the 5 years.**

At the top of the Daily Factors section, we’ve inserted a formula for each day of the week which calculates the standard deviation. As expected, Monday’s standard deviation (0.28) is higher than Wednesday’s: 0.17.

**1-20 The next step is crucial: for each day, we calculate how many standard deviations from the mean they are.**

The next step is key: we calculate how many standard deviations each day’s value is away from its mean. In the highlighted example, Monday, Dec 31, 1990’s daily factor is 1.09. 1.09 is 0.25 higher than Mondays’ mean of 0.84. 0.25 is 90% (0.25/0.28) of the Monday’s standard deviation of 0.28: “0.9” is shown in the highlighted Cell T11. The same calculation is performed for all the days of the period.

Note the colorful conditional formatting in this Section 3. I wanted to bring attention to how close or far each day is from its mean. And I have done so by using the criteria that is set up at the top of this section. Days that have low standard deviations, defined here as being anything less than 1, are highlighted in green. Days with a very high standard deviation count were defined as being above 2, and are highlighted in orange. Days that fall between 1.5 and 2.0 standard deviations are highlighted in pink, while those between 1.0 and 1.5 have no fill. Just like that, we can quickly see how well “behaved” the data is each day and each week.

**1-21 We determine which weeks to allow, and which to eliminate, by setting a “Maximum Number of Standard Deviations” to allow.**

The next step determines whether or not we are going to “count” a given week. In highlighted Cell AC4, we have entered a value of “1.0”. This value represents the maximum number of standard deviations we will allow in order for a day and week to be counted. If the number of standard deviations from the mean for a given day and week are less than the 1.0, it will be counted; if not, it won’t.

**1-22 We then evaluate each day of each week, and we only count those weeks where every day came within the maximum standard deviations allowed.**

In Column AC, a formula is inserted, and copied down, that asks of each day of that week, whether or not the number of standard deviations from the mean is less than the maximum number permitted, as defined in Cell AC4. We can quickly see (thanks to the colorful formatting) that the only week meeting this criteria is the week of Jan 21, 1991. It’s is the only week where every value in Section 3 is “green”, is less than the 1.0 maximum. Accordingly, that week has a “1” in Column AC, the other weeks displayed here do not.

Notice that the count of the allowed weeks is summed at the top of Column AC, in Cell AC1: of the 261 total weeks over this 5-year time period, 185 met the standard deviation criteria, and were counted.

**1-23 Our Final set of Equated Day Factors pick up only the weeks that are counted, that meet the Maximum Standard Deviation criteria.**

Section 4. of the “EDF” tab picks up those weeks that meet the criteria. The highlighted week of Jan 21 picks up the Daily Factors from Section 2; the other weeks displayed here have zeroes. The “Final EDFs” section is filled by simply multiplying the Daily Factors from Section 2, by the “Count?” value in Column AC.

**1-24 So what does this look like? The chart displays the outcome where the weeks counted all have factors within 1.0 standard deviations of the mean.**

We earlier noted that 185 weeks met the 1.0 standard deviation criteria. What do those weeks look like? Our chart here shows us. The new set of Average Daily factors (or EDFs) is highlighted. Note that Monday’s average has increased a bit – from 0.84 to 0.89; getting rid of all those “0” values helped increase it. The “band” shows the range within which each day had to fall in order for that day and week to be counted.

**1-25 If we increase the Maximum Standard Deviation criteria to 2.0, many more weeks get counted.**

We originally set the maximum standard deviation criteria to 1.0. What happens if we increase that value? Here we have increased the value in AC4 from 1.0 to 2.0. As you can see, many more of the weeks are counted. Note that the total weeks counted has increased from 185 to 220 (Cell AC1).

**1-26 If we broaden the net, and allow up to 2.0 Standard Deviations, more weeks are picked up, with a notable concentration of higher values on Friday.**

The chart here shows the impact of broadening our net to 2 standard deviations. More weeks are picked up. Look at Friday: only a few lower values are picked up, but there are quite a few higher values on Friday. Not surprisingly, the average factor for Friday has increased – from 0.99 to 1.01.

**1-27 The Maximum Standard Deviations to allow is a judgement call, and does not have to be a whole number. Here we “compromise” at 1.5.**

There is nothing sacred about the maximum standard deviation value being a whole number. Indeed, I generally like to use “1.5” as a nice compromise value. Statistics will tell you that a standard deviation of “1” picks up about 68% of the values, while “2” standard deviations pick up 95%. 1.5 happens to pick up about 87%. Here we are applying this rule to each day of the week, and requiring all 5 days meet the criteria, so the numbers work out to be a bit less. But you get the idea.

The number of standard deviations that is used is very much a subjective decision. There is no one right value. I’ve generally found “1.5” works pretty well when I “eyeball” the data and decide where I want the axe ultimately to fall. This chart here for the EDFs looks pretty good. We’re not quite as top-heavy on Friday, so its average slightly drops back, to 1.00, still above the 0.99 we had with 1 standard deviation. The other days look pretty clean as well. Wednesday has a couple of values right by the max – but there are a few lower values to help offset it.

**1-28 A comparison of the resulting Equated Day Factors shows they are all very similar, except for the Simple Average.**

I’ve added a chart here to show how the different criteria work out. You don’t need to do this ordinarily; I’m doing this here to help show what kind of difference it makes to “play” with the maximum standard deviation criteria. As you can see, all 3 lines, using 1.0, 1.5, & 2.0 standard deviations, are all very similar. Friday sees the widest range of result; the other days hardly change at all.

The biggest difference is from our original factors, where all data was allowed. The low Monday value is notable, along with a correspondingly higher amount for Tuesday thru Thursday. Observing this comparison, I feel all the more comfortable settling with the 1.5 standard deviation criteria.

**1-29 Equated Day Factors are calculated for the other 5-year periods since 1991.**

We now calculate the EDFs for all the other 5-year periods. This is NOT a daunting task. I timed myself, and it took me literally 15 minutes to create the other 5 time periods and to tweak the standard deviation criteria. All you need to do is:

1. Copy the tab (EDF9195) and change the tab name.
2. Change the title in Cell A1.
3. Change the 1st Monday date in Cell A11.
4. Change the formula in Cell B11 so it picks up the correct amount for the 1st day of the covered period.
5. Copy B11 thru the rest of the period (basically to the Range B11:F271).
6. That’s it. Almost.

**1-30 “Tweaking” the maximum standard deviation value can help eliminate some dangling outliers.**

You should always check out what the data looks like. For instance, when you look at the chart for the 1996-2000 time period, you’ll note there are a few dangling outliers that are almost on their own. This is where “playing” with the standard deviation count can be useful. Again, this is very much a judgement call, but I like the outcome better when I use 1.3 as the maximum. The 4 highlighted outliers are removed by making this change. Note that the averages hardly change at all – in fact, rounded to the nearest hundredth they remained the same. You only see a change when rounding to the thousandth.

**1-31 Equated Day Factors are calculated for every 5-year period. The pattern is much flatter across the week in the recent 2011-15 period.**

As mentioned, EDFs are calculated here for every 5-year time period. As you can see, they did change a bit over time for the NYSE volume data, most notably in the 2011-15 timeframe.

How many different time periods should you check with your data? Again, this is something of a judgement call. I would try to at least have 1-2 years for a time period range, unless there is a distinct change in pattern going on and you want to narrow the time period to see if you can determine exactly when that change took place – which, by the way, should give you a helpful clue in answering the important question of why it changed.

(I’m not showing it here, but I did look more closely at the NYSE history and found the pattern seemed to start flattening around 2007 – that made sense, for that was about when the Great Recession began, and sales volumes increased enormously.)

I would recommend trying to split out your history into perhaps 3-5 time periods. This will help give you some sense of how the pattern changes over time.

**1-32 The general pattern has changed somewhat over the years, with the last decade seeing activity spread more evenly across the week.**

If you wanted, you could stop here. But I like to see if it’s appropriate to further summarize the history, to see if a very similar pattern applies to more than one period, and to summarize the broader time period accordingly. Again, this is a judgment call; it is very difficult to anticipate what types of pattern changes you may see over time.

But I noticed for the NYSE sales volumes, for the earlier history from 1991 thru 2005, there was a similar pattern, with Mondays generally quieter than later periods, and Wednesdays relatively busier. So I averaged those three time periods. I then averaged from 2006-forward, putting much less weight on the final period as it only had data for the 1st 9 months of 2016. Notably, the 2016 pattern is much more similar to the earlier 1991-2005 pattern. This would be worth keeping an eye on, for if it continues for another 6-12 months or more, it would merit capturing this final period separately.

Why call this out? Because the latest set of average factors will be the EDFs we will want to use when it comes time to forecast. But in the meantime, the 2006-Pres seems to represent a good set of EDFs for forecasting purposes.

This completes the estimation of the factors for the day of the week. Next, we turn to the holidays…

**Chapter 2: Normalizing the Data – Holiday Factors**

We next turn to holiday factors, estimating how much performance changes on holidays, and the days immediately preceding them & following them.

**2-2 Holiday Factors describe how the activity on holidays, and the days immediately surrounding them, compares with “normal”.**

Holiday factors, like the EDFs, are an index that compares activity against “normal”. Normal activity has an index of exactly 1.00. Holiday factors measure the degree to which the holiday and days surrounding are relatively quieter or busier than normal for that given day of the week for that time of year.

As example, with Memorial Day we can see that the stock exchange is closed on the holiday itself, that it is about 25% quieter than usual the prior Friday, and that it sees a slight activity lift the day after the holiday, with a slight drop-off the two days after that.

**2-3 Why bother?**

Developing holiday factors takes a bit of effort. However, I believe you will find the effort much reduced by using the templates provided here. Nonetheless, it’s not unreasonable to question why go through this bother. Ultimately, you should find this extra effort worthwhile for 3 key reasons.

First, you will gain a better understanding of how your monthly results can be impacted by the holidays. For example, as we shall soon see, July sales may be as much as 5% higher or lower simply due to what day of the week July 4th falls.

Second, you will be able to obtain more accurate estimates of the impact of certain internal programs, like marketing campaigns or process improvements, by reaching a more accurate read of how holidays impact your reported results.

Finally, the additional effort should give you greater confidence in the entire reporting and analysis you perform for your organization. The more predictable adjustments you can make to your data, the better you isolate what is going on underneath the data, better answer that primary question “how are you doing?”. And hopefully gaining that greater confidence rubs off with management having greater confidence in the information & analysis you provide them.

**2-4 What do you do if you have no daily data to work with? Your best.**

Before we begin walking through how holiday factors are derived, you might be asking yourself whether you need to go through this exercise if you don’t have daily data, or your data is very incomplete and of limited value. As we saw with the EDFs that compared the relative level of activity across the week, you’ll want to at least try to come up with some crude estimates of what happens in and around the holidays. Again I’d recommend sitting down with appropriate staff and talking through rough estimates of how behavior changes.

You might find seeing the NYSE holiday factors developed here of some use. You can note some of the patterns – like how the impact from July 4th is greatest when it falls midweek, as compared to when it falls on a Tuesday or Thursday, and so on. Certainly the behavior will likely be different for your organization, but hopefully the *patterns* you observe here may give you some clue or idea of what you experience.

**2-5 There are two categories of holidays to estimate: Fixed Day and Fixed Date.**

There are two categories of holidays that need to be estimated, and which require different treatment. Fixed DAY holidays always occur on a specific day of the week – like Monday for Memorial Day & Labor Day, or Thursday for Thanksgiving. Fixed DATE holidays always occur on the same calendar day of the year – like July 4th, Christmas, Halloween, and so on. We’ll take a look at these two separately, starting with the easier Fixed Day holidays.

**2-6 The table shows the Holiday Factor calculation for Martin Luther King Day, 2016.**

The table here shows the calculation of holiday factors for Martin Luther King Day, for 2016 alone. (On the same tab, to the right of this table, similar tables are developed for each of the other years from which data and calculations are drawn.)

We’re going to walk through each section of the worksheet so you understand how the holiday factor calculation works.

**2-7 The headings identify the holiday, day of week, and year.**

The top of the table simply has headings identifying the holiday, the day of the week it falls, and the date and year for which calculations are being made.

**2-8 The data is sorted in columns, with the holiday placed in the middle; this ensures picking up activity change for the 3 days *before* & *after* the holiday.**

Below the headings comes the data, with data sorted in columns by day of week. The day of the holiday is ALWAYS the middle column – this way when we bring in one full week of data, we can do so for the 3 days before, and the 3 days after, the holiday. Note how in Row 9, the actual number of days from the holiday (from -3 to +3) is placed – the prior Friday is at “-3” because it occurs 3 days before the ML King Monday holiday. These numbers will later be picked up by formulas seeking the sales data.

**2-9 Past holiday dates are manually entered in Column A.**

Column A is where the dates of the ML King holiday are placed for each of the last 9 years, with the most recent year first. Thus, in 2016 the holiday fell on January 18 (Cell G11), while it was on the 21st of January in 2013 (Cell G14).

This model is set up to estimate holiday factors based on 9 years of historical data. While there is no real magic to this figure, it should represent sufficient data points to derive a reasonably accurate estimate of the holiday factors; and it does not involve going back too many years – at least not for fixed DAY holidays.

You might be wondering how you determine the date in past that a given holiday fell. For the major holidays affecting the NYSE, those past dates will be found in the “Holidays – Fixed Day (& Fixed Date)” files. But what about for other holidays like Veteran’s Day? You will find a “Calendar” file on the website that runs from 1990 to 2020 and should help you determine the exact historical holiday date each year you run. If that doesn’t do the trick, I’d Google the holiday and you’ll doubtless find a listing somewhere.

**2-10 The “Week of” column identifies the date for each of the 4 weeks preceding & following the holiday.**

When we estimate the impact of a holiday, we want to compare the level of activity for the holiday, or the days immediately surrounding it, with the weeks before and after the holiday. We want to keep comparisons to a time period that is close to the holiday so we lessen the problem of seasonality. That is, if we’re looking at ML King Day for example, we don’t want to compare January activity with what is going on in March or April or later because by then normal activity may be significantly different to what occurs in and around the month of January. Seasonality will still always be somewhat of an issue here, but the problem can be lessened by confining our comparisons to just the 4 weeks preceding & following the holiday week. The “Week of” column here, column B, picks up the date of the holiday itself in the middle week – highlighted in blue - and then calculates the date of the weeks before & after by adding or subtracting “7”.

Note that the “Week of” date here is using the date of the holiday itself, not the Friday that is 3 days prior to the Monday holiday.

**2-11 The “Week of” column and the “Days from Holiday” row are used to pick up the appropriate data from the NYSE Daily Data file.**

Array formulas are used to populate the actual data for the entire period. The formulas know what day the data should be by picking up the “Week of” date from Col. B at the beginning of the row, and adding or subtracting the “Days from Holiday” value up in Row 9. So the highlighted Friday value of over 3 billion, is for December 18, 2015, calculated as the week of date of 12/21/15 (Cell B11) plus the “-3” from the Days from Holiday row (Cell C9).

**2-12 The “Include Week” column is used to manually identify if and when there is a week you do not want to count.**

The “Include Week” column is a manual entry. A “0” is inserted here for anytime you encounter a week that you don’t want to have included in the calculations. Here the weeks of Christmas and New Years are both excluded by inserting a “0” in Column J. We don’t want to bring in the data for those weeks because performance would have been so different in and around this major holiday.

We also never want to count the holiday week itself, for we are interested in how it compares with the OTHER weeks surrounding.

I considered dropping the week of Presidents’ Day at the bottom as well, again for fear of being skewed by different behavior that week. However, looking at that data, it doesn’t appear that activity differs much that week from the level of activity of the other preceding weeks. Also, I will later find that Presidents’ Day indeed elicits little change in performance, so I am safe to leave that week in, a good thing given that we’ve already taken out the two weeks of Christmas.

**2-13 The “Holiday Indices” section calculates the degree to which activity in the prior & following weeks differed from the activity during the holiday week.**

Okay, now we can calculate the “Holiday Indices”, how the holiday value compares with the weeks surrounding it. Looking at the highlighted example, we compare here how the Tuesday after the holiday compares with the Tuesday two weeks before. The Tuesday value of 1.571 billion is compared with, or divided by, the 1.135 billion volume from two weeks before. The ratio, 1.571 over 1.135 comes out as “1.38”: the Tuesday after the holiday is 38% busier than the Tuesday two weeks prior. Note that the holiday value here is being compared with the value two weeks before, NOT the other way around – to determine how the holiday compares with normal activity, we need the holiday value in the numerator.

**2-14 Looking at the Tuesday following the holiday, we see how that day’s volume (1.571 billion) compares with the volume on other “nearby” Tuesdays.**

We can see this holiday Tuesday is consistently higher than the other Tuesdays.

**2-15 Notice how the factors are so much higher compared to the Christmas weeks; that’s why we excluded those weeks.**

Not just Tuesday, but all the other days of the week behave quite differently during the two weeks of Christmas – which is why we excluded those weeks by entering a “0” in the “Include Week?” column.

**2-16 The “Determining Outliers” section identifies outliers. Just as was done with EDF’s, standard deviations are used to exclude unwanted data.**

Now we get to the trickier part – determining outliers, the activity we do NOT want to pick up when calculating holiday factors because the values differ substantially from “normal”. As we did with the EDFs, we will calculate the number of standard deviations from the mean for each value. If the difference is too high, the value will be excluded. The standard deviation for each of the 7 days of the holiday week is calculated in Row 43.

However, note a key difference here from the way calculations were done for the EDFs: there we excluded the entire week when any given day was found to be an outlier. Here we will only exclude the outlier day itself.

**2-17 We again establish the number of acceptable standard deviations, and employ ranges for conditionally formatting the output.**

As before with EDFs, we manually input a “Maximum Allowed Standard Deviations”, again using 1.5. We also employ conditional formatting to highlight how well or poorly each value compares with the mean.

**2-18 The “Prelim Avg” line calculates the simple average of the Holiday Indices *for the weeks to be included*.**

What is the “Mean” that will be used for calculating how much a given value differs from it? The “Mean” is the simple average for the weeks we are including. Recall that we wanted to exclude the two weeks of Christmas. Thus, the Preliminary average for Tuesday will be the average for the two weeks before & the 4 weeks following that “holiday”. Why are the excluded weeks excluded from the Preliminary Average? Because they would likely cause the standard deviation to be much higher, and we are interested in comparing the holiday days with only those weeks that we anticipate will represent an appropriate “normal” for that day at that time of year; the excluded weeks do not meet that criteria.

**2-19 We compare the Tuesday following the holiday with the Tuesday two weeks prior. That week’s factor (1.38) is almost two times (1.9) the standard deviation (0.14) from the mean (1.12).**

We return to the Tuesday two weeks prior to the “holiday” Tuesday. You can see it is quite a bit higher than the other values. Given a simple average of 1.12 (Cell G41), a standard deviation of 0.14 (Cell G43), its 1.38 value (Cell G33) is 1.9 standard deviations from the mean (Cell W33). Conditional formatting highlights this higher value in pink.

**2-20 So do we count that Tuesday? No, because the number of standard deviations from the mean (1.9) is above the maximum allowed (1.5).**

Given that it is more than our set Maximum Allowed Deviations of 1.5, this value is excluded from the calculation by being assigned a value of “0” (Cell O33), instead of “1”. Note that formulas are being used to make this determination; there is no manual input.

**2-21 We’re now ready to calculate the final factors for 2016. A SUMPRODUCT formula is applied that multiplies the indices by the Count value of 0 or 1.**

We now apply a SUMPRODUCT formula that picks up all the holiday indices, and all the Count, yes or no values, to arrive at a final weighted average. The Final Average of “1.06” (Cell G42) here is the average for the 5 days that were counted: one from the week before (Cell G34), and the other 4 from the 4 weeks following (Cells G36:G39).

**2-22 These are the final set of factors for 2016. Do they make sense? Wednesday’s value looks awfully high, especially as Tuesday was lower.**

We now have final values for 2016 alone. Are they reasonable? It makes sense that the Tuesday following is a bit higher. But the Wednesday factor of 1.30 looks awfully high. We will want to later see if that is typical, or is instead unique to 2016 alone. That’s why we ideally look at 9 different years of behavior – so there is a way to exclude outlier values, such as potentially this Wednesday, mechanically.

**2-23 The model is set up to repeat the factor calculations for each of the 9 years of holiday identified in Column A.**

The model is set up to perform the same set of calculations for the prior 8 years as was done for 2016, the most recent year available. Each of the other 8 years picks up the dates that were manually input into Column A. It may be difficult to see from the figure, but each year’s table of calculations is offset by exactly 26 columns to the right. Thus, while the 2016 data for Friday was found in Column C, you will find the 2015 data in Column AC, the 2014 data in Column BC, and so on, through the 2008 data in Column HC.

**2-24 Weighted average values are calculated for the 9 years of factors we have to work with.**

We are now ready to calculate a weighted average set of holiday factors, using the 9 years of data we’ve gathered. Let’s again walk through the different parts of the calculation.

**2-25 The weighted average factors are pulled in from each of the 9 years that were evaluated.**

The “Annual Holiday Values” section picks up the weighted average values for each of the 9 years. Notice how the 2016 values in Row 51 are the same as we earlier saw when we walked through calculating that year’s set of holiday factors.

**2-26 As before, a simple average and standard deviation are calculated for the 9-year dataset.**

Next, simple averages and standard deviations are calculated for the 9-year dataset. We don’t need to worry yet about exclusions because there is no obvious year we want to count out – we will wait for the calculations to do that for us.

**2-27 Weights are assigned to each of the 9 years, with consideration given to assigning less weight to older years.**

Weights are assigned to each of the 9 years. In a fairly arbitrary fashion, I decided to give somewhat less weight to the older years, so the 1st 5 years are fully counted, and the prior four will get slightly smaller weights, from 90% down to 60%. If you had reason to distrust or little value older years, you may want to assign even lower weights to these older years.

**2-28 Again, the standard deviation from the mean is employed to determine whether to count a given value.**

We again calculate how many standard deviations from the mean each value is, highlighting the values that are close, and far, from the mean. And we again use 1.5 as the “Maximum Allowed Standard Deviations”. Notice how that apparently unusually high value of 1.30 for the Wednesday following the holiday in 2016 does indeed pass the review; in fact, at only 0.7 standard deviations from the mean, it easily met the criteria.

**2-29 Note the weights applied to values that are counted: they are the same amount as the weight assigned in Col. A.**

Values that are within the Maximum Allowed Standard Deviations are counted. But note how the count value is less than 1 for those years that were given less weight in Column A.

**2-30 SUMPRODUCT formulas combine the Annual Holiday Values and the Count values to obtain our Final Weighted Average values.**

And again, SUMPRODUCT formulas pick up the Annual Holiday Values, and the Count Values, to arrive at a set of Final Weighted Averages for the holiday and the 3 days before and after.

**2-31 But do these final weighted average values make sense? It appears this is just a relatively busy week, perhaps because it’s at the start of the year.**

We again want to ask of the results, “do they make sense?”. The higher value for the preceding Friday seems plausible. Perhaps there’s a rush to get in purchases before the holiday weekend. And a busier Tuesday makes sense. But the Wednesday & Thursday values still seem a bit high. Evidently this is just a busier week, perhaps associated with it being the beginning of the year.

**2-32 The other fixed day holidays are run. Below is the look at Thanksgiving in 2015, identified here as a Friday (!) holiday.**

Other Fixed Day holidays are next examined. Thanksgiving is shown here. Notice that it is identified as a “Friday” holiday, not “Thursday”. I did that for two reasons. The main reason was that I wanted to easily capture the following Monday, for I expected it to be busier than usual after what is for most folks a 4-day weekend. I also did it this way because I separately took a look at the Monday prior to be sure it wasn’t significantly different from the norm. If it had been I would definitely want to pick it up. How? Through the somewhat tedious approach of treating Thanksgiving as a Thursday holiday AS WELL AS a Friday holiday, and then bringing the two separate pieces together when the holiday factors are all summarized. Fortunately, Monday was close to a 1.00 factor, so I felt safe to not give it special treatment, and be given its own holiday factor.

**2-33 Let’s look at the changes made for this holiday. First, the title is modified, as well as the day of the week.**

Let’s look at the changes we make to the model to handle Thanksgiving. These are worth walking through so you know what changes you’d need to make if you wanted to add a holiday not shown here in these files. As you’ll see, it’s pretty simple. First, we input the holiday name and the day of the week it falls at the top.

**2-34 Next, the critical holiday dates are input. Note that the dates here correspond to the date of the holiday, still treating it as a Friday holiday.**

Next we input the critical holiday dates. Again, be sure you are entering the date of the holiday itself. Here, the dates are for the Thanksgiving Friday as we are treating it as a Friday holiday to ensure the following Monday gets calculated.

**2-35 Next, the data is reviewed to see if any weeks should be excluded, as is done here for Christmas.**

Then we look at the weeks before & after the holiday to see if there are any weeks that should obviously be excluded. Not surprisingly, we found it necessary here to exclude Christmas week, by manually inserting a “0” in Cell J19.

**2-36 One important change may be necessary, depending on whether the holiday falls on a different day of the week: day of week headings may need change.**

Note that sometimes you will need to make one important set of changes. You may need to modify the headings for the day of the week, in Range C10 thru I10. Here the headings have been changed, making sure that the holiday is in the middle. We’ve identified Thanksgiving as a “Friday” holiday and here make sure the heading is lined up accordingly.

**2-37 The seven standard changes made for Thanksgiving are the same as you would make if you need to add any other holiday.**

The 7 standard changes highlighted here for Thanksgiving are the same as you should expect to make for any other holiday: the holiday name and day of week, the holiday dates, the days of the week headings, the exclusion weeks, the tab name, and the Maximum Allowed Standard Deviations. That’s it, that is all one needs to do to set up a new holiday. The only thing remaining is to review the reasonableness of the results to determine whether or not any additional tweaking is called for.

**2-38 The output is much as one would expect, with Thursday’s holiday factor at 0, the Friday value very low, and the following Monday well above 1.0.**

Let’s take a quick look at the results for some of the holidays that were examined, starting with the Thanksgiving holiday we’ve been tracking. Looking at the results here, they look very much as one might expect. Thanksgiving Thursday is of course at “0”; Friday is very quiet, as one would expect, while Wednesday is also quieter though not as quiet as Friday. Meanwhile, the following Monday indeed sees the heightened activity we’d expect to find.

**2-39 Memorial Day is an example of why you must review the Annual Holiday Values to ensure they are all reasonable.**

Memorial Day has some issues. Look closely at the Annual Holiday Values, and see if you can guess what they are.

**2-40 Some of the values are much higher. Why? Because they fall on the last business day of the month, & activity is always higher that day, every month.**

Some of the values are much higher than most. Why? Because Memorial Day comes near the end of the month. All these higher values correspond with their falling on May 31. And, not surprisingly, activity on the last business day of every month is typically well above normal. So we’re seeing the impact of the last day of the month.

We do NOT want to confuse that with what is being measured here, how the days following Memorial Day are busier or quieter simply because they come immediately after Memorial Day. That they sometimes correspond with the last day of the month is a SEPARATE matter.

**2-41 Deleting these cells not only ensures they won’t be counted, but also helps keep down the standard deviation.**

How to deal with these special days? Here I have chosen to DELETE the cell formulas. I’m doing this, not only because I want to make sure they are not counted, but also because I want to keep down the standard deviation, so we maintain a “tighter” fit with the remaining data. Previously, Tuesday and Thursday had relatively high standard deviations, of 0.25 and 0.19. Now they are down to a much more desirable level of just 0.07 and 0.08.

Lowering the standard deviation ensures a better “containment” of outlier values. Warning: the holiday template files do NOT have any cells deleted – you will need to look at your output and determine if there are any particular days you wish to exclude by deleting the cell formulas, as was done here.

**2-42 Columbus Day needed to have the 2008 values removed. They were all well above the other years’ values, thanks to that year’s stock market crisis.**

Columbus Day was a problem back in 2008. That of course was the year of the stock market crisis. We can imagine the highly volatile days following the collapse of Lehman Brothers would lead to unusual holiday values for Columbus Day that year. Accordingly, all the data for the week is deleted, and the standard deviations fall. At least, they fell for Friday, Monday, & Tuesday. Wednesday & Thursday hardly change, so perhaps we could have left those days alone but I think it more consistent and appropriate to just exclude that entire year of data.

**2-43 Thanksgiving also saw a high value, here on the Monday following the holiday in 2015. But it was not counted, by formula.**

Having observed how Memorial Day had some very high values when one of the days following the holiday was the last day of the month, I want to take another quick look at Thanksgiving. Remember how we had a pretty high value for the following Monday in 2015? Sure enough, that proved to be an outlier, likely consistent with that being the last day of the month. I could have deleted that value, but I left it alone as that value did get excluded since it was more than 1.5 standard deviations from the mean, and because the number of standard deviations for Monday, 0.13, was very similar to the number for the other days of the week. But it would also have been fine to have deleted that Annual Holiday Value.

When faced with any decision like this – do you delete a piece of data, what years or weeks do you exclude, what do you set the Maximum Number of Standard Deviations at, etc., try to resist anxiety over what the “right” way to do it may be. First, there is never an absolutely right way to do things, it is the nature of any analysis that there will be necessarily some judgment involved. The key thing to keep top of mind is what you are trying to accomplish. With holiday factors you are simply trying to determine how the level of activity on holidays and the days immediately surrounding them, compares with the normal level of activity for that day of the week for that time of year. Period. Just keep that in mind and you’ll end up close enough to what you’re after – the effect of the holiday on performance levels.

**2-44 All the Fixed Day holiday factors are picked up in a separate “Sum” tab; Final Results modify the values, where deemed appropriate.**

A “Sum” tab draws in all the final weighted average holiday factors for each of the Fixed Day holidays examined. “Initial Results” are our initial calculations. A place to modify them, if we see fit, is added and called here “Final Results” – these final results will be the ones picked up later when we normalize the data.

**2-45 The holidays are identified and the “dates” presented as the number of days before & after the holiday itself.**

Let’s quickly look at this table. The holiday names are identified in Row 10. The “dates” are identified in Column B, noted simply as the day of the holiday and the 1, 2, or 3 days before and after the holiday.

**2-46 Final Results modify those days where it doesn’t seem to make sense that there is any change due to the influence of the holiday.**

We now get into the realm of somewhat subjective decision-making. We could just leave all the numbers as is. And that should generally be fine. I decided here to make some very small modifications. Presidents’ Day had a value of approximately 1.00 for the Tuesday after the holiday, while Wednesday & Thursday come in at 1.01 and 0.97. They are all awfully close to 1.00, so I went ahead and manually input 1.00 in the Final Results section. It hardly makes a difference, but it just seemed slightly “cleaner” to use a 1.00 factor that implies there is no impact from the holiday. Similarly, I manually fixed the Thursdays following Labor Day & Columbus Day to 1.00. Whether I do this or not, really doesn’t matter. I just wanted you to see that it could be done, and that conditional formatting highlights any occasion that the initial result is changed.

**2-47 Generally the results make sense. ML King values are notably higher but may be due to start of year.**

Otherwise, the results generally look quite reasonable, and while they may sometimes be slightly higher or lower than one might expect, none are very striking. With the possible exception of ML King Day. Those values are quite a bit higher than the rest. And notice how the total for the week (Row 28, or 18) is higher than all the other days – except Columbus Day which DID see activity on the holiday, just lower. As noted earlier, perhaps this is a product of ML King Day coming early in the year. I don’t know, but given that it consistently sees higher activity across those days, I’ve chosen to just leave all it’s days alone.

**2-48 Before leaving Fixed Day holidays, I want to call out a special exception unique to the NYSE: the 3rd Friday of the last month of each quarter.**

Before leaving this section on Fixed Day holidays, we need to look at a special “holiday” that is unique to the NYSE, or perhaps the securities industry generally. Sales volumes are substantially higher on the 3rd Friday of the last month of each quarter (i.e. March, June, September, December). I call the “3rd Friday” a holiday because sales consistently behave differently on that occasion. What makes it even more unique is how dramatically it has changed over time. For years, sales were about 30% above average that day, but starting in 2008, they began to increase, and since 2012, sales volumes that day are almost 2.5x larger than normal for a Friday. The 3rd Friday in December is shown separately here because it occurs during the Christmas period, falling somewhere between Dec 15 & Dec 21 – I wanted to call it out to make sure it wasn’t behaving much differently than the other 3rd Fridays; and it hasn’t.

**2-49 Holiday Factors were assigned the 3rd Friday, with values increasing over time.**

We want to develop holiday factors that capture the bump in activity that occurs on these days, or “holidays”. When a “1-Year Moving Average” line is added to the chart, it becomes apparent how generally flat the behavior of these factors has been over time. A “Factors to Use” line was added to indicate the proposed holiday factor. Note how it is steady at 1.30 from 1993 thru 2007. It then bumps up to 1.50 in 2008, and up again to 2.50 in 2012. It looks like the factor has recently been falling fairly steadily, so that a lower value may be appropriate for future years, but for now I’ve kept it simple and decided to simply hold the factor steady at 2.50 from 2012-forward.

**2-50 An additional set of Holiday Factors were assigned the 4th Friday in June, with values increasing over time.**

To further complicate matters, I found the fourth Friday in June, and only June, also saw significant increases in sales volume. The pattern over the years is a little different. First of all, there is no “holiday factor” to speak of until 2004; thru 2003, activity on the 4th Friday in June differed little from normal. But it bumped up significantly in 2004, and sees additional bumps in 2008 and 2012. I again came up with holiday “factors” that capture those bumps across the years.

**2-51 So how do you determine whether or not a “special day” should be treated as a holiday?**

This unique behavior of the 3rd Friday calls attention to the question of how to treat certain days unique to your organization, when year in and year out performance is significantly higher or lower on or around a particular day. Perhaps you always run a big annual “Clean-out” sale the last weekend in August. Or you like to close your offices for an annual picnic in July. Should these annual events be treated as “holidays” with holiday factors developed for them? It depends.

**2-52 So how do you determine whether or not a “special day” should be treated as a holiday?**

The first criteria for deciding whether a given day should be treated as a holiday is that the occasion should indeed be an annual event, one that occurs every year, and preferably around the same time of year. What if you hold a certain event every other year? I would recommend treating that as an event, and not as a holiday. You can examine the impact the occasion every time you hold it, but just don’t count it as a holiday.

**2-53 So how do you determine whether or not a “special day” should be treated as a holiday?**

The next criteria is whether or not it is a public holiday. If it is widely acknowledged as a public holiday, like the traditional holidays noted on any calendar, it should probably be treated as a holiday. The public is aware it’s a holiday, and their behavior will change to some degree, and to some level of predictableness. But if it is a “private” holiday, one of your own choosing, then treatment will depend on a third criteria.

**2-54 So how do you determine whether or not a “special day” should be treated as a holiday?**

Perhaps the most important criteria for whether to treat an annual event as a holiday is determining whether the impact is due to forces within your control. So, for instance, when you close your offices every year for a picnic you will lose 1 day of activity. Period. This should be treated as a holiday, just be careful to develop different holiday factors should these picnics be held on different days of the week, or perhaps even different times of year.

On the other hand, the impact of an annual sales event will likely fluctuate due to forces other than just the calendar. Sales may be greater or less due to the effectiveness of the sale you hold – how well is it advertised, how big were the price cuts, and so on. In these cases, annual events should NOT be treated as a holiday; instead they will be captured by your monthly seasonal factors, and better yet, by isolated analysis of the impact of these events on your organization each year.

**2-55 Fixed Date holidays are a special category of holidays as they require a different set of factors for each day of the week they fall.**

We now move on to look at fixed DATE holidays. These are a special creature for they will require a different set of factors for each of the 7 days of the week that they can fall on. As you might well imagine, for example, activity will be quite different if July 4th falls on a Sunday instead of a Wednesday.

**2-56 Looking at July 4 as example, there are two notable differences with how Fixed Date holidays are handled.**

Looking at July 4th as example, we note there are two key differences to draw attention to. 1st, we only have a few years to draw on. The example here is for when July 4th falls on a Monday. It did that in 2016; you have to go back to 2011 to find the next occasion having a Monday 4th. Over the past 25 years, the 4th fell on Monday on just four occasions. It is not shown here, but because there are only 4 occasions, the calculations only go the right three more times, not 8 more times as was the case when we were calculating Fixed DAY holidays.

2nd, there will be a separate tab for each day of the week – the tab shown here is for Monday. Note that regardless of what day of the week the holiday falls, holiday factors are being developed for the first 7 calendar days of July – by happy circumstance, July 1 happens to be 3 days prior to the holiday.

**2-57 Setting up Tuesday, and the other days of the week, only required four changes.**

Once the Monday tab was constructed, the others are quick to follow, and only require a few modifications. Critically, each day of the week will have a different set of years from which to draw data. Next, the headings are changed – the day of the week identifier in Cell F1, as well as the tab name. Finally, the headings for the Day of the Week are revised, making sure that the holiday is always the middle column; thus Cell F10 has “Tue”, while for Monday Cell F10 had “Mon”.

**2-58 Weighted average values are “tweaked” by giving less weight to older years, and by tightening the number of allowed standard deviations.**

We then examine the weighted average values section. Again less weight is given to older years, here reducing the weight for the 3rd oldest & 4th oldest dates. Of course, notice how we only have four years of data to draw on. This is why we wanted to have 25 years of data to work with.

I also modified the “Maximum Allowed Standard Deviations”. You may recall that 1.0 standard deviation is generally associated with picking up about 68% of the values, while 1.5 picks up close to 85-90%. By tightening the Maximum to 1.0 here, I’m increasing the likelihood that at least 1 outlier might be removed. This seems quite appropriate as we only have 3 or 4 years of values to work with.

**2-59 For any other holidays you may need to add to those in the template files (e.g. Valentine’s Day), there are still only a few places you’ll need to change.**

So what do you do if you need to add another holiday to those that are found here in this file? This shows you what changes you’ll need to make: the choice of years drawn on (Cells A11:14); the date of the holiday (Cell G1); and the holiday title, both at the top of the worksheet in Cell C1, as well as the tab name. Finally, you’ll need to be sure the day of the week (Cell F1) is being identified appropriately for each of the 7 days.

**2-60 To capture the wider impact of Christmas, the model tracks holiday factors for all of Christmas week, as well as the week before & after.**

We now turn to Christmas. T.S. Eliot called April the cruelest month. For the analyst, December probably merits that title. Before I developed this model, I found the Christmas period would take almost as much time to analyze as the rest of the year put together. This was because it’s impact was spread out over such a lengthy period, it was a Fixed Date holiday so 7 different sets of holiday factors were required, and it marked the end of the month & the year so often there would be a big sales push or other activity and influence going on as yearend approached. It could also be, and still can be, difficult to sort out the effect of the holiday versus the effect of it being December; i.e. the degree to which sales may be down due to the holidays as opposed to basic seasonality associated with the end of the year and the start of winter.

Modeling it as is done here has greatly helped simplify and speed up the process. The trick is to treat the holiday as three separate weeks of holiday period – the week of Christmas gets holiday factors, of course, but so will the week before and the week following. In effect, we will develop factors for the calendar period of December 15 through the following January 4.

How do we go about this?

**2-61 The data for Christmas will always be for the same time period, though an extra week was added to provide a bit more data for comparison.**

As was the case with July 4, because we always have the holiday as the middle day of the week, it works out that the data pulled will be for the same calendar period, regardless of which day of the week the holiday falls. Thus, the data starts at Nov 24, exactly 4 weeks and 3 days before Christmas Day. However, I added another week at the bottom of the dataset – I figure it helps to have an extra week added on to compensate for the extra weeks we lose because of the lengthened impact of the holiday.

**2-62 The week Christmas falls is handled the same as any other fixed date holiday, with Xmas week activity compared with the weeks before & after.**

The calculation of the holiday factors for the week of Christmas is virtually identical to how all other holidays are handled. There are only two notable differences: 1st, as just mentioned, an extra week of data has been added on; 2nd, there will be an extra week that is excluded, the week following Christmas, because that week is impacted by New Year’s Day. The week *before* Christmas however, was *not* deleted. I have found that frequently this week does not behave much differently. I still calculate holiday factors for it as there is often some slight change in activity level, but generally not significantly so.

Finally, note that we often need to drop Thanksgiving week as that week too can skew the results.

**2-63 Indices are then also calculated for the week before Christmas, and the week after.**

The big difference with handling Christmas is that in addition to Christmas week, we want to develop factors for the week before Christmas, as well as the week after. This is actually a fairly easy revision to make. All we need to do is to add on two more sets of Holiday Indices and Determining Outliers calculations, and then make a few key revisions.

First, the formulas in the Holiday Indices section are changed to ensure every week of data is being compared with the highlighted week – that’s why you see “1.00”’s in the highlighted weeks above: because the “holiday” is being compared with itself.

Second, the “Count” cells for the Week Before Christmas were deleted. This was done because we don’t want to have this week included in the calculations – again, we want to know how this week compares with the OTHER weeks, not with how it compares with itself. The “Count” formulas did not need to be deleted in the Week After Christmas section because that week was already excluded up above (in the Included Week section at the top of the worksheet).

Finally, I inserted the blue highlight for the different weeks being analyzed to ease visualizing what is being performed with the calculations.

**2-64 Weighted averages are calculated for each of the three weeks, drawing from the past experience and giving less weight to older years.**

Similarly, the “Weighted Average Values” section also had two more weeks of summary calculations inserted. So we now have holiday factors for Christmas, and for the full 10 days before and 10 days after Christmas.

**2-65 A “Sum” tab pulls in the results for the Fixed Date holidays, and provides a place (Final Values) to insert manual overrides.**

A “Sum” tab pulls all the results for the July 4 and Christmas Fixed Date holidays. It again has a place to insert manual overrides of the preliminary results.

**2-66 When July 4th falls on a Monday, the impact is quite negligible. But Friday sees a definite drop the prior day, and bump the following Monday.**

Let’s take a quick look at the results for July 4. If we compare the 1st day of the week vs the last, in other words Monday vs Friday, we see that the total activity is almost identical – just under 3.8 days of activity. The Monday holiday sees a bit of a drop the prior Friday, while the Friday holiday sees a considerable drop on Thursday but some catch-up increase the following Monday. These seem reasonable.

**2-67 As might be expected, business really drops off the Monday before a Tuesday 4th, and the Friday following a Thursday 4th.**

Tuesday & Thursday holidays also see similar total impacts, though the totals are almost a half-day less than they were for the Monday & Friday holidays. As one might expect, the Monday before the Tuesday 4th sees a big drop in activity, as does the Friday following the Thursday 4th.

**2-68 But a Wednesday has the greatest impact on sales, with an additional day’s worth of business lost in total.**

Wednesday has the biggest overall impact. At 2.98 days of activity for the week, it implies that not only is the full Wednesday holiday lost, but the drop in activity the other days of the week totals another full lost day.

**2-69 The total impact of the 4th is smallest when it falls over the weekend.**

Curiously, and perhaps not surprisingly, the least impact occurs when the holiday falls on a Saturday or Sunday. Evidently, even though the Friday or Monday is off, the rest of the impact is lessened because the holiday itself is one day further away. This, like what we saw with the other days, all seems reasonable.

**2-70 There will be occasions when you want to manually revise the final result. The “Final Values” section picks up the Initial Values, & highlights changes.**

As earlier mentioned, there may be occasions when you think no holiday impact should be measured, that the factor should be 1.00, implying no effect from the holiday. Thus, for example, the Tuesday following the Saturday 4th would not seem to merit seeing any change in activity due to the holiday itself. There is always likely to be some level of noise across the days, and this seems an appropriate candidate to manually revise to a simple “1.00”. Note that conditional formatting highlights the manual change so you are aware later that a manual adjustment was made.

Generally speaking however, I usually prefer to just leave the factors alone. The factors previously were so close to 1.00 that it hardly makes any difference. And do we really “know” the holiday has no impact? Ultimately this is a judgement call. And I would recommend that usually you just leave them alone unless something just plain doesn’t look right.

**2-71 The pattern across the week is very distinctive, and makes sense. Such a pattern could easily be repeated for other enterprises, or fixed date holiday.**

Before leaving the 4th, I want to draw your attention to this chart. It is distinctive that there is such a lovely balance across the days of the week. I highlight this because you may not have daily data, and you’re wondering what the impact would be from the various holidays. July 4th’s pattern here may give you a helpful idea of how to treat holidays at your firm; i.e. that greater holiday impacts will be found as the holiday falls closer to the middle of the week.

The other point worth noting here is the significance of the difference in month lengths simply due to what day of the week the July 4th holiday falls. If volumes drop by 1 day if the holiday falls over the weekend, and drop by up to 2 days if the holiday is midweek, it implies that all else being equal, volumes for the month of July will be as much as 5% greater or less due to this “minor” calendar effect. It speaks to the importance of going through this effort of estimating holiday factors.

**2-72 All 3 weeks of results are brought in for Christmas. The high values are highlighted but left unchanged – they’re the 3rd Friday at quarter-end.**

Finally, let’s take a look at Christmas. All 3 weeks of factors are brought into the summary tab. You can see that some of the days have been outlined in the left-hand Initial Values table, and changed to 1.00 in the “Final Values” section. These are all 3rd Fridays, when as we saw before, activity is much increased. In fact, the activity was more than double in recent years. These factors here are relatively lower, at 1.28 to 2.02, because they represent averages calculated using data from the past 25+ years. Accordingly, rather than treating them as a “Christmas” holiday, I have converted them here to 1.00, and will instead pick up this holiday factor separately as 3rd Friday activity.

**2-73 Monday’s Jan 3 value was lowered – there were a couple of unusually high values in the history, meriting a manual override to reduce it.**

The rest of the factors look reasonable with one notable exception: the January 3rd following a Monday Christmas & New Year’s, which would be a Wednesday. It doesn’t make sense to me that this day would be so busy – there is nothing I’m aware of to suggest otherwise, and the other days the holiday falls see nothing like this level of activity. When I went back and looked at the underlying data, it looked like it was just one of those things – an outlier that merits adjusting. Rather than dropping it entirely to 1.0 however, I decided here to drop it to just 1.10. I don’t like making this kind of change, but I would be more uncomfortable leaving it as it is. The “1.11” value on the 4th following the Monday holiday also seems a bit high and inexplicable; however, it isn’t much above 1, so I decided to leave this alone.

**2-74 The pattern across the week for Christmas has a markedly high Monday total, though the rest of the week is close to what would seem likely.**

This summary chart looks at the total impact of the Christmas holiday season for each of the 7 days of the week that the holiday can fall. It’s striking that as much as 3.5 to almost 5 full days are “lost” due to the holidays. The totals look reasonable for the most part. But Monday’s high total stands out. I would have expected the total to be closer to about 10.7 or so, such that it is higher than Tuesday, but lower than Sunday, implying that the least impact occurs when the holiday falls on the weekend – just as we saw with July 4th. I took another look at the original data, and nothing really jumped out at me. So I reluctantly leave it as is, in the hope that it won’t matter too much in the grand scheme of things.

**2-75 While a lot of work, developing the holiday factors will be very helpful for analyzing the data.**

And with that, we’re done with holiday factors. Yes, they are rather painful to go through. But it is worth it as you end up with a much stronger sense of how long the months truly are, and in turn, can better gauge how well you’re doing at any given moment in time. It will particularly help when you try to measure the impact of internal programs, like marketing campaigns or process improvements that occur at or near any holiday. Any firm that tracks data daily must have a sense of how their performance changes when the numerous holidays across the year occur. The holiday factor calculations will account for those changes.

The factors will also be key when we start putting together the normalization factors, to which we now turn.

**Chapter 3: Normalizing the Data – Adding It All Up**

Now that the Equated Day Factors & Holiday Factors have been established, we’re now ready to move on and normalize the data. We’ll be using normalized data when we embark on seasonally-adjusting the data.

**3-2 Normalizing monthly data refers to the process of adjusting each month’s data so that every month is of equivalent length.**

Normalizing the data refers to the process of adjusting time-series data so that every period is of approximately equal length. Perhaps the most commonly used time-series period is monthly, so we’ll refer here to just monthly data; though obviously, the same concept applies to efforts to adjust daily data, hourly data, and so on.

**3-3 How do we normalize the data?**

How do we normalize the data? It’s a fairly simple process involving four steps. First, the EDFs and Holiday Factors are combined to arrive at Net Daily Factors. Next, these daily factors are added up for each month in order to arrive at the approximate “true” length of each calendar month. Third, each month’s length is divided by the average month length; the result is a normalization factor which measures the degree to which any given month is “longer” or “shorter” than average. Finally, the original monthly actuals are divided by their normalization factor to arrive at the normalized data.

Let’s quickly walk through each of these steps.

**3-4 In order to calculate the Net Daily Factors, we need to bring in the developed Equated Day Factors (EDFs) & Holiday Factors.**

The 1st step in calculating the Net Daily Factors is to bring in all the Holiday Factors and Equated Day Factors into one place. On the website, you’ll find a file called “Normalizing Data Template”. The “Inputs” tab has all the EDFs and Holiday Factors that were developed for the NYSE volumes. To use this file yourself, all you need to do is to replace the factors here with the factors you develop with your own data; just write over all the numbers in the “Input” tab. If you really have no idea about a given number, perhaps you just give it a “1.00” implying the given day or holiday is “average”.

**3-5 Formulas in a “Calc” tab pick up the EDFs & Holiday Factors for the entire covered period.**

The 2nd tab in the “Normalizing Data Template” file performs the calculations of the net day lengths. Columns A through E have some key basic information: the calendar date in Column A, the day of the week, the “Week of” date that identifies the Monday date of any given day, the month, and the original data.

Columns G & H have the equated day & holiday factors of particular interest to us. The EDFs are simply inserted at the start of the period; formulas pick up the factors from seven days before to fill in all other days. Note that there are two sets of factors used in this file, one set for 1991-2005, and another from 2006-forward. If your factors are for different periods, or if you only have one unique set of factors, say, then you’ll need to make sure that the formulas in this “Calc” tab pick up *your* factors. Perhaps you have developed several different sets of EDFs. If so, just add any additional sets to the right of the two sets shown in the “Inputs” tab. And then, in the “Calc” tab, go to the start of each period and insert a formula for the 1st 7 days that pick up the seven days of equated day factors from the “Inputs” tab. You’ll find that the EDF column has been conditionally formatted so that whenever a new set of factors is introduced, the cells are highlighted yellow and bordered – to help make them stand out.

**3-6 Net Daily Factors are calculated by simply multiplying each day’s EDF by its Holiday Factor; summing them arrives at the “true” Net Month Length.**

The Net Daily Factors combine the EDFs & Holiday Factors. It’s an easy operation: the EDF is multiplied by the Holiday Factor. If either of these is “0”, the Net Daily Factor will be “0”. When the Holiday Factor is “1”, which is the case most of the time because most days are not influenced by a holiday, then the Net Daily Factor is the same as the EDF.

When you sum up all the “Net Daily Factors” for any given calendar month, you arrive at the “true” Net Month Length. For the January 2016 numbers displayed here, the Net Month Length is 19.45 days, a relatively short month thanks to the New Years’ & ML King holidays. How much shorter than average was January? We need to look at all the months and see.

**3-7 Monthly and annual totals are calculated for the entire period, along with overall averages.**

The third tab in the “Normalizing Data Template” file is an “Output” tab. Here is where all the month totals are calculated, and an overall average is calculated. For the 1991-2020 period, the average month length was 20.67 days, per Cell C423 highlighted here. You can see at the bottom of the table (Rows 391-on) that every year is of slightly different length; this is the result of the subtle influence of the equated day factor for one day of the week occurring one more time than usual. And, of course, every four years there’s an extra day with leap year.

**3-8 Monthly data can be put into a table to more easily observe how month lengths vary over time.**

It’s informative to see how month lengths compare over time. I constructed a table that allows an easy look at all the month lengths. The January 2016 total of 19.5 days, which we saw earlier, is highlighted in Cell AJ11.

Note that this table shows the month lengths BEFORE inserting the holiday factors for the 3rd Fridays and for the 4th Fridays in June. Those are such a special case that I excluded them to make this comparison view more meaningful, and less skewed by this unusual “holiday” that the NYSE has.

**3-9 Year lengths vary slightly, and leap years are not necessarily the longest.**

Let’s take a look at some of the more unusual comparisons. First up is this odd comparison of the 2011 and 2012 totals. You’d think 2012, with a leap year, would be longer; and yet its total is almost one day shorter than 2011. How did that happen? If you look at the side-by-side comparisons of the calendar months, you can see that 2012’s February is indeed 1 day longer than 2011. But then 2011’s March is approximately one day longer than 2012’s. 2012’s longer May is offset by 2012’s longer June. And 2011’s much longer September is offset by 2012’s much longer October. With January, April, July, and November of roughly equal lengths, that leaves December to “break the tie”. And as 2011’s December is one day longer than 2012, it follows that the year length for 2011 is also one day longer.

But that still doesn’t explain how the entire longer year of 2012 winds up shorter than 2011. A closer examination of the calendar explains it. 2011 had 365 days, with Saturday being the one day that happens 53 times (while the other days have 52 occasions). 2012 had 366 days, with Sunday and Monday occurring 53 times. Saturday and Sunday are obviously weekends, so there is no length added to the year for these days. But the 366th day of 2012 is a Monday, so you’d think that would translate into an extra day length for 2012. Not so. Why? Because in 2012, New Year’s Day falls on a Sunday, which means the Monday is off, while in 2011 New Year’s Day is a Saturday which means the prior Friday, which occurred in December of 2010, is off. Because Monday, Jan 2, 2012 is a holiday, there is no extra time gained in 2012 for having the leap year. So that explains why 2011 and 2012 should be of approximately equal length. How is it 2012 ends up shorter? Because in 2012 July 4 fell on a Wednesday, while in 2011 it was Monday. As we saw in the Holiday Factors chapter, activity is almost 1 full day less when the 4th falls on a Wednesday, rather than a Monday.

Whew. That’s a pretty long explanation, but it is interesting to see how it is that a longer year can end up shorter, or vice versa. The devil is in the details, and sometimes those little details can make a difference so you want to be able to understand how they do.

**3-10 Month over month lengths can change significantly.**

It is striking to see how much month lengths can vary from one month to the next. Here we see July 2007 had just under 20 days while August 2007 was almost 3 days longer. And then in 2008, August winds up one day shorter than July. Such can be the whims of the calendar effect, and why it can be so important to recognize the “true” length for any given month.

**3-11 Month lengths can also change dramatically year-over-year.**

As I emphasized in the summary overview, many organizations are fond of measuring how well they do in a given month by looking at how they did as compared with the same month one year before. The perils of such comparisons are evident when we look at these same two pairs of calendar months. Notice how July 2008 is almost 2 days longer than 2011, while August 2008 is 2 days shorter. In each case, one month’s length is about 10% different than the other. Clearly, taken on its own, you would expect YOY results to look much better, or worse, simply because of this calendar effect. Which is why it is so important to adjust for month lengths when making YOY comparisons, a practice which I see rarely performed. But then, as was further discussed in the summary Introduction, YOY comparisons are fraught with other perils that still render it an operation best done sparingly, and with caution.

**3-12 Almost half the time, year-over-year month lengths change by 5% or more; more than 10% of the time, they change by 10% or more.**

It’s informative to see how frequently, and by how much, YOY month lengths can vary. This chart shows the distribution of YOY month length changes for NYSE sales. It shows that less than 25% of the time, YOY changes are almost zero. And it shows changes are still only 1-2% about 12-14% of the time. But notice how month length changes are 5% higher more than 15% of the time, and 5% lower about 13% of the time. And a total of almost 10% of the months change by 9-10% or more, year-over-year. In fact, more than half the time the YOY change is 4% or higher.

Nonetheless, how often do you see reports that carefully calculate the percentage change for a given month compared to one year prior? A report might read something like: “August 2016 sales were strong, up 8.36% over last year.” Too bad the statement neglected to mention that August 2016 was almost 10% longer than the prior August. Not only does the statement present a false description of how the two months compared, but it’s also guilty of an utterly inappropriate precision. “8.36%” is a figure that suggests a level of accuracy that is clearly completely unwarranted. And yet countless reports published every day are guilty of this, of what I call the “patina of precision”; the tendency to show figures with extra digits which suggest a level of precision that just isn’t there. Much better in this instance to have instead written something like: “August 2016 sales were 8% higher than prior year; however, the month length was 10% longer than last August (as we were open 23 days instead of 21).” Better still of course, is to present all your data in seasonally-adjusted form and speak to that, but I get ahead of myself.

Let’s leave all this with one point abundantly clear: a straight calculation of year over year changes is a lot less meaningful when you take into account these kinds of changes in month lengths.

**3-13 Normalization Factors compare each month’s Net Length with the Average Net Month Length.**

Once you know the “true” lengths of each month, calculating Normalization Factors is easily done: they are simply each month’s Net Month Length divided by the average Net Month Length. We saw that January 2016’s net month length was 19.45 days, and the overall average month length was 20.67 days. So the Normalization Factor for January 2016 divides the 19.45 days by the 20.67, to come up with a normalization factor of 0.94.

**3-14 Normalization Factors are calculated for every month for the entire period.**

This calculation is made for all the months, and is performed here in Column D. The sum of the factors each year is also shown.

**3-15 Normalizing the data requires dividing each month’s Actual amount by its Normalization Factor.**

When you have the Normalization Factors it becomes a similarly simple exercise to then normalize the data. The original actual amount is divided by its Normalization Factor to arrive at the Normalized amount. In January 2016, actual NYSE sales totaled 29.393 billion shares; when adjusted for the low 0.94 factor, the sales bump up to a normalized amount of 31.236 billion.

**3-16 Original Actuals are normalized for every month for the entire period.**

Again, sales are normalized for the entire period, and annual totals are summed.

**3-17 So what impact is made on the Original Actuals when the data is normalized? Here are the Actuals.**

So what kind of difference does it make to normalize the data? We start with a chart of Actual monthly sales for the NYSE from January 2011 through October 2016. You can see quite bit of volatility from month to month, not to mention a huge spike in August 2011.

**3-18 While not always the case, normalizing the data usually helps smooth out some of the volatility in the data.**

Normalizing the data appears to definitely help reduce the volatility. The normalized data is not “smooth”, nor would we expect it to be when we consider we are following activity that we know from watching the nightly news can be subject to significant impacts and volatility every day. Nonetheless, it does appear less volatile than the original data. Some high values are made lower, and some low values have increased. The spike of August 2011 is less impressive, Of course, sometimes the data looks noisier, as is the case near the end of 2012. But that is more the exception, and perhaps it’s the case that those months are more accurately portrayed as being more volatile than the original.

And with that we are done with normalizing the data. Again, it may seem like an awful lot of work to make changes that appear less significant when charted as they are here. But I have found that these adjustments are still well worth making. Each month, people look at monthly reports to see how they did that month; normalizing the data can render reported values more accurate by enabling the reader to make more appropriate and accurate comparisons with other months; (of course, such comparisons will be much better when they’ve been seasonally-adjusted as well). These adjustments are a very important prelude to calculating the monthly seasonal factors. And finally, making these calculations will make you that much more comfortable in your reporting and analysis of “how you are doing”.

**3-19 An aside on Retail Trade: Many in this industry like to divide the “months” into weeks of 4-4-5. There are some issues with this approach.**

Before leaving this whole discussion about month lengths and normalizing the data, I would like to address one particular industry that I know can often handle month lengths quite differently than simply examining the traditional 12 calendar months. I refer to retail trade, the large department store chains and other enterprises that typically have major activity occur on weekends, and as a result use a quite different calendar. Their calendar often comes in the form of the 4-4-5 calendar, referring to how the 1st two months of each calendar quarter are measured as exactly 4 weeks long, while the 3rd month is made 5 weeks in length. This certainly helps work toward eliminating the significant impact of the weekend sales by simply ensuring that any given month will always have precisely 4 or precisely 5 full weekends, thereby seeming to make year-over-year and month-to-month comparisons easier and more accurate.

But I have some concerns with this approach. 1st, sometimes a 4-week month may not pick up the first calendar day or last calendar day of the month, and as some people may receive a paycheck on the last (or first) day of the month, you may miss capturing that extra sales observed on these pay days. Or, a 5-week month may encompass two month-ends and be guilty of slightly overstating sales performance.

Secondly, Decembers will obviously be especially vital, but when those Decembers can end a few days before New Year’s, or a few days after, you may have trouble making a true apples-to-apples comparison of one December versus another.

While we’re on the topic of Christmas, it must also be noted as a 3rd concern, that as we saw with NYSE sales, the day of the week this holiday falls can make quite a difference. There’s no reason to believe one wouldn’t see similar substantial differences in retail trade, if not greater. No problem if the analysis has been made that clarifies & quantifies the impact of the day of the week the holiday falls. If.

Another potential problem with the 4-4-5 calendar is that some holidays may not work well under this approach. For example, Memorial Day weekend may sometimes be in the 4-week month of May, and sometimes in the 5-week month of June. Or possibly much more problematically, Thanksgiving AND Black Friday may fall in November, or December. Now quite possibly this is not an issue as stores carefully align the calendar to always have these holidays fall in one particular month. But this can be tricky to do consistently when you consider so many holidays near month-end: Easter, Memorial Day, July 4th, Labor Day, Thanksgiving, Christmas & New Years.

Fifth and finally, of course, there is the unfortunate need to add an extra week every 5 or 6 years to keep the calendar on track with the 365+ calendar year. Everyone will have a way of handling this. But to what degree do these adjustments skew the reading of annual or year-over-year results?

I’ve never had occasion to work with daily data at the retail level, so I honestly couldn’t say how much of a problem this is, or how well it is typically addressed. But it seems to me that 4-4-5 has enough issues that it would certainly be well worth the effort to at least consider a different approach, an approach like what I have been outlining here.

At the very least, for those working in retail trade, or for those simply reading about how Christmas sales fared this year compared to last, hopefully this discussion makes clear that such comparisons should probably be taken with a grain (or two) of salt.

**Part II: Seasonally-Adjusting the Data**

In this Part II, seasonally-adjusting the data is explained, a process which takes the normalized data and adjusts for seasonality, the typical cyclicality across the 12 calendar months of the year. An initial set of seasonal factors are calculated in Chapter 4. Chapter 5 will walk through the construction of the trend model that adjusts the history for growth & events, in order to arrive at a “final” set of seasonal factors. Chapter 6 walks through estimating trend, capturing your best estimate of how the data trends over time using manually input estimates of the growth rates and events.

**Chapter 4: Initial Seasonal Factors**

Now that the data has been normalized, so every month is of approximately equal length, we’re ready to start seasonally-adjusting the data through the development here of initial seasonal factors. These initial factors will be determined by formula.

**4-3 The initial seasonal factors are obtained almost entirely by formula, while the final, growth-adjusted seasonal factors will entail more manual intervention.**

While the initial seasonal factors are obtained by formula, the final seasonal factors will entail manually entering estimates of growth rates & events. The initial set of seasonal factors will help improve the growth & event estimates, for the more we can adjust the data for predictable impacts of the calendar effect and seasonality, the clearer & easier the read will become of how the data is trending & shifting over time.

Nevertheless, be aware that calculating these initial seasonal factors is not something that you have to do. You will find it will assist the procedure for developing the final seasonal factors, but it is not necessary to calculate these initial seasonal factors.

It’s also important to recognize that as the initial factors are obtained entirely by formula, it can provide a practical approach for handling numerous data sets.

**4-4 What we’ll cover in this chapter:**

After this introduction to initial seasonal factors, this chapter will walk through the method for estimating initial seasonal factors, and then review the results and consider some simple modifications that may merit being made. We’ll look at how you may want to change the format of how the calculations are laid out in the workbook. And finally we’ll consider using initial seasonal factor calculations alone for those occasions where you may have many sets of data you’d like to seasonally-adjust.

**4-5 The quick calculation of initial seasonal factors involves taking the normalized data, indexing it, removing outliers, & computing a weighted average.**

The development of the initial seasonal factors is a fairly quick and simple procedure, outlined in the Excel spreadsheet here – and which we’ll go through step by step. At a high level, the process involves taking the normalized data, indexing that data, calculating standard deviations and expressing each data point’s distance from the mean in terms of the standard deviation. As done before, applying a Maximum Number of Allowed Deviations leads to certain years being excluded as they have one or more months of data falling outside the accepted range. A weighted average is computed that simply picks up factors for only those years where all 12 months are accepted.

**4-6 The calculation of the initial factors starts with bringing in the normalized data, and presenting it in columnar form.**

The first step in the procedure is to start with the data. Here the normalized data is brought in, and laid out in columnar form, with the months in rows and the years in columns. The table shows the data for 2001 thru 2016; the years 1990-2000 have been hidden from view to better display the methodology.

**4-7 Next the data is indexed each year.**

The normalized data is indexed, with each month’s value expressed as a percentage of the average monthly value for that given year. Thus, January 2001’s index is 1.01, obtained by dividing the January value of 26.5 million by the 2001 monthly average of 26.3 million.

**4-8 Simple averages and standard deviations are calculated for the time period(s).**

Over to the right hand side of the indexed data is the calculation of the simple average and standard deviations. For the 2001-16 period, January’s mean is 1.01 and the standard deviation around that mean is 0.07.

**4-9 Then, each month’s distance from the mean is calculated, with conditional formatting added that highlights the “strays”.**

As we did with the calculation of EDFs and holiday factors, we express each month’s value in terms of how far it is from the mean, with that distance expressed in terms of the number of standard deviations. Let’s look at the highlighted values for January 2007: the index is 0.87 (Cell S31), which is 0.14 less than the 2001-16 January average of 1.01 (Cell AI31). The standard deviation for January is 0.07 (Cell AL31), so as January 2007 is 0.14 from the mean it is approximately 2.0 standard deviations from the mean (Cell S51). Conditional formatting highlights that value in yellow, indicating the value is somewhere between 1.5 and 2.0. Had the value been above two, it would be highlighted in orange. There are mostly green colored cells, so most index values are less than 1 standard deviation from their month’s mean. Note that 2008-10 has lots of significant outliers – not surprising when we recall the financial crisis in the fall of 2008.

**4-10 A “Count” table establishes whether any value is outside the Maximum Allowed Standard Deviations; if all 12 months pass the test, the year counts.**

We want to determine whether a given month’s index value falls within the accepted range. This is done by comparing the “Number of Standard Deviations from the Mean” value with the highlighted “Maximum Allowed Deviations” value. The Maximum has been set at 2.0 (Cell AM44), so all months that are in orange, indicating they are over 2 standard deviations from the mean, are not counted. The “Count Data” table at the bottom highlights the months that failed to meet the criteria. Only those years where all 12 months meet the criteria are counted, indicated by a value of “1” in Row 85.

**4-11 Weighted averages compute the average for all counted years, for each time period. These are the Initial Seasonal Factors.**

A weighted average formula, in the form of a SUMPRODUCT formula in Excel, computes the average index values for those years that are counted. For January, the average for the accepted years (2003-07, 2012, 2013, & 2015) is 1.00 (Cell AI71); this is the initial seasonal factor for January for the 2001-16 time period.

**4-12 The 2001-16 data is plotted; for the most part, the averages look reasonable.**

All the years of data are plotted so we can see what these initial seasonal factors look like. The dotted blue lines indicate the high range and low range for each month. Whenever a month’s value is above the high range or below the low, that month and the rest of that year, are removed from the calculation. The remaining years’ averages are the highlighted Weighted Average. These averages, the initial seasonal factors, generally look quite reasonable – no remarkable figures, though August is clearly the quietist month of the year, and December the busiest. Keep in mind that this is the normalized data, so December values have been raised somewhat thanks to the holiday factors (i.e., December is typically a shorter month, due to the impact of the Christmas holidays; but the lower normalization factors for the month lead to higher normalized values for December).

**4-13 While the weighted average for most months falls within the thick of the data points, August’s average is above it. Why?**

If we look more closely at each month’s weighted average, we’ll notice that the August average is above a significant concentration of values that fall below it. The other months do not behave like this, but instead generally have their weighted average falling within the same vicinity of the concentration of most of the data points. Why does August behave this way?

**4-14 August has one data point that is just below the high range of allowed deviation.**

Notice the highlighted “High Value”: it is well above the weighted average, and is just barely below the dotted blue high value range. If the allowed range was slightly lowered, this high value would fall outside the accepted range, leading to a reduced weighted average for August. Note that there are no low values in August that are near the 0.60 minimum allowed value for the month.

**4-15 To effect what years are counted, we can modify the Maximum Allowed Standard Deviation value; here we drop it slightly, from 2.0 to 1.9.**

There’s an easy way to remove the unwanted “outlier”: we revise the Maximum Allowed Deviations, dropping it here from 2.0 to 1.9.

**4-16 Lowering the allowed deviations in 2007 causes the August (as well as January & November) value to not be counted; so the entire year is not counted.**

That slight drop in the Maximum causes August 2007 to not be counted; it also causes January and November to not meet the slightly tighter criteria.

**4-17 Dropping 2007 from the calculation causes the August average to drop, and is now closer to most August data points; other months’ values change slightly.**

With 2007 being dropped from the Weighted Average calculation, August’s average drops, from 0.94 to 0.89. Though still above the concentration of August data points, at least it is now right at the top of them, not well above. The other months weighted averages also change of course, though none appear to change very much.

**4-18 Revising the overall Maximum value is the preferred approach for addressing “near outliers”, values that seem to skew monthly averages.**

This can be a nice simple technique for addressing initial seasonal factors that may have any month(s) that appear somewhat higher or lower than they appear they should. As with much of statistics, though everything is calculated in a very precise manner, there will invariably be occasion to apply some level of judgement to the procedure. Such judgement and modification should always be done appropriately, and with care. We want the data to speak for itself, we do not want to manipulate the data to get it to tell us what we want to hear. But this August “outlier” seemed, in my judgement, to be in need of reduction, so that the average was closer to the values that a substantial majority of the data seemed to be coalescing at. By simply modifying the Maximum Allowed Deviation value, and by modifying it to a modest degree, we strike upon a fair and relatively unobtrusive way of obtaining initial factors that properly reflect the data.

**4-19 Initial seasonal factors were also calculated for 1990-2000. Curiously, the monthly values are much tighter, and growth near yearend is steeper.**

You may have noticed that the years 1990-2000 were hidden from the displayed spreadsheets, but that a weighted average calculation for the period was nonetheless being performed. The chart here shows how those years look. What first strikes me is how much tighter the data behaved. The High & Low values (set at 1.9 standard deviations) are pretty close to one another, certainly much more so than was the case with the 2001-16 data. The pattern here is also quite different. The 1990-2000 data has a relatively strong finish to the year, while June & July are well below the average for the later period. All the weighted averages for 1990-2000 seem to fall fairly close to where each month’s concentration of data points, so no need to adjust again the Maximum Allowed Standard Deviations.

**4-20 If you have fewer years of history, you may want to play with the format to have everything fit nicely on one page, with one screen look.**

I’ve found that it’s always advantageous, if possible, to display all of the pieces of a key calculation in a way that they all fit on one screen. That should certainly be feasible with the initial seasonal factor calculations should you be using a limited number of years of data. In the example displayed here, we only have 7 years of data. This example shows one way of arranging the spreadsheet so that all the data and calculations are visible in a single screen view. How you format and arrange all the data is of course entirely up to you.

**4-21 However, with fewer years of data, you will very likely need to reduce the Maximum Allowed Standard Deviations.**

One thing to be aware of, however, when you have fewer years of history to work with, is the need to evaluate the value used for the Maximum Allowed Number of Standard Deviations. You will likely want to reduce this, for a value of 2.0 (or more) may often lead to all data being used. It will often be the case that you’ll see there is a need to make the accepted range a little tighter so that at least one or two years of data are removed from the weighted value calculation. This will necessarily be a judgement call, and it is impossible to anticipate how your particular data will behave. Just be prepared to need to adjust the Maximum so that some useful discretion is applied to your data.

**4-22 What if you have literally hundreds or thousands of sets of data? You may be tempted to develop seasonal factors for each set entirely by formula.**

I mentioned in the introduction to this chapter that another valuable feature of the initial seasonal factor calculation method is that, as it’s all formula, it can be fairly easily applied to a large number of sets of data. And certainly in this age of Big Data, you likely have almost innumerable quantities of data to work with, and to which you may wish to apply some form of seasonal adjustment calculation. Data can be sliced and diced in so many ways, with such ease. But when it comes to looking at time series data, the key thing you have to guard against is using too thin a set of data, where the numbers are so small and the data experiences such volatility, that seasonal adjustment is not only rendered meaningless, but can add a level of apparent precision that is completely inappropriate – again, the patina of precision that can mislead people to overestimate the level of accuracy regarding the results.

**4-23 The problem with seasonally-adjusting data at a very granular level is that you will almost always encounter much greater volatility.**

I strongly recommend you resist the temptation to resort to formulas alone to seasonally-adjust your data. The problem is that as you break out your data into smaller and smaller subsets, the behavior of the data becomes more and more erratic. For example, let’s look at home sales in the state of Wisconsin. Wisconsin has 72 counties. This chart shows indexed home sales for Florence County, a small county in the northeast corner of the state. Many months have a value of “0”, other months’ values vary. Such a dataset is far from unusual. Certainly many expense accounts will often have “0” values interspersed with other months that may have very small or occasionally very large values. It is almost impossible to arrive at meaningful seasonal factors for such erratic data.

**4-24 Grouping your data to some degree will usually address this issue, resulting in cleaner data with clearer patterns.**

Where you do deal with tiny subsets of data, I recommend that, whenever appropriate, you endeavor to combine the data with other similar sets of data, so that you may arrive at a dataset that behaves much better, yet represents an appropriate proxy for the original data. Here, Florence County home sales are combined with the rest of the North Region of Wisconsin, as defined by the Wisconsin Realtors Association (the source for this data). The data appears to have a similar pattern from year to year. How similar? Let’s run it through the initial seasonal factors calculation model.

**4-25 Note how much tighter the seasonal pattern is for North Wisconsin, as compared with Florence County.**

Not surprisingly, the initial seasonal factors for the North Region are much tighter than they are for tiny Florence County. This example may seem extreme, but I’m confident you will find instances in your own data where the variation is even more exaggerated.

**4-26 When faced with ill-behaved granular data sets, generally you do best to group your data, applying the group total’s factors to all sub-categories.**

In short, when possible, try to combine volatile data with other meaningful data of a similar nature, to arrive at a set of seasonal factors that is both useful and appropriate. Thus, the North Region seasonal factors here could be applied to each of the 18 counties listed.

Of course, when you deal with very volatile data, applying “appropriate” seasonal factors can still be tricky, for the underlying data is still tiny and volatile, and your seasonal factors will likely do little to improve their appearance. That’s the way it goes. You can only do your best. But should there be an interested party in your company that wants to closely examine your tiny datasets, you can try to meet their demands by using seasonal factors for an appropriate combined set of data that includes the specific data of interest. Just don’t expect much….

**4-27 But when you have an erratic and outlier set of data, with no real “home”, it’s probably best to just assume no seasonality at all.**

When you have an erratic set of data, and particularly if it doesn’t have natural comparables, your best bet may be to assume that there is no measurable seasonal pattern, in which case each month is simply given a seasonal factor of “1.0”. Nothing wrong with coming up with no seasonal pattern; indeed, it can be most appropriate not to have one, and inappropriate to come up with a pattern that’s really just a mess with little apparent explanation or meaning.

The calculation of initial seasonal factors is a first step, and a fairly crude one. Nonetheless, even this will generally help smooth your data behavior and could prove useful even when applied to hundreds or thousands or more sets of data.

Just recognize that if any set is ever questioned, or greater scrutiny and analysis is requested by an interested party, you will ultimately want to take the next step: to closely, and manually, examine your data for growth and events. This will lead to a more accurate and “final” set of seasonal factors. To which we turn next…

**Chapter 5: Growth-Adjusted Seasonal Factors Model**

This chapter describes the model to be used for estimating trend and developing the final, growth-adjusted seasonal factors. The focus of this chapter will be the model structure, an explanation of why and how it works the way it does. The next chapter will walk through the trend estimation process itself which culminates in the final growth-adjusted seasonal factors.

**5-2 This chapter will walk through the model structure that will be used to arrive at the growth-adjusted seasonal factors.**

This chapter walks through the structure of the Trend Model. It begins with an introduction reviewing why we need to adjust the data for growth before arriving at a final set of seasonal factors. The chapter then walks through the three basic parts of the trend model: the Inputs, the Calculations, and the Trending.

**5-3 Growth is removed so the seasonality measure only captures seasonality, nothing else.**

Why do we need to adjust the data for growth and events before we calculate seasonality? The importance of adjusting for growth is apparent in this simple example. Here there is some seasonality across the year, with summer busier than the rest of the year. But notice how because of growth, December (as well as October & November) is consistently busier than January (& February). If you were to estimate seasonality without adjusting for the growth, you would falsely conclude December is significantly busier than January.

**5-4 Removing growth reveals the “true” relationship between January and December (& all other months).**

After we adjust for growth, you can see how the data behaves when growth is taken away. Note how January & December in Year 2 are at a much more similar level – no longer does December come in much higher than January, thanks to our adjusting the data for growth.

**5-5 Adjustments need to be made for events as well.**

We also need to adjust for events. Here we have a 15% bump in September that we would want to adjust for. The adjustment would reduce the last 4 months of this year by the 15% bump. Just as with growth, adjusting for events helps ensure the measured seasonal pattern captures fluctuation across the typical year that is due to seasonality alone.

**5-6 After adjusting for growth & events, the “true” underlying seasonal pattern emerges.**

Once the data has been adjusted for growth & events, as well as the calendar effect, we have a clearer picture of the “true” seasonal pattern across the year. It’s quite striking how much cleaner the pattern is after adjusting for growth & events. Obviously I acknowledge that rarely will the pattern each year be as clear and steady as it is here. “Noise” was much reduced for this hypothetical dataset in order to help clarify the necessity of adjusting for growth, & events.

**5-7 The Excel model for developing the growth-adjusted seasonal factors, and for estimating trend, has three parts.**

The model I’ve created for adjusting the data for growth is called, appropriately enough, the Growth-Adjusted Seasonal Factors Model. It is divided into three parts, laid out over three tabs.

**5-8 The “Inputs” tab pulls in all outside data that will be used in the Trend model.**

The “Inputs” tab draws data in from other files that will be used in the model. This is also the section where each new month of data would be added.

**5-9 The “Calc” tab is where most of the calculations for the model take place; it is designed to be left alone for the most part.**

The “Calc” tab performs most of the calculations in the model. Except for monthly updates, it should never require revision.

**5-10 The “Trend” tab is where the Growth Rate and Events estimates are developed.**

The “Trend” tab is the heart of the model; it is where the trend estimates of Growth Rate & Events are manually entered. A Trend Chart is there to assist with developing the growth rate & event estimates. The “Trend” tab is also where the final seasonal factors are calculated.

**5-11 The “Inputs” tab contains all the key information drawn from other files that will be used to estimate trend.**

The “Inputs” tab draws in the key data, generally brought in from other files. This is where you bring in Actuals, and where new updates come in. It is also where the normalization factors are brought in, as you can see from the table here.

Note that the normalization factors go through 2020 – you’ll want these when you develop your forecasts.

**5-12 The “Inputs” tab also brings in the seasonal factors – the initial factors already developed, and the growth-adjusted factors to be developed in the model.**

The initial seasonal factors, calculated in the last chapter, are brought in. For clarity, I also have the other two sets of seasonal factors here, so they are all in one place. The set of factors labeled “None” here is where there is no seasonality, so all the factors are simply 1.00. This is a set you may want to try if you want to see how the normalized data looks without adjusting for any seasonality. The final “Growth-Adjusted” sets of factors here are determined in the “Trend” tab.

This is all I’ve placed in the “Inputs” tab. But be aware the Inputs tab is a place where you’ll want to park other outside info that you may want to work with as you trend and analyze your data – such as economic data, or competitor info, or price changes, or whatever.

**5-13 The “Calc” tab is where all the calculations are performed. Except for updating monthly data, this tab need never be touched.**

The “Calc” tab is where most of the calculations take place. It is designed to never need changing or tweaking; that’s all done in the “Trend” tab. The exception is that each month some formulas are copied down so that the model updates in a way that keeps the Trend chart “clean” – we’ll see what I mean when we later walk through the “Trend” tab.

The display of the tab here shows the 1st year of calculations, then hides everything until January 2016. Ordinarily, all the months are on full display. I’ve formatted this way of course, so you can see what is going on at the start & finish of the time period; all the “hidden” months have the same formulas as those visible here. As you can see, the “Calc” tab is split into five sections. We’ll now walk through each of these sections to explain what they are about.

**5-14 The first part of the “Calc” brings in the original data, allows for adjustments if needed, and calculates the normalized monthly amounts.**

The “Actual & Normalized Data” section brings in the original actuals from the “Inputs” tab. An “Adjustments” column (Col C) has been inserted. Why? Because I’ve found on occasion that sometimes you need to perform a manual adjustment to the data, often necessary due to an accounting issue. Perhaps the data for the last day or so of the month came in too late to be counted so one month is artificially a bit low and the subsequent month too high. This issue can be obvious on a chart when you see a low spike one month, and a high spike the next, or vice versa. If you can get accounting to fix it great, but often you can’t, so this column is a place to insert a manual modification. But try and be sure to have your adjustments add to exactly zero, unless you know the grand totals for your firm are running lower than the correct amount.

The adjustments are added (or subtracted) form the original data to arrive at adjusted data, which is then normalized using the normalization factors brought in from the “Inputs” tab.

**5-15 The next section seasonally-adjusts the data, using one of the 3 sets of seasonal factors.**

The “Seasonally-Adjusted Data” section calculates the seasonally-adjusted amount for each month. As you can see, though, it has 3 different sets of seasonal factors to choose from: “None”, “Initial”, & “Final”. Which set is to be used is decided in the “Trend” tab. This “Calc” tab just brings in your selection and assigns the appropriate set of seasonal factors to use. The three sets are as follows: “None” (Col. G) is the assumption that there is no seasonality; every month has the factor 1.00. “Initial” (Col. H) is the set of Initial Seasonal Factors, as was described & developed in the last chapter. This column could be used for any possible set of factors that you want to work with initially. As we mentioned earlier, you could pick up your developed factors for Product A, to apply as a useful starting point for Product B. The “Final” seasonal factors, in Column I, are brought in from this file’s “Trend” tab. Note that these values are in blue format; this is my way of recognizing the values that will be coming in from the “Trend” tab.

Column J shows the seasonal factors that are being used. The highlighted Cell J8 at the top identifies the chosen seasonal factor set. Again, this choice is made on the “Trend” tab; the number determining which set of factors will be picked up by formula.

Column K is the desired seasonally-adjusted amount; it is calculated by simply dividing the Normalized Data (from Column F) by the chosen Seasonal Factor (from Column J). Finally, Column L is a running 3-month average. As we’ll see when we chart the data in the “Trend” tab, this smooths the data and can help determine the general trend and behavior. Note that these last 2 columns are blank from January 2017-on. This is so the chart stops with the last full month of data that is available, We don’t want the chart to plunge to zero in February because there is no data yet for that month.

**5-16 The 3rd section estimates the data trend, using the Growth Rates and Events that will be estimated in the “Trend” tab.**

The next section of the “Calc” tab calculates the estimated trend. Let’s be clear about what is meant by “trend” here. The “trend” is a rough approximation of how the data is behaving over time. Even after seasonal adjustment, data will show a fair amount of noise. The trend endeavors to cut through that noise and identify the approximate behavior or trend. You’ll better see what is meant by this when we include the estimated trend in our chart on the “Trend” tab.

The trend estimate has two components: growth rate and events. The “growth rate” describes the slope of the data on an annual basis; “events” describe the step functions that occur. Estimates of growth rate & events are performed, manually, in the coming “Trend” tab.

Notice the highlighted “Starting Level” in Cell O8. This is an initial estimate of the level the data is at as we begin this process.

**5-17 Each month’s Estimated Trend is calculated by taking the prior month’s value and applying to it 1/12th of the growth rate, and all of the event estimate.**

The estimated trend is calculated by picking up the prior month’s trend value and then adding on growth & events. Thus, the estimated trend amount at the end of May 2016 is 26.036 million. The annual growth rate estimate for June 2016 is 1.0%. We only want to add on one month of growth so we multiply the starting value of 26.036 by the 1.0% growth rate, and get 0.260 million; we then divide that by 12 to get 0.022, which when added to the initial 26.036 gives us a total of 26.058. Occasionally there are events – these are treated differently in that they are taken whole. So in July 2016, the starting amount of 26.058 million is grown by the 1.0% annual rate: adding another 0.022 million gets us to 26.080 million). The event is a significant -12.0% drop: 12% of 26.080 is 3.130. When 3.130 is subtracted we get the 22.950 million shown here in Cell O197.

**5-18 Each month’s Estimated Trend is calculated by taking the prior month’s value and applying to it 1/12th of the growth rate, and all of the event estimate.**

The last column in this section is the “Estimated Trend Values”. This column takes our Estimated Trend (from Column O), and puts back in the noise of the calendar effect and seasonality. Why bother with this? Because it is absolutely essential that our estimated trend line is at the approximate same level as the actuals, in the long run. Each month the estimated trend is going to be a bit above or below the actuals; we want to make sure that over time the lower and higher amounts offset one another, such that the total actuals and total estimates are almost the same. However, the comparison will be made with the estimated trend values rather than the simple estimated trend itself. Although the estimated trend is a simplified line describing how the data is behaving over time, it is still ultimately describing how the data behaves with the noise put back in. We want the actuals, with all their noise, and the estimated trend values, with its noise, to approximately match in the long run. How are the estimated trend values calculated? By picking up the estimated trend amount from Column O and multiplying it by the Normalization Factor (from Column E) and the Seasonal Factor in use (from Column J). Thus, for July 2016, the estimated trend of 22.950 million (Cell O197) is multiplied by the normalization factor of 0.96 (Cell E197) and by the seasonal factor of 0.98 (Cell J197) to arrive at the estimated trend value amount of 21.596 million (Cell P197).

**5-19 The 4th section is the forecast, which takes the estimated trend and extrapolates it using the assumptions for future growth rate & events.**

The 4th section is where the forecasts are calculated. The values here do not “start” until the 1st month for which we do not have data; thus the column is blank until February 2017. The forecast trend is labeled here as “extrapolated” trend. Why change the name? Because in some organizations, the person or department running this model may not be the same as the department responsible for the “official” forecast. Perhaps the Sales department is on the hook for that, and perhaps you’re in the Financial Analysis area, so you aren’t responsible for the official forecast. Instead, others look to you to help understand how the data is trending (at least they certainly should be looking to you for that). When you are generating projections of trend and trend values, you want it to be clear your figures are not official and not meant to replicate what the Sales department is putting together. I have found it useful to clarify the distinction by referring to the projections as “extrapolated” trend and “extrapolated” trend values.

The numbers here are derived in the exact same way as the estimated trend and estimated trend values. Thus, the extrapolated trend takes the prior month’s trend and adds to it using the estimates of growth rate and events. The formula makes sure that in the very 1st month, the extrapolated trend picks up the prior month’s trend from Column O; in subsequent months, it will pick up from Column Q. Thus, the Extrapolated Trend for February 2017 picks up the Estimated Trend for January 2017 (23.065 million in Cell O203) and applies the annual growth rate estimate of 1.0% (Cell M204) to it, to get to 23.084 million (Cell Q204); there is no event estimate for the month (Cell N204 = 0%). In March, another 1.0% growth estimate (Cell M205) is applied to the 23.084 million (Cell Q204) to arrive at 23.103 million (Cell Q205).

Meanwhile the extrapolated trend values take the extrapolated trend and multiplies it by the noise of the calendar effect and seasonality.

**5-20 The 5th & final section picks up the past & projected trend, as well as the past actuals and extrapolated trend values.**

The 5th and final section picks up past & future values and puts them all in one column. Thus, for trend in Column S, it picks up the estimated trend from Column O up through January 2017, then shifts over to extrapolated trend from February 2017 on. Meanwhile, in Column T, past actuals are picked up from Column B and extrapolated trend values from January 2017 forward are picked up from Column R. Why do this? It all has to do with the charting that we’ll see on the “Trend” tab. Rather than have two separate lines, one for the past and one for the future, we have one line describing past & future combined; and then we’ll superimpose a yellow background to distinguish the past from the forecast.

**5-21 The 5th & final section picks up the past & projected trend**

The 5th and final section picks up past & future values and puts them all in one column. Thus, for trend in Column S, it picks up the estimated trend from Column O up through January 2017, then shifts over to extrapolated trend from February 2017 on. Meanwhile, in Column T, past actuals are picked up from Column B and extrapolated trend values from January 2017 forward are picked up from Column R. Why do this? It all has to do with the charting that we’ll see on the “Trend” tab. Rather than have two separate lines, one for the past and one for the future, we have one line describing past & future combined; and then we’ll superimpose a yellow background to distinguish the past from the forecast.

**5-22 The 5th section also picks up the past actuals & the projected extrapolated values.**

Meanwhile, in Column T, past actuals are picked up from Column D and extrapolated trend values from February 2017 forward are picked up from Column R. Why do this? It all has to do with the charting that we’ll see on the “Trend” tab. Rather than have two separate lines, one for the past and one for the future, we have one line describing past & future combined. The chart will then superimpose a yellow background to distinguish the past from the forecast.

**5-23 Finally, the 5th section calculates the difference between estimates & actuals to perform a running total of how well the estimated trend line fits actuals.**

The final column in the “Calc” tab calculates how much our running estimates vary from actuals. In June 2001, for example, the actuals were 24.674 million (CellD16, and Cell T16). The estimated trend value for the month was 25.106 (Cell P16). The difference between these two is 0.442, which is rounded here to 0.4 (Cell U16).

**5-24 The bottom of the “Calc” tab calculates annual totals for all the columns.**

Before we leave the “Calc” tab, I want to point out that there are annual totals calculated at the foot of the tab. These simply sum up the totals for each year. Note however, that the totals for some of the columns have a grey background. That’s because these are averages for each year, rather than a sum. The average Normalization Factors in 2001 was 0.983 (while the sum total would be a bit under 12). The average seasonal factors are exactly 1, and always should be, by definition. The average growth rate in 2001 is 0.0% - every month used that rate, while future years have varying rates. Note that the 1-Time Events are a simple sum, as that more closely describes their effect – while each month’s growth rate is divided by 12, the event estimates are taken whole.

And with that we’re done with describing the “Calc” tab. Again, this tab will rarely be used going forward. It’s where almost all the calculations are taking place, to be sure. But except for monthly updates for charting purposes (which we’ll later walk through), we won’t need to look at this tab again.

**5-25 The “Trend” tab is divided into three sections.**

We now move on to the “Trend” tab. This table shows the entire tab, which can be broken out into three sections. Although the Seasonal Factors Calculation is the largest section, once it is set up there is little to do with it (just like the “Calc” tab). Indeed, you may want to park it on a separate tab. I’ve kept it here out of habit, and as I find it helpful to see what’s going on down below. But the heart of this tab is at the top, with the Trend Estimates and accompanying chart. Let’s walk through each of the three sections in turn.

**5-26 The “Estimating Trend” section is the steering wheel for the model; it’s where you insert your monthly estimates of the growth rate and events.**

The Trend Estimate section is the primary driver of this trend model. This is where you input your estimates of trend, expressed in terms of the growth rate and events. It is also where you can compare how well your estimated trend line matches up with the running total for the actuals.

Please note that for the purpose of showing how this model works, we are going to proceed with some very simple figures shown as placeholders. Thus, you can see that the growth rate and events are left unchanged throughout the 2003-15 period. We will walk through updating the model in the next chapter. This image shows the entire 2001-17 period. We really don’t need to see all these years together, so going forward we’ll just focus on the 1st 4 years, from 2001-04. And, we’ll key in on the month of September 2001 to observe more closely how the model works.

**5-27 The Growth Rates are the estimates of the annual rate at which the data is growing; conditional formatting highlights changes in the rate.**

Let’s look at the Growth Rate section. Three points to make here.

1st, the growth rates are annual, i.e. each month’s growth rate is the estimate of the annual growth rate at that time. The growth rates here, which are just placeholders for the moment, indicate the sales are growing at an annual rate of 2.0% from January 2001 thru June 2002; the rate drops the following month to 1.0%.

2nd, the growth rate is conditionally formatted, with an outline & yellow fill whenever the growth rate changes; that’s why July 2001’s rate is highlighted (Cell D17), as well as our beginning rate in January 2001 (Cell C11).

3rd, the growth rate is expressed at the 1/10 of 1% level. I’ve almost always found this is the most appropriate level of precision to work at. Rarely will you be able to be more precise (& use 1/100 of 1%). And usually you can at least get the growth rate accurate to perhaps a ½% level; i.e., you can express the rate as 1.5%, or 1.0 or 2.0%; being more precise, such as 1.7%, yet alone 1.72%, risks suggesting a level of accuracy that is usually absent.

**5-28 The Starting Point is the approximate level at the outset of the period being modeled.**

The Starting Point is your estimate of what level the data is at as the model begins to trend the data. This model is for the period 2001-Present. Thus, the starting point here is for the level as the year 2001 begins. It has been estimated here at 25.0 billion (Cell A25), though we’ll later be playing with this (& the growth rates & events) when we walk through estimating trend.

**5-29 Events are the estimates of when a step function takes place, when there is a sudden shift in the level of the data. Most months do not have events.**

Events are those occasions where the level suddenly & significantly shifts up or down. Most months do not have events. Indeed, if you have an event every other month or so, you risk just following the data around wherever it’s going and you lose a sense of how the data is trending over time.

We can see that September 2001 has a 30% event input. Of course, that’s the month of 9/11, so not surprisingly the sales shot up that month. There is a -20% drop shown here for October 2001. This is just a placeholder, but it suggests that much of the September spike is a one-off, that sales soon drop back closer to their prior level. On the other hand, we see a 10% increase in July 2002. This indicates a shift that is “permanent”, an increase that is held for a long period of time.

**5-30 The Actual vs Estimate section compares each year’s totals and calculates the difference between the estimated trend and actuals.**

The Actuals vs Estimate section performs a very important function: it compares how your estimates line up with actuals. Ideally, the estimates and actuals will be almost identical. They won’t exactly match, but they should be close. I try to maintain a difference of less than one-half of one percent each year. That way I know I’m not far off in any given year. That while the estimated trend line may be above or below the seasonally-adjusted actuals for a period of a couple-few months, the trend line has been set such that other periods will be below or above in turn. I want the differences to offset one another over time, and I try to ensure this is occurring each and every calendar year. This table is my way of checking how I’m doing.

In this example, 2001 saw 307.5 billion shares sold; the estimate is at 306.4 billion, a difference of 1.1 billion, or 3/10 of 1%. Not bad, though again, the line that’s there is just a placeholder for the moment. Which explains why the differences in 2002 thru 2004 are much greater, and not acceptable. Those will be addressed and reduced as we walk thru the trend estimation process.

**5-31 The Total Actual vs Estimate table shows how actuals and estimates compare over the entire modeled time period; it also compares current year-to-date.**

In the bottom right corner of the Estimating Trend section you will find a “Total” section that compares actuals and estimates for the entire time frame. In this instance it compares the totals for the entire 2001-16 period (in Column Q). These are all the years for which we have complete years of data; if the current year is a partial year, it should not be included in these totals. The total difference here should be very close to 0, and would hopefully have a total difference that is less than 1/10 of 1%, or certainly at most 1/10 of 1%. This helps ensure that while yearly differences may be fairly slight, that there is a tendency for the differences to offset over time, such that the total differences are truly tiny, on a percentage basis. Column R shows the year-to-date totals for the most current year. As we’ll see when we walk through updating the model for January 2017, this will be a place to compare the differences for the partial 2017 year.

**5-32 While hidden here because it isn’t applicable, there is a section here for “Other Info” that may provide a useful reference.**

Occasionally, when you’re walking through the trend estimation process, you may find it helpful to have certain other time-related information at hand. Of particular note might be price changes, or marketing campaigns, or whatever. You could use this “Other Info” section as a place to have that info at hand, as reference.

**5-33 The Seasonal Factors Calculation section takes the Normalized data, and adjusts for growth & events.**

Now we turn to the Seasonal Factors Calculation section. In this section, we will start with the normalized data, then perform a series of calculations to adjust that data for the growth & events that have been input above. The exact formulas used here can be found in the “TrendModel” Excel file attached to this website. We’ll quickly walk through each of the steps, and arrive ultimately at our final Growth-Adjusted Seasonal Factors.

**5-34 We start the seasonal factor calculation procedure by bringing in the normalized data.**

We start off the seasonal factor calculation by bringing in the normalized data. We work with the normalized data, not the actuals, because we want data that has been cleaned of the calendar effect, data where every month is of approximately equal length. Here we see September 2001 is quite a bit higher than the other months. The actual sales for the month weren’t that high, but the exchange was open only 15 days that month, as it was closed for 9/11, and the following three days.

**5-35 To adjust the data for growth, we bring in the estimates for both the growth rate & events.**

We want to adjust for growth, so we begin by bringing in our estimates of the annual growth rate and events. The annual growth rate is divided by 12, in order to capture the impact of one month, such that January thru August 2001 all show monthly growth rates of 0.2%. The event estimate is picked up whole. Thus, for September 2001, we can see that 1/12th of the 2.0% annual rate has been picked up and combined with the 30% event figure. (Note that the formula is such that the amount here is slightly more than the sum of the two.)

**5-36 Next, the growth is accumulated over the year.**

Next, we take our monthly growth rates (including events), and cumulate them, so we can properly get at how much higher (or lower) the year ends than it started. We also calculate the average growth rate for the year, which is the simple average of the 12 months of growth rates. The cumulative growth rate jumps in September 2001 as the prior month’s rate is increased by 30.2%.

**5-37 The cumulative growth rates are then indexed, giving us the factors we need to adjust the data for growth and events.**

The cumulative growth rates are indexed, by adding 1 to each month’s cumulative rate and dividing it by 1 plus the average growth rate. September 2001’s index is (1+ 32%) divided by (1 + 4.6%): 1.320 / 1.046 = 1.26. These indices capture the level of growth and events that is occurring across the year, and will be the factors used for adjusting the data for growth in the next step.

**5-38 The normalized data, from above, is adjusted for growth, using the Indexed Cumulative Growth.**

The normalized data is adjusted for growth (& events) by dividing each month’s normalized sales by the Cumulative Growth Index. For September 2001, the normalized sales of 35.1 billion are divided by the 1.26 index to arrive at 27.8 billion. Note how this amount is now much closer to the other months’ amounts – the growth index has taken out the sales spike. Note also that a simple average figure has been calculated for each year – in Row 193.

**5-39 The growth-adjusted data is then expressed as an index.**

Now that we’ve adjusted the data for growth, we’re ready to zero in on determining the seasonal factors. We start by indexing the growth-adjusted data. September 2001’s growth-adjusted sales of 27.8 billion are divided by that year’s 26.3 billion average, to arrive at a factor of 1.06.

**5-40 In order to determine what years we wish to exclude, each month’s factors are averaged, and an acceptable low & high range is calculated.**

We are now ready to determine the weighted average seasonal factors – we just need to know what years we wish to exclude as they contain a month(s) that is outside the acceptable range. As before, we achieve this by calculating a simple average & standard deviation for each month, then set an acceptable range based on the Maximum Permitted Standard Deviations. For September, the average across the 16 years happens to be 1.00, and the standard deviation is 0.10. We’ve set the maximum deviations at 2.0 (in Cell R199), so the bottom of the accepted range is the average (1.00) less 2 standard deviations (or 0.80) and the high end is the average plus 2 standard deviations (or 1.21). Looking across the years, we see only one occasion, 2008, where a September factor falls outside the range; conditional formatting highlights this outlier.

**5-41 A “Data Accepted?” table is added that flushes out those years where one or more months fall outside the accepted range.**

Again, as we’ve seen before, a table is inserted that picks up whether or not each month falls within the accepted range. If it doesn’t, that month gets a “0”, and the year also gets a “0”, meaning it will not be included in the weighted average calculation.

**5-42 A weighted average is calculated, picking up all the accepted years. This is our goal: the growth-adjusted seasonal factors.**

The last step in the calculation is to determine the weighted average for each month, using those years where every month met the acceptable range. Picking up the 8 accepted years – outlined in blue - for September, we arrive at a weighted average seasonal factor of 1.01 (Cell AC209).

**5-43 Note that these final growth-adjusted seasonal factors are the set that are picked up in both the “Inputs” tab and “Calc” tab.**

This may be redundant, but I wanted to remind you that these final factors are what are being picked up elsewhere in the model. These are the “Growth-Adjusted” seasonal factors displayed in the “Inputs” tab. And they are the “Final” factors picked up in the “Calc” tab.

**5-44 We now turn to the chart that will be used to track our estimation of trend.**

We’re now ready to move on to the 3rd and final key part of the “Trend” tab: the chart showing the seasonally-adjusted data, and our estimation of that data’s trend. Again, we’ll walk through this step-by-step.

**5-45 One of the most important aspects for trending the data is the choice of seasonal factors to use.**

Absolutely central to estimating trend is the choice of the set of seasonal factors to employ. The choice is a manual input that you insert in the highlighted Cell AG1. As the note to the left (Cell Y8) indicates, a “1” means there is no seasonality, that all the factors are 1.00; a “2” picks up the Initial set of factors that we developed in the last chapter; and a “3” picks up the Final set of factors – this Final set is the growth-adjusted seasonal factors we are developing on this worksheet. And they are the same as we described on the last page, which are found near the foot of this “Trend” tab.

Here’s a rule of thumb on which set of factors to employ: if you have nothing worth working with, use a “1” to assume no seasonality; if you do have something to work with, be it an initial set of factors you’ve developed, or a final set of factors you developed for some other set of data, use a “2” here. This initial set is what you work with while you walk through the FIRST iteration of your trend estimate. Once you’ve estimated trend, you should have an improved set of seasonal factors as they will have been adjusted for growth – accordingly, use a “3” to employ the Final growth-adjusted set.

**5-46 Focusing on the chart, we start by setting the time range to the 1st three-four years of the series.**

We now turn to the chart. The first thing to point out is that we want to initially have the chart focus on just a few years of data at a time. If we tried to estimate trend on 16 years of data at once, we’d have a difficult time seeing clearly how the data is behaving. To adjust the visible time range, you simply want to right-click on the X-axis, choose “Format axis”, and then set the “Minimum” and “Maximum” for the range of dates you want to focus on. Here we’ve focused on four years, 2001 through 2004. I also include Jan 2005 so it’s clear to me that ALL of 2004 is included.

**5-47 Focusing on the chart, we start by setting the time range to the 1st three-four years of the series.**

Similarly, you will want to set a Minimum and Maximum range for the Y-axis. Here the values should be set such that you can see the top and bottom ends of the range of data you’re working with, for the time frame you’re working with.

**5-48 The Actuals are the first set of data to chart. They are colored light gray to de-emphasize them; they’re here as an FYI.**

We’ll now go through the different lines depicted on the chart. First is the Actual data, shown in very light gray. The light color is used because we won’t really be tracking it very much; it’s here mainly as reference, as an FYI. But note that it is labeled “Actual (& Extrapolated) Values”. When we later look at forecasts, we’ll see that this line extends into the future, displaying what our forecast trend looks like after we’ve re-inserted the calendar effect and seasonality.

**5-49 A dotted blue line is added, showing the seasonally-adjusted data.**

The next line on the chart is the seasonally-adjusted data. It’s shown here in dotted blue. You can see that the line is certainly somewhat smoother than the original actuals. This line depicts the normalized data that has been seasonally-adjusted using the “Initial” seasonal factors we chose at the top of the “Trend” tab.

**5-50 A thicker, solid blue line is added for the 3-month moving average of the seasonally-adjusted data.**

Recall that in the “Calc” tab we added a 3-month moving average for the seasonally-adjusted data. This was a way to help smooth the data and make it somewhat easier to read its behavior. That data is added here in the form of a thick solid blue line. I’ve emphasized this line because it is the one most used for observing how the data is trending. Why not just eliminate the dotted monthly seasonally-adjusted line? Because we’ll find the monthly is needed to better pinpoint the timing of when shifts occur in the data.

**5-51 Finally, the Estimated Trend line is added to the chart, colored in red to emphasize its distinction.**

The final line to add to the chart is our Estimated Trend line. This will be the key line, showing our estimates of how the data behaves over time, based on our inputs of growth rate & events. The line is in red, to help it stand out more. Note again that it is labeled the “Estimated (& Extrapolated) Trend” line. We’ll see it stretch out into the future when we get to the far end of the chart. At the moment this trend line is not tracking very well with the seasonally-adjusted data. We earlier used “placeholders” to get the line started. In the next chapter we will walk through playing with the growth rate & event estimates so that this trend line much better corresponds with the behavior of the seasonally-adjusted data.

**5-52 The heart of the estimation trend analysis is performed at the top of the “Trend” tab.**

That’s it for walking through how this model is structured. We saw how the “Inputs” is where we brought in our monthly data and normalization factors. We saw that the “Calc” tab is where most of the calculations are performed, and we’ve now seen how the “Trend” tab works. Going forward, our focus will be working with the top of this tab, where in one view we are able to see our estimates of the growth rates & events, how our estimates compare with the actuals, the set of seasonal factors we’re using, and a chart displaying the key info we wish to track. This page shows the entire 2001-17 time period; when we start estimating the trend, we’ll hide the later years so as to better focus on those years being trended.

**Chapter 6: Estimating Trend**

The last chapter laid out the structure of the model to be used for developing the final, growth-adjusted seasonal factors. In order for the final seasonal factors to be “growth-adjusted”, one needs to establish the growth rate, and events, that drive the data over time. In other words, one needs to estimate trend. This chapter walks through the process of estimating trend, an iterative and somewhat subjective procedure that relies on your estimates of growth rates & events.

**6-2 This chapter will walk through the manual process of estimating trend. It will also describe how to modify and update trend estimates.**

This chapter will walk you through the procedure for estimating trend, using the NYSE sales volumes as example. We’ll start back in 2001, and walk through 2016. We’ll focus though on the 1st few years to get a “hang” of how it works, then more quickly go through the subsequent years, pointing out notable exceptions or specific suggestions on handling certain traits the data may possess.

Once the trend has been estimated, you will have a set of final, growth-adjusted seasonal factors. You’ll need to review the period again to see if any further tweaking of trend is called for.

We’ll then walk through updating the model, for when a new month of data has come in.

And then we’ll conclude with some final notes on the whole procedure. Let’s get started.

**6-3 Development of the “Final Seasonal Factors” involves adjusting the data for growth & events.**

As mentioned, in order to develop the final set of seasonal factors, you need to adjust the data for growth and events.

**6-4 When you estimate growth & events you are estimating trend.**

In effect, “estimating growth rates and events” is the same as “estimating trend”. The trend is defined as the combination of the growth rate, (or slope of the data), and events (or the step functions). We’ll walk through how you can estimate these for any set of data, using the NYSE sales as example.

**6-5 The development of growth-adjusted seasonal factors, and the estimation of trend, is a feedback loop.**

At the risk of belaboring the point, I want to make it clear that as you estimate trend you are developing & tweaking your growth-adjusted seasonal factors. But these factors then help inform how the data is trending. It’s a feedback loop, with each informing the other. We’re not going to go back and forth with them endlessly, but we will walk through one full round: we’ll start with the initial seasonal factors, estimate the trend to arrive at growth adjusted seasonal factors. And with these growth-adjusted seasonal factors we’ll revisit the trend estimates again, revising them where needed. Of course, if you don’t have a set of initial factors to start with, there will likely be a need for much more revision. Once the trend has been re-estimated with the growth-adjusted seasonal factors, we’ll have in hand our final set of seasonal factors and seasonally-adjusted data.

**6-6 There are a number of “Rules of Thumb” to bear in mind as we walk through the process of estimating trend.**

As we walk through the process of estimating trend, it will be useful to keep in mind a number of “rules of thumb”. Like all such rules, they are meant to be broken on occasion, but generally they’re helpful guides to the procedure.

1st, try to limit trend changes (by revising the growth rate or inserting an event) to 3-4 times per year. This is an effort to avoid just following the bouncing ball; there will always be some level of volatility; you’re trying to capture the key changes when they occur, as well as reflect the overall growth trend for the metric. I’ve found that a change in any one direction is generally not worth pursuing unless either it lasts 3 months or more in length, or it is so large that you have to pursue it, even if for just one month, in order to get your annual total estimates and actuals to closely match. If changes are inserted too frequently, you can lose sight of the underlying trends. It’s a variation on the theme of “seeing the forest for the trees”.

2nd, make sure the estimated trend values each year come close to the actuals (by reviewing the annual comparisons in Rows 51-54).

3rd, sometimes, in order for the actuals and estimates to come close, it may be necessary to follow a “one-off” type of event, like a sudden one-month spike.

4th, events are usually much more common than changes in growth rates, so most of the trend “tweaking” will likely involve playing with events rather than growth rates.

5th, you will likely want to chart and track only a few years at a time, and separately look at the entire period as a whole. Thus, with the NYSE data, we will estimate trend for only 3-4 years at a time, then look at the entire 2001-16 period after completing the trending.

And finally 6th, and perhaps most important, the trend estimates are a best guess, based on how the data appears to behave and, significantly, your own knowledge of what may be causing the data to behave as it does. This is as much an art as it is a science. No one knows what the true trend is, you can only make your best educated guess at it.

**6-7 Ideally, the trend estimate inputs, the actual vs estimate comparison, and the chart, all just fit onto a single screen view.**

As we walk through the process, we’ll work with all the key info available in a single view: growth rate & event estimates, the annual actual vs estimate comparison, the set of seasonal factors in use, and the chart showing how everything looks over time. Note that here we have limited that “view” to the 1st 4 years of data; the columns for the other years of inputs are hidden, and the X-axis for the chart shows only 2001 thru 2004.

**6-8 And we’ll start with a clean slate, with growth set at 0%, no events, no Start point showing on the chart, & with our using the “Initial” seasonal factors.**

Let’ start with a clean slate, with growth rates and events at zero, and the start point lowered so we can’t see the red “Estimated Trend” line. Note that we begin with the “initial” set of seasonal factors in use (Cell AG1).

**6-9 At first glance, it looks like this period can be described as basically flat, with a step function after 9/11 & another in mid-2002.**

When I first look at a given time period, I want to determine what the broad themes are over the period. Is there a general upward or downward trend with a couple of bumps along the way? Is there a point at which the level of activity clearly and significantly shifts up or down? Is it fairly cyclical and requiring a number of shifts to capture?

The 2001-04 period appears to fit that latter description. Broadly, there are three levels of activity, roughly drawn in the chart here: the 1st lasting until 9/11; the 2nd for the ensuing year, and the 3rd, from summer 2002 thru most of 2004.

**6-10 But such a depiction leaves many significant gaps between the estimated trendline and the actuals.**

The problem with this interpretation is that the trendline would frequently have large gaps from where the actuals are. It’s evident that we’ll need to insert several step functions, and perhaps a change or two in the growth rate, in order to better reflect the data behavior over the period.

**6-11 Inserting a few more step functions and a couple of changes in growth rate appears to align better with the data behavior.**

Looking more closely at the data out in 2003, it looks as though it was broadly declining over the period, after the spike mid-2002. It also looks like having trend drop during the middle months of 2001, and having it bump up for the 1st half of 2004 would be appropriate. Perhaps something a bit closer to some of the step functions inserted here will do the trick. Let’s start with something like this in mind, and walk through the actual trend setting.

**6-12 To capture this overall interpretation, we start with the growth rate remaining set at 0.0%, and the Start point bumped up to 25.5 B.**

We begin by setting the Start Point at 25.5 B, accomplished by inserting that amount in the highlighted Cell A25. A flat trendline should appear, at the 25.5 billion level.

If it does not appear, be sure you are hitting the “F9” Calc button, one of the function keys at the top of your keyboard. I generally have all my models set at “Manual Calc”, which requires the “F9” key be tapped each time you want Excel to update the calculations. Why do I do this? Because often, especially as we walk through this trendline estimation process, I want to be able to insert several changes before I observe their combined impact. Or perhaps I make changes in a different part of the worksheet and I want to return to a chart to see what happens when those changes are implemented. Using Manual Calc is a very useful way of enabling one to do that. If you still prefer to have Excel always automatically update the calculations, you can do so by changing the Calculation setting, done by going to “Formulas” in the menu bar, then “Calculation Options”, then clicking on “Automatic”.

**6-13 It looks like we would do well to drop the trendline in May 2001, to a level just above 24 B; we do this by inserting a “-5.0%” event in May.**

Next, let’s drop our trendline in May 2001. I tried a couple of different numbers, and have settled here on -5.0%, inserted in highlighted Cell C35 – it’s an event, and it occurs in May 2001.

**6-14 With 9/11, the level jumps to around 28 B, achieved by inserting a +15% event in September.**

Then 9/11 happens, and sales jump enormously. Before we try to get September 2001 itself right, let’s get to the level we want our trendline to be at for the months following. Inserting a +15% event seems to be about right.

**6-15 But the 9/11 spike was huge, so the September bump is increased to +45% to match, then a -20% drop is inserted in October, to get back to the 28 B level.**

Of course, the September spike is enormous, so large that we want to capture it separately. Note that we can’t always go after every spike, or else we’d have the trendline bouncing around every other month or so. But this spike merits pursuing on its own, for three very important reasons: 1st, the jump is substantial, well above 10% or so from where the general trend is at. 2nd, after this spike occurs, sales are at a different level than they had been prior to the spike. And 3rd, there is not an equivalent offset in the opposite direction in the month or months immediately following. Often one month can come in very high, say 10 above where it had been previously, and the subsequent month may come in about 10 lower than the prior level. In such an instance, perhaps there is evidently an accounting adjustment going on, that activity is inadvertently being attributed to the 1st month rather the 2nd. Whatever the reason, offsetting activity generally does not merit being pursued by the trendline.

But this is clearly a case where we want to spike sales up to get close to our seasonally-adjusted amount for September. This is accomplished by inserting a +45% event for September; we can then approximately drop down to the ensuing level by inserting a -20% event for October.

**6-16 This looks pretty good for 2001, but notice that there’s a 0.9% difference between actuals and estimates for the year. How can we lower this?**

This looks like a pretty good capture of 2001 activity, but notice that our total estimate is still 2.8 billion, or almost 1% higher than actuals (Cells C53 & C54). It would be nice to try to get this gap to at least something less than a positive or negative 0.5%. We’ll need to do some tweaking.

**6-17 The difference is lowered by further reducing the May event, from -5.0% to –5.5%.**

We can see that in the May-August 2001 period, three of the four months are below 24 billion, only July is above 25 billion. A slight lowering of the trendline seems in order, done here by changing the May event, from -5.0% to -5.5%. Now we have a reasonably acceptable gap of 1.7 billion, or about 0.5%.

**6-18 2002 starts fine. A +14.0% event is inserted in June to capture much of the increase; growth is dropped to -8.0% in July to align with growth thru 2003.**

Now we move on to 2002, and as it works out, 2003 as well. I’m thinking a line that goes through the two green-circled areas would fit the data nicely: it seems to capture the general trend over the period; the two gaps between the circles nicely offset one another, with the early 2003 high gap about evenly matched by the mid-2003 low gap. The “high” gap I’m referring to here is how the trendline is above the actuals in early 2003, and the “low” gap is how the line is below the actuals in mid-2003.

I played around with the numbers a bit, but was able to get this nice-fitting line by making the two changes highlighted in the table here: a 14.0% event in June 2002 (Cell D36), along with a change in the growth rate to -8.0% in July 2002 (Cell D17). Note that I first worked on getting the slope right. That is to say I played with the growth rate to make sure it nicely paralleled the overall decline I wanted to capture. Once I had this slope right, by using a -8.0% growth rate, I then played with the June event until I was able to get that trendline to pass through the two circled areas in the chart, accomplished by an estimated event of +14.0%. Note that this 14% bump very nicely brings the trendline just slightly below the level of our seasonally-adjusted line: in June 2002, the seasonally-adjusted actual is just above 32.0 billion, while the trendline is just below that level.

**6-19 Clearly we need to capture the July spike, done here by inserting a +30.0% jump in July. But where does August’s precise-looking -23.1% come from?**

July 2002 is another instance of a monthly value that is well above all the others. We’re going to want to capture it, or otherwise we will have a trendline that in total will fall well below the actuals. So I again played with the numbers and found +30% nicely brings the trendline up to the seasonally-adjusted actual level for July. I then want to bring our trendline back to the level was at previously. In fact, I want to bring it back to precisely the level that corresponds with the line that I had before I inserted the July increase. The value “-23.1%” does just that; but how did I come up with it?

**6-20 Employing the Event Reversal Formula ensures the trendline returns to the identical level it was at before.**

There is a VERY useful formula I frequently employ whenever I want the trendline to pick up where it left off, to have the trendline return to the precise level it would have been at had the event not been inserted. That “Event Reversal Formula” is highlighted here. It is simply 1 divided by “1 plus the Event %”, minus 1. In practice the formula here would be 1/(1+30%) – 1, which would be 1/1.3 -1, which is 0.769 – 1, which is -0.231, or -23.1%. Hopefully the highlights here make clear that the formula worked in such a way that the line goes thru August at the same level it would have done had the 30% event in July not been inserted. The nice thing about using a formula for August, in Cell D38, is that should I later adjust the July bump, the August reversal will automatically adjust along with it.

**6-21 The July event, and reversal, help bring the total estimate and actual for the year to nearly match; but it’s still high – for the second year in a row.**

I like the outcome, but our actuals vs estimates are still off. Indeed, not only is 2002 high by almost 1%, but 2001 was also high, so we’re clearly still setting a trendline above where we’d like it to be. More tweaking is called for.

**6-22 We begin addressing the 2-year highs by further dropping Oct 2001, reducing it from -20.0% to -21.0%; the Jun ‘02 event must be raised to +15.5% to offset.**

The first thing I did was to drop the post 9/11 level slightly; by changing the October event from -20% to -21%. That brought the later trendline down, so I had to bump up the June 2002 event: from +14.0% to + 15.5%. That fixes 2001 very nicely, with the gap now less than a billion, at just 0.2%. 2002 is still a bit high but certainly much better.

**6-23 We insert a -6.0% drop in Dec ‘02, later correcting it in Mar ‘03. The 2002 totals almost match perfectly, but now the 2003 estimate is too low.**

This is an admittedly subjective process. One could be content to leave our line as it is. But I’d like to try to get a better fit here, so I’m going to do that by inserting two more events – the 1st to capture the 3-month drop in sales from December 2002 to March 2003, and the 2nd to capture the 3-month bump in sales, from May to August 2003. We begin by inserting a -6.0% drop in December 2002 (Cell D42), and reverse it using our “Event Reversal Formula” in March 2003. You can see how the line in March 2003 picks up where it left off before the December 2002 drop was inserted. And you can see that the Actual vs Estimate totals are now almost identical for the year 2002. 2003 is too low, but that will hopefully be corrected when we insert a mid-2003 bump.

**6-24 So we insert another event, a +5.0% increase for 3 months, starting in May ‘03. Now the 2003 totals are close as well.**

Inserting a 5.0% increase in May 2003 (Cell E35) looks about right; it is later reversed, using the Event Reversal Formula in August (Cell E38). Now the Actual vs Estimate gap total for the year is again down to a very low 0.2% (Cell E54). On to 2004…

**6-25 A +14.0% event to start 2004, combined with a -8.0% drop in June, captures the spike. The -8.0% underlying annual growth rate is still left unchanged.**

We start off 2004 with a +14.0% in January (Cell F31); we then drop June 2004 by -8.0% (Cell F36) to get to the approximate level sales in mid-2004. So far, we’ve left the growth rate unchanged, at -8%.

**6-26 The level in mid-2004 is almost identical to what it had been at the end of 2003, suggesting the -8.0% growth rate should perhaps be changed.**

Notice how the level before the January 2004 jump is almost exactly the same as the level after it ends, in June thru September 2004. This suggests that instead of continuing the -8.0% growth rate over this period, we should flatten it instead, starting in January 2004.

**6-27 Here the growth rate is changed in January to 0.0%, and the event is dropped, from +14.0% to +12.0%.**

So we flatten the growth rate by setting it to 0% in January 2004 (Cell F11). We then drop the January event, from +14% to +12% (Cell F31). And we get our trendline back to the December level by leaving as is the Event Reversal Formula in June 2004 (Cell F36).

Is this what is really going on??? Do we know that this is the “true” trend growth rate and performance over this period? Of course not. This is a best guess. But it already looks like we are seeing a pattern of the market experiencing various ups and downs, only to have activity return to its prior level. We saw it at the start of 2003, in mid-2003, and again here in mid-2004. It’s a pattern I’m inclined to continue pursuing, until I see otherwise.

**6-28 To finish off 2004, a +11.5% spike is inserted in Oct, and growth is shot up to +25% the next month. The ‘04 total difference is close, & offsets ‘01 & ‘03.**

We now finish off 2004 by inserting an 11.5% increase in October (Cell F40), and by bumping the growth rate the following month to a heavy 25% (Cell F21) to align with the new slope. Two points to make. 1st, I would have preferred a 12.0% bump in October, for that would have exactly matched the January bump, and meant the level in October would be in line with where it had been from January thru May. But as the trend increases from there it’s not as meaningful that it precisely align. Secondly, the 25% growth rate was applied not only as that is how the data is sloping at the end of 2004, but because I snuck a look at 2005, to which we’ll turn next.

Before leaving this, note that the 2004 actual vs estimate comparison is low by just 0.3% (Cell F54), and that the 1.1 billion difference (Cell F53) exactly matches the sum of the 2001 & 2003 differences (Cells C53:E53).

**6-29 That’s a lot of work, and is admittedly quite subjective. But it aligns well with trend, the total ‘01 – ’04 difference is small, & shifts are limited to 4 per year.**

Well, that was quite a bit of work. And that’s the nature of this process. It’s very iterative, and you’ll often find that when you look at a trendline you’ve set you’ll want to tweak it further. Whatever you do, you just want to make sure it makes sense, that you’re not inserting too many shifts in the trendline. And perhaps most importantly, that you are able to explain WHY the data shifts when and how it does. You’re going to want to be able to point to changes you’ve implemented on a certain date, or major moves made by your competitors, or some outside event (like 9/11) that explains the behavior and is consistent with the level of change you are inserting with your growth rate and event estimates.

Frankly, when I look at these first four years, there a few more changes than I would like. The growth rate was changed 4 times (including the initial rate of 0%), and there are no fewer than 13 event values. However, about 5 of those events are event reversals, so you could think of the change and its reversal as one event instead of two; that helps. Nonetheless, about 4 changes a year is as much as I’d like to see. Invariably there will be the odd year that might see 5 or 6 changes or more, but that should be the definite exception, and hopefully you find that 2-3 changes at most each year prove to be the rule.

**6-30 Except for substantial volatility in late 2007, the 2005-07 period shows remarkably steady growth.**

From here on, let’s not walk through all the details of how the trend is developed. Instead, we’ll focus on any observations that may prove helpful when it comes to approaching the trending of your own data.

This figure shows the 2004-07 period. The Y-axis was expanded to accommodate the much higher numbers in 2007, and the legend was moved to the left so the lines are all visible. Note that the X-axis was revised to also include 2004, and that the Growth Rate & Events columns show 2004, as well as 2005-07. You want to continue to show the last year of the prior period you were working on, for it may be necessary to go back and slightly revise it. That was the case here, for I found I needed to bump up the growth rate – from 25% to 28% (Cell F21). That in turn called for dropping the Oct 2004 lift from 11.5% to 11.0% (Cell F40); that combination only increased the actual vs estimate difference by another 0.2 billion (Cell F53).

It’s unusual to have such a high growth rate (28%), sustained over a 7-month period. One could try to insert an event to get the trend level up, but I chose to use growth alone as the values increased so steadily over the period. After that, growth “slows” to about 11%, a rate that nicely fits the data from mid-2005 through the end of 2007. The line is a remarkably tight fit with the data for the next 2+ years, with only a couple of 4-6 month dips required. It’s the 2nd half of 2007 that starts going crazy. Perhaps not coincidentally, the Great Recession started officially in Q4 2007. I inserted a 2-month surge (Cell I37:I38), and dropped it back down again (Cell I39). But clearer insight on how to handle this will probably come with looking at 2008-on.

**6-31 From mid-2007 thru mid-2009, sales swung enormously. The trendline attempts to pick up the swings while avoiding using too many changes.**

Not surprisingly, 2008 and 2009 are quite the roller coaster of activity. The financial crisis sparked by the Lehman bankruptcy on September 15, 2008 causes unprecedented stock sale volumes that roll into October. But activity quickly relatively subsides as the year ends. February-May 2009 see another smaller surge, and after that the sales volumes really decline, with only a moderate surge in spring 2010.

It’s obviously very difficult to decide how to characterize and trend the data when it behaves with such volatility. I didn’t want to chase after every big spike or plunge; nor did I want to just try to find a single shift or two that would construct a line that captured the period’s averages. I’ve tried to find some compromise between the two with the events inserted here. Hopefully you aren’t working with data susceptible to this kind of erraticness; this is clearly the nature of the subject here – stock market sales volumes.

Note that I shifted the growth rate to a -10% in December 2008 (Cell J22). I did NOT choose that figure because it so closely mirrored the prior one; I chose it because it seemed to so well fit the general trend from here on, especially the trend from mid-2009 thru the end of 2010. It’s hard to tell, but I applied the Event Reversal Formula in June 2010 (Cell L36), to offset the May 2010 increase (Cell L35), and in August 2010 (Cell L38), to offset the April 2010 increase (Cell L34). By applying two reversals to two lifts, I am able to have the line in August 2010 pick up from where it had left off in March 2010.

Again, this is a tough period to capture. This is an instance where the 3-month moving averages seem to be very useful, for the trendline shifts here try to follow the 3-month averages, while maintaining small differences in the annual actual vs estimate comparison.

**6-32 Sales continue to plunge thru 2013, despite another short spike in the summer of 2011. But by the start of 2012, stability appears to have returned.**

2011-13 shows a continued slide in activity, but one that definitely seems to show signs of leveling out by the start of 2012. August 2011 was another spike of a few months’ duration, but otherwise sales volumes are not only down but experience much more steadiness. Almost boring, if you will, after all the upheaval of the prior 5 years. I normally wouldn’t “chase” the smaller spike that occurs in June 2013, but I found I needed to go after it to achieve a good match in actuals vs estimates. It still was about a 14% bump (Cell O36), and importantly, it precedes a drop of -18.0% to an even lower level in July 2013 (Cell O37).

**6-33 Sales volumes bottomed out in late 2014, and have grown in spurts since.**

We now finally bring our trendline up to date. (Recognize that this is being written in January 2017, so obviously to the degree you’re reading this later, that much more history will have been added.)

Sales see another spurt of activity in early 2014, but drop down even further in May. But then sales take off in the fall of 2014, and continue to show impressive step increases from there, culminating in the short-lived heights of over-30 billion at the start of 2016. Of course, these “heights” are far below the 40-75 billion levels experienced in 2007-09. But the levels notably drop back down again and seem to hover around the 25 billion level fairly steadily for most of 2015 and 2016.

Note that I switched the growth rate from a -3% to a +1% in October 2014. This corresponds with the overall trend in sales, which were declining up till then, but have been growing slightly but somewhat steadily since. I might just as well have used 0% as the growth rate, for it is very difficult to discern whether or not there has been actual “growth” in sales activity since the beginning of 2015. Again, this is what makes tracking something like stock market activity so tricky – it’s very hard to determine what the UNDERLYING trend is. This is especially the case for me as one who has not had occasion to work in past with investment clients who could speak with authority on what and why sales behave as they do. Again, hopefully it is much easier for you to be able to identify what drives the behavior of your specific data.

**6-34 Estimates vs Actuals matched almost exactly for the total 2001-16 time period.**

It’s very nice to see that the total 2001-16 actual vs estimates are almost identical (Cells Q58:Q61). This should not be surprising really, for I was always endeavoring to not only have each year match pretty closely, but to also avoid a period of several years where the totals would be consistently higher or lower.

**6-35 For the entire 16-year period, there are a total of 60 growth or event changes, close to 4 per year.**

It’s worth noting how the growth rate and event change counts performed over this entire 16 year 2001-16 time period. My preference is to try to keep changes to about 2-3 per year. I don’t feel too guilty about averaging almost 4 for this time frame, given the nature of what is being tracked here. It really isn’t that surprising that so many shifts were needed to closely match the trend of the market’s sales, but not too closely.

**6-36 The new weighted average seasonal factors are similar to the initial, but June is higher, and the average relies heavily on the later years.**

Now that we’ve completed our first attempt at carefully trending the data, we can review how the new set of seasonal factors looks. We head down to the bottom of the worksheet and can see from the chart that the “Final” set looks similar to the “Initial”. However, it is somewhat disappointing that while all but one of the 2010-16 years are used toward the weighted average, only 3 of the prior 9 years meet the criteria, and the criteria is set at a generous 2.0 maximum allowed standard deviations (see Cell Z199). However, this seems reasonable given how volatile the market behaved, especially in the 2007-09 time frame.

It’s also noteworthy how much the June factor increased (from about 1.02 to 1.05). What caused this? Recall that when we looked at Holiday Factors, we developed special factors for the 3rd Friday of the ending month of each quarter (March, June, September, December), as well as an additional factor for the 4th Friday in June. I chose to not pick up these factors when I normalized the data for use in this Trend Model. I did so in order to avoid a bit of confusion over month lengths, and year lengths, increasing over time – that is, these factors lengthened the year, and the factors have been getting higher and higher over time. For instance, until 2004, the 4th Friday in June had no apparent impact, but since 2012, the factor has averaged 2.6, implying an extra 1.6 days being added to June.

**6-37 Seasonal factors are split out into two 8-year periods, resulting in a significantly lower June factor in the earlier period.**

Because the 4th Friday in June is so significant, especially in recent years, and because the average relies heavily on the more recent years, it would seem we would do well to split out the period from which the seasonal factors are drawn. That instead of having one set of factors for the entire 2001-16 period, we split it into two shorter periods. That’s what has been done here, with averages, standard deviations, and the accepted range all being applied to each of the 2001-08 and 2009-16 time frames.

Each period now has 6 of their 8 years applying towards the average factors. June (and December) are notably higher in the later period, while curiously January & October are higher in the earlier period. We’ll use these two sets of factors when we seasonally adjust the data, which we’ll turn to next.

**6-38 The two new sets of seasonal factors are added to the table of seasonal factors found in the “Inputs” tab.**

With our two sets of seasonal factors that have been adjusted for growth and events, we are now ready to modify the trend. We do this by applying the new factors to the calculation of the seasonally-adjusted sales. 1st, we need to add the two new sets of seasonal factors to the “Inputs” tab. Previously, the Seasonal Factors displayed here picked up one set of factors alone from the “Trend” tab. Now, the table still picks up that first set, and labels it for the 2001-16 period (Cell I10). Columns for the other two periods were added.

**6-39 Formulas in the “Calc” tab are updated to pick up both new sets of seasonal factors.**

Next we turn to the “Calc” tab, where we need to revise the “Final” seasonal factors so that they pick up the appropriate set of factors. Accordingly, the formulas for January thru December 2001 are revised to pick up that set. From 2002-08, formulas are simply picking up the prior year factors, so those years are taken care of. But from 2009-on, we need to insert formulas in the 2009 months that pick up the 2009-16 factors from the “Inputs” tab. That’s it; now the model is all set for modifying the trend, using our new “Final” two sets of seasonal factors.

**6-40 To start modifying the trend, we go back to having the original 2001-04 period on display.**

Modifying the trend will involve updating the “Seasonal Factors in Use” to pick up the “Final” factors. So we go to the “Trend” tab, revise the visible columns and chart to display the 2001-04 time period. Note that the “Initial” set of factors is currently in use.

**6-41 The “Seasonal Factors in Use” is revised to pick up the “Final” set of factors; the seasonally-adjusted data changes, though not by much.**

To update the “Seasonal Factors in Use” we insert a “3” in Cell AG1; this picks up the “Final” factors. You have to go back and forth a bit, but you can see that using these new factors results in a slightly different line for the seasonally-adjusted data.

**6-42 So long as the “Seasonal Factors in Use” is set on the “Final” factors, the seasonality measure is “live”, changing with every revision of growth & events.**

We’re now going to need to tweak our estimate of growth rates and events. What’s going to be different is that every time we revise estimates of growth & events, the seasonal factors will change. The “Final” set of factors is “growth-adjusted”, and revising the growth rates & events revises the seasonal factors. In effect, we’ve now gone “live”. However, unless a major modification is made to growth or events, it shouldn’t change the seasonal factors much at all.

**6-43 Event estimates were slightly modified to better capture the trend in the seasonally-adjusted data.**

There wasn’t much to do here, but I bumped up the September 2011 spike (from 45% to 47%) and had it drop from -21% to -22% the next month. I also tweaked the bump in mid-2002, and the drop in mid-2004. Nothing much, just tweaking.

**6-44 The growth rate and event estimates are modified for the entire 2001-16 period.**

The tweaking is continued for the entire 2001-16 time period. With the NYSE sales, the alterations were minimal: events or growth rates were changed by perhaps a half or full percentage point or two, from time to time, as needed. An eye was kept on how closely the actuals and estimate totals compared each year. The final result here again has the total 2001-16 comparison showing a very close match (Cells Q60 & Q61).

**6-45 We now turn to updating the model, how to handle new monthly data as it comes in.**

Now that we’ve walked through how to estimate trend, and modify it, we turn to updating the trend. If you’re dealing with monthly data, as we have in this example, at the start of each month new data will come in and the model will require updating. Here we can see that the January 2017 data has now become available (Cell B203); the formulas in the “Inputs” tab ensure that the Model file automatically updates.

**6-46 The “Calc” tab shows the new actuals coming in. The “Seasonally-Adjusted Data”, “Estimated Trend”, & “Forecasts” all need to be updated.**

We now go to the “Calc” tab, and observe that the formulas in the “Actual & Normalized Data” section A, all update. But we’ll need to update formulas in the “Seasonally-Adjusted Data”, “Estimated Trend”, & “Forecasts” sections. Accordingly, the formulas in K202:L202 are copied down to K203:L203, to update the “Seasonally-Adjusted Data”; the formulas in O202:P202 are copied down to O203:P203, to update the “Estimated Trend”; and the formulas in Q203:R203 are deleted, so that the first set of “Forecasts” start with Row 204 (February 2017).

**6-47 Updating the “Calc” tab can also be accomplished by simply copying down all the formulas from the prior month.**

An alternative way to update the formulas in the “Calc” tab is to copy down the entire row, from Column A thru Column U. However, be aware that if you do that you have to be careful in the month of January that you revise the formulas in Columns M & N, in order that they correctly pick up from the “Trend” tab. Thus, the formula in Cell M202 is “=Trend!$R22”. When you copy that formula into Cell M203, you’ll need to change it from “=Trend!$R23” to“=Trend!$S11”. Similarly, the 1-Time Events formula in Cell N203 will need to be changed from “=Trend!$R43” to“=Trend!$S31”.

**6-48 After updating the data, and expanding the chart to show 2017, a number of further tweaks in growth & event estimates seemed appropriate.**

After updating the data for Jan 2017, I shifted the years covered by the Trend chart to include 2017. I also ended up making a number of small tweaks to the growth rate & event estimates, the largest perhaps being the choice to capture the 1-month spike in August 2015 (Cell Q38). In the wake of the inauguration and new administration, I thought it appropriate to expect a bump in sales for a few months. Accordingly, I bumped sales in February, and have them drop back down in July. Time will tell just how wrong my forecast proves to be.

**6-49 We can see what the entire 2001-16 period looks like. After a long & marked decline, activity has been slowly rising in recent years.**

I inserted a larger overall trend chart in the “Trend” tab to be able to observe the entire 2001-16 time period at a glance. It’s obviously difficult to clearly see the specific month-to-month gyrations, but it does give one a good sense of the overall picture. Such a look can be very useful to get a clearer perspective on how any given metric has behaved over the long run.

**6-50 Here’s a quick review of the “Rules of Thumb” to follow as you go through the process of estimating trend.**

Now that we have walked through an example of estimating trend, it will be useful to review again some “rules of thumb” to keep in mind during the process. Again, like all such rules, they are meant to be broken on occasion, but generally they’re helpful guides to the procedure.

1st, try to limit trend changes (by revising the growth rate or inserting an event) to 3-4 times per year. You want to avoid just following the bouncing ball; there will always be some level of volatility; you’re trying to capture the key changes when they occur, as well as reflect the overall growth trend for the metric. Changes in any one direction are generally not worth pursuing unless they either last 3 months or more, or as we occasionally saw with the NYSE sales spikes, they are so large that you have to pursue it, even if for just one month, in order to get your annual total estimates and actuals to closely match.

2nd, total estimated trend values each year should come close to the actuals.

3rd, as we just mentioned, in order for the actuals and estimates to come close, it may sometimes be necessary to follow a “one-off” type of event, like a sudden one- or two-month spike.

4th, events are usually much more common than changes in growth rates, so most of the trend “tweaking” will likely involve playing with events rather than growth rates.

5th, you will likely want to chart and track only a few years at a time, and separately look at the entire period as a whole. Thus, with the NYSE data, we estimated trend for only 4 years at a time.

6th, and perhaps most important, the trend estimates are a best guess, based on how the data appears to behave and your own knowledge of what may be causing the data to behave as it does. This is as much an art as it is a science. No one knows what the true trend is, you can only make your best educated guess at it.

**Part III: Trending Daily Data**

We are now going to look at seasonally-adjusting data on a DAILY basis. Many companies track and follow certain key information every day. These are the organizations where reports are issued on a daily, or perhaps weekly basis. Such reports are not possible unless data is available on a daily basis. And if reports are being produced daily or weekly, there is clearly interest to know how you’re doing now – today, this week – not just how you’re doing for the month overall. Given the greater sense of immediacy & urgency, you’re going to want not to just retrieve the data itself; you’re presumably going to want to interpret it as well, to be able to provide a knowledgeable explanation of how you are doing. And to do that, you’re going to want to seasonally-adjust your data, on a daily basis.

In this Part III, we will describe how to trend daily data. If your organization doesn’t capture daily data, you can skip this section. Chapter 7 describes how to develop Day of the Month (DOM) factors that reflect the pattern across the month. Chapter 8 then goes through the process of trending daily data across time.

**Chapter 7: Day of Month (DOM) Factors**

So far, we have produced indices that measure how relatively busy you are across the week: the equated day factors. And we’ve measured how activity varies in and around holidays. We now turn to how activity varies across the calendar month. Often, there is a fairly distinct pattern across a given month. Sales, for example, often peak at month-end. You’re going to want to be able to adjust your data for the known & measurable cycles that occur across the calendar month. For that you need to develop factors for each day of the month, what will be referred to here as Day of Month Factors, or just DOM Factors, for short.

**7-3 Companies that track & follow data on a *daily* basis will want to determine day-of-month (DOM) factors that measure pattern across month.**

When we look at the NYSE’s daily sales activity across 2016, we can see there is a spike at the end of each month. Of course, we can also see that there is a fair amount of volatility from day to day, even though we’ve normalized the data, adjusting it for the day of the week and holidays.

**7-4 Day of Month Factors identify the comparative level of activity across the calendar month; (e.g. how busy is 1st day of month compared to average?).**

What are DOM Factors? They are indices that measure the level of activity by day of month, as compared with the average daily activity. Thus, in the hypothetical example shown here, the 1st day of the month has a DOM Factor of 1.05, meaning it is about 5% busier than the average day of the month. In this example, the factors slowly decline throughout the month, then start to increase as month end nears. The final day of the month is clearly the busiest.

**7-5 Why determine day-of-month (DOM) Factors? DOM factors will enable you to obtain more accurate measurements of performance & event impacts.**

Why do we measure DOM Factors? Given this is a fair amount of effort, and produces results that will likely be “messier” than monthly data, it’s an important question to ask.

1st, you want to be able to provide an informed answer to the question “how are you doing?”. When you’re tracking, AND following, data on a daily or weekly basis, there is clearly interest in knowing how things are going. Being able to adjust for the day of the month will help ensure you provide a more accurate reply.

2nd, when you have DOM factors, and you observe a sudden shift in the data, you can be more confident that the shift is genuine and reflects some change that is occurred. And of course, you’ll be able to more accurately measure that shift.

3rd, companies that closely follow daily data often do so because change is happening with enormous frequency. In such an environment, it is often the case that the ability to be able to measure the impact of an event, such as a Marketing campaign, is there for only a few weeks. The window soon closes, for some other event soon occurs which will mask the impact of the original event. You might start a big marketing campaign on the 10th, and a major competitor, or even your own firm, suddenly introduces a drastic price cut on the 20th. When such a follow-up event occurs, it becomes much more difficult to measure accurately how much sales changed due to the Marketing campaign versus the price cut. And therefore, you’re going to want to able to try to estimate the impact comparing activity before the campaign launch with the activity from the narrowly-ranged dataset of the 10th through the 19th of that month. Not much to go on, but you have to do your best with what you’ve got. In such an instance, you will want to make sure you’ve appropriately adjusted for how much busier or quieter you are simply because of what day(s) of the month you’re examining.

**7-6 This chapter will walk through the process of estimating DOM Factors.**

This chapter will walk through the process of estimating the DOM factors. After this introduction we’ll look at how you manipulate your data so that every month examined is of equal length, even though that’s not the case with the calendar. Then we’ll walk through developing DOM Factors, by calendar month. And then we’ll review those results and draw some concluding observations.

**7-7 A central challenge with developing DOM factors is that months have different lengths. Even the same calendar months in different years can vary.**

One of the main challenges with developing and applying DOM factors is that months come in different lengths. Obviously the calendar itself has 7 months of 31 days, 4 of 30 days, and 1 that’s 28 or 29 days in length. And that’s if you have data all 7 days of the week. But often, like with the NYSE, there are days of the week when no activity occurs; this leads to even greater variability in month length. If yours is a business that’s only open 5 days a week, then at most you can have 23 business days of activity. And you can have as few as 18 business days in a month – which occurs when a 2-day Thanksgiving holiday falls in a month where there are 5 full weekends.

Even the same calendar month can vary substantially – here we see March 2011 had 23 business days, while 2 years later March had only 20 business days as the markets were closed on March 29 for Good Friday.

**7-8 When we index the daily data, and spread it across the possible 23 business days of the month, shorter months align poorly.**

The problem with the varying month lengths becomes apparent when we chart the data. For example, the 20th business day in 2011 was fairly quiet, at 20% below average, while in 2013 the 20th business day was the last day of the month, and sales were 25% above average.

**7-9 To fix this issue, we want to manipulate the data so that the “missing” days are all dropped mid-month.**

To address this issue, we want to manipulate the data so that the “missing” days are dropped from mid-month, or put differently, that the “extra” days for the longer months are to be found in the middle of the month. Why the middle? Because generally this is where the least amount of pattern change is found across a given month. Usually the most distinctive highs & lows (especially highs) are found at the beginning & end of each calendar month. We want to capture those distinctive patterns. Also, we want to be sure that we are entirely consistent with how we go about assigning the “missing” days – that they always occur at the same time of the month.

**7-10 To begin the process of manipulating the data, we bring in the Daily data, and normalize it.**

We’re going to fairly quickly walk through the process of developing the Day-of-Month Factors. It starts with normalizing the data, which requires a bit of work to set up properly. So to begin, we bring in the daily data, in a tab we’ll call “Data”. Actually, all I needed to do here was copy in the data from the “Normalizing Data Template” file. Let’s walk through the columns here.

Column A shows the calendar date. Note the starting row here of Row 3665. It’s that high because this file actually begins back in 1991; I’ve just hidden those 1st 10 years.

Column B shows the day of the week.

Column C, which had the date of the Monday that started the week is hidden as we won’t need to use that.

Columns D&E identify the calendar month and the day of the month.

Column F has the original sales data.

Column G is blank, while columns H & I bring in the Equated Day Factors and the Holiday Factors.

Column J calculates a Net Daily Factor as the product of the EDF & Holiday Factors.

Column K has the Normalized Data, taking the original data in Column F and dividing it by the Net Daily Factor in Column J.

As we walk through this process, we’ll follow what is happening to January 30, 2001, as an example. Here we can see its original total of just under 1.15 billion shares (Cell F3694) becomes a normalized total sales of just over 1.136 billion (Cell K3694).

**7-11 A table is created that sorts the data by day of the month, and calendar month.**

We next turn to a tab called “Table23”, where we’ll perform some calculations that will enable the data to be sorted in a format much easier to work with. It’s called “Table23” as it is designed for organizations where the maximum number of business days in a month is 23. This chapter will focus on “Table23”, but the file also has a “Table 27” & “Table31” so companies with 6-day & 7-day workweeks can pick up data from those tabs instead.

In the Table we see here, the normalized data, expressed in billions of shares, is brought into a table where there are 31 rows for the days of the month, and there are columns for every month for which we have data. For display purposes, July 2001 thru June 2016 are hidden.

**7-12 The manipulation process begins by identifying whether or not a given day has had activity.**

We start the manipulation process by identifying whether or not a given day has activity. January 30 had activity, and gets a “1”. The total each month, in Row 83, tells us how many days of activity there were that month.

**7-13 We then calculate the cumulative days of activity.**

Then we calculate how many days of activity have so far occurred in the month, being sure not to count the days with no activity. January 30th was the 20th day of January 2001 to have activity.

**7-14 We then bring in the actual normalized sales volumes for each day of activity.**

Now we bring in the normalized sales volumes, but sort them according to their day of activity. We earlier saw that January 30th was the 20th day of activity. This table picks up the 1.136 billion in sales that occurred on Jan 30, and aligns it with the 20th day of business activity.

You can see this table is shorter than the others, because now we are only dealing with business days, and the NYSE has only a maximum of 23 business days in the month. Note that of the 12 months on display, only August 2016 had a full 23 days of activity.

**7-15 A “Data Manipulation Table” is constructed that lays out how we want to arrange the data, based on how many days of activity occur in the month.**

Now we determine how we want to manipulate the data. A “Data Manipulation Table” is constructed that lays out how we want to have the data arranged, based on how many days of activity occur in the given month. We want the table to place the “missing” days mid-month. That’s what we have done here, being careful to always have every month end at the 23rd day, and to have the missing days right in the middle of the month. That’s why the cells are blank, for the “missing” days of the month.

For months with 21 business days, like the January 2001 month we’ve been tracking, we will want to pull the data sort from Column GR. And we would want the 20th business day of that month to be shown as occurring on the 22nd day of the month.

**7-16 A VLOOKUP formula picks up the day of activity from the “Data Manipulation Table”.**

We next use VLOOKUP formulas to identify the day of activity number that we wish to assign to each day. Our Jan 30 example was the 20th business day of the month, and was assigned the number “22” in the Data Manipulation Table. Accordingly, we find it here in the 22nd row. And we see there are no numbers for the two “missing” days that have been placed midmonth.

**7-17 An array formula then brings in the normalized sales amounts for each day of activity.**

Array formulas are then used to bring in the normalized sales amounts. Our Jan 30th example has the 1.136 billion shares showing up on the 22nd, the second to last day of the month. We also use this occasion to calculate the total sales for each month, and the average daily sales. Those average daily sales will be used for indexing. We can see that Jan 30’s sales of 1.136 billion is a bit less than the 1.279 billion average daily sales in January 2001 (Cell B236).

With this table, we are now done with manipulating the data so that the “missing” days of each month are placed mid-month. We can now go on to calculate the DOM factors.

**7-18 Rather than calculate one set of factors to apply to all months, we will calculate DOM factors for each of the 12 months of the year.**

Ordinarily, one might think that it would be sufficient to estimate just one set of Monthly DOM factors, representing an average behavior for all months. But there’s a big problem with that approach – it can be very tricky to determine how to transition from one calendar month to the next. If August, for example, is quieter than July, how do we transition from July 31 to August 1? We will consider this issue much more closely when we are done with estimating the DOMs. For now, let’s just take it as a point of faith that it is more appropriate to estimate DOMs for each of the 12 months of the year.

**7-19 Calculation of the DOM factors has four parts, as laid out below.**

The figure here shows the entire calculation of the DOM factors for the month of January. Let’s walk thru each of the 4 parts of this calculation.

**7-20 The data for all January’s is brought in, as an index.**

We begin with the data, as one almost always does. Section A has brought in the data from the “Table” tab, calculated as an index. Thus, we can see here the index of 0.888 for the second to last business day of January 2001 (Cell C32). This figure was obtained by dividing the sales for January 30, 2001 (1.136 billion) by the average sales for January 2001 (1.279 billion). Note that there are no “0’s” for the missing days. Instead, I went through and manually deleted these formulas so only the days with data will show. Why? Because when I calculate the averages and standard deviations for each day across the years, I do not want to count those years where there is no data.

I have also summed the indices each year, and checked it against the total days counted when the data was being manipulated; that’s what the bottom row, Row 39, is about - it’s a check to ensure our index calculations total appropriately.

**7-21 The simple average, standard deviation, and an upper and lower bound are then calculated for each day of the month.**

As ever, we want to determine what data to accept, and what not. As we have in past, we calculate a simple average and standard deviation. These are calculated for each of the 23 business days in the month. Of course, January only has a maximum of 21 business days (due to the New Year’s Day & ML King holidays). So there are no calculations made for the 11th & 12th of the month.

As before, we work with a maximum number of permitted standard deviations of 1.5, (shown in highlighted Cell V9). That is applied to the average and the standard deviation to arrive at the accepted Lower & Upper bounds for the data. Thus, for the 22nd, the average is 1.03 (Cell U32) and the standard deviation is 0.12 (Cell V32); one and a half standard deviations would be 0.18. When 0.18 is subtracted from the 1.03 average, we have a lower bound of 0.85 (Cell W32), and when it is added to the average we get an upper bound of 1.21 (Cell X32). Our 0.888 value for January 2001 (Cell C32) is within that range, and therefore is “accepted”.

**7-22 An “Accepted Values” table determines whether or not each day’s index comes within the accepted range.**

Armed with the Lower & Upper bounds for each day, it is a simple calculation to determine whether or not each day’s value is accepted. In Section B, formulas apply a “1” to all days falling within the accepted range. As we just saw, the 22nd business day of January 2001 was accepted.

If you look carefully at the table, you’ll see that there is not one single year in which all business days fall within the accepted range. Even 2005 just misses as the 10th fell out of range. Because every year fails to always meet the criteria, when it comes to calculating a weighted average, we will pick up any and all years where the data was accepted.

**7-23 Weighted averages are calculated for each day, and a final set of DOM factors determined.**

Once we’ve identified the days whose values are accepted, a simple weighted average calculation gives us our DOM factors.

Note that the weighted averages in Column Y are a preliminary calculation. We want the factors to exactly average 1.00.as there are 21 days of activity, we want the weighted average to total exactly 21. This is accomplished here by multiplying each day’s preliminary average by the ratio of the total days (21, per Cell Z9) divided by the preliminary average total (20.73, per Cell Y35).

**7-24 Except for a strong final day, most of the January DOM factors hover around 1.0.**

Charts greatly assist in visualizing and understanding what the results look like. The weighted average line here is actually a simple average of all the points that fall between the Lower & Upper Bounds. There is a bit of noise in these results. There may be cause for the 2nd day of the month to be down a bit, coming right after the start of a month. But I can think of little cause for days 5, 9, and 20 to be lower than their adjacent days. Days 9 & 10 are notably lower than all others. Meanwhile, the several days preceding month end show a bit of an uptick in activity. And of course the final day sees a surge. These should do for now.

Before leaving this, note the seasonal factors displayed above the chart. These factors are imported from the “TrendModel” file we developed in the last chapter. The 1.04 factor becomes of interest when we look next at February.

**7-25 February’s DOM pattern shows a distinct decline across much of the month.**

The approach taken for calculating January’s DOM factors was repeated for February, and all the other months. There’s quite a distinct pattern here in February – a long decline across much of the month, with some pickup the last few days before a usual strong result the final day.

Notice that February’s seasonal factor is around 1.00, about 4 basis points below January. What’s interesting about this is what February’s pattern suggests about the transition from January to February. Again, this will become clearer when we put all 12 months side-by-side, (which we shall soon do). What this declining pattern suggests is that February begins as strong as January ends, perhaps even a bit stronger, and then it proceeds to decline across the month. So what? Well, think about how February’s factor is lower than January. How does this drop play out? If every month had the exact same DOM pattern, it would suggest that there is a sudden drop or lift at the start of each month. But that doesn’t make sense. You’d expect to see some kind of smoother pattern to take place as one month transitions to the next. February’s pattern fits with a transition where the decline is occurring across the month, not at the month’s start. And as we’ll next, March continues this pattern of decline.

**7-26 March’s DOM shows a continuation of decline across the month.**

March shows a continuing general decline in the factors across the calendar month. Indeed, even the last day’s factor is lower than usual, the first time we’ve seen the final day come in below 1.1. This is also the first time we’ve had a full set of data across all 23 business days. We have fewer years of data to work with for the factors for the middle days: the 11th & 12th business days of the month. But their results are certainly consistent with the factors preceding and following. The 13th is low; clearly we are never going to be able to escape some noise from one day to the next.

Note that again March’s seasonal factor is lower than February, and again by about 3 or 4 basis points. The February & March DOM patterns suggest there is a steady decline from January thru March, with no sudden drop or jump as one month flows into the next.

**7-27 Unique perhaps to the NYSE sales data, the data surrounding the 3rd Friday is extremely erratic when it’s “holiday factor” is removed.**

Before leaving March, I should point out something that is not initially apparent, and that is quite unique to this NYSE sales data. You may recall there is a big surge in sales on the 3rd Friday of each of the quarter-ending months (i.e., March, June, September, & December). No such sales increase is evident here because we adjusted for the 3rd Friday when we normalized the daily data. However, while this daily data adjusts for the 3rd Friday factors, the monthly seasonal factors we developed in Chapters 4 & 5 did not. Isn’t that inconsistent, to adjust in one instance but not the other?

No, because these factors change so much over time, almost doubling between 2007 & 2012. All the other “normal” holiday factors are pretty consistent across the years, indeed across the decades. Not so the 3rd Friday which is why we did not adjust for them when developing the monthly seasonal factors, and when we trended the monthly data over the years. But as you can see from the chart on this page, the daily data responds badly when it is not adjusted for the 3rd Friday factors.

This is a very unusual situation, and hopefully unlike anything you will encounter with your own data…

**7-28 April sees some decline the first half of the month, and a distinct jump midmonth. But it also has an apparent outlier on Day 11.**

We now turn to the April DOMs. I’ve shown the entire tab here because I want to explain the treatment for the apparent outlier value that occurs on the highlighted 11th day of the month. It doesn’t make sense that the value would not only be 15% higher than the 10th day and prior, but that it is 5-10 basis points more than the days following it. It’s probably hard to see, but the culprit here is the factor from 2013. Falling in the middle of the month as it does, there’s only 3 years that have values for the 11th, and the value for 2013 is 1.448; it is literally off the chart. However, it still gets picked up as it is not outside the Upper Bound. There are two ways we can get rid of this outlier.

**7-29 One way to get rid of the Day 11 outlier is to lower the number of allowed standard deviations.**

The 1st way we can get rid of this outlier is to drop the maximum allowed standard deviations – by narrowing the accepted range we can ensure the 2013 value is removed from the weighted average calculation. That is done here by lowering it to 1.15; yes, it’s a bit of a funny number, but it was the highest I could go while still succeeding in lowering the Day 11 value. Of course, when we do this, the upper & lower bounds are tightened, and the resulting pattern across the entire month is modified. But if you go back and see the prior pattern, you can see the change is still fairly modest. And we’ve succeeded in getting a “better”, more consistent value for Day 13.

**7-30 The other way to get rid of the Day 11 outlier is to manually remove the 2013 value, accomplished by inserting a “0” in the Accepted Values table.**

The other way we can get rid of the Day 11 outlier is to manually insert a “0” in the Accepted Values table. By assigning a “0” to 2013’s Day 11 value, the formulas do not pick up the 2013 value. So again the Day 11 value is much reduced, as we wanted. By using this approach, we maintain the 1.5 maximum allowed standard deviations. And by doing that the pattern changes little from the one we started with; all we’ve done is to drop the Day 11 value, which in turn lifts all other days proportionately. You can see this if you go back and look at the original results.

**7-31 Generally, the preferred approach for dealing with outliers is to modify the allowed range as needed.**

So which approach shall we go with? My preference is for the tightening the allowed range by lowering the maximum allowed standard deviations. This is a “cleaner”, and seemingly less “manipulative” way of adjusting the calculations so that they remain statistically consistent, albeit with the lowered maximum. But I wanted to show the other technique, of manually inserting a “0” in the Accepted Values table, because there may be occasions when this has a better result. Better not because the pattern is closer to the pattern we may “want” to see, but better because the pattern makes more sense.

We now see April has a pattern of slowly dropping during the 1st half of the month, then jumping about 5 basis points right after midmonth. April’s seasonal factor is around 1, or about 3 or 4 basis points above March’s. So here’s an instance where the change in the monthly seasonal factor is apparently due to a sudden shift in activity, rather than a trend that slowly extends from one month into a next. And what might explain this sudden shift? Well to begin with, of course, we’re looking at the days immediately following April 15, infamous in the U.S. as the day taxes are due. So perhaps there is some degree of freeing of investor inhibition to be active in the market. Also, midmonth of the 1st month in a new calendar quarter is when companies often start releasing their quarterly results, which typically see an uptick in investor activity in response to how the company has performed in the prior quarter, whether for better or worse. January also showed a bit of a jump going from the 1st half of the month to the latter half. We’ll have to wait and see how July & October behave to see how consistently quarterly announcements lead to a general increase in sales.

**7-32 May again sees a general pattern of decline across the month, though not as marked as we saw in March.**

We’ll quickly go through the rest of the months, remarking on matters or adjustments of note. May is generally unremarkable, showing some decline across the month, and a sharper than usual increase at month-end. Note that the maximum allowed standard deviations was again modified. Here I actually increased it slightly, in order to have the highlighted lower value for Day 11 be counted; it fell just outside the accepted range when the maximum was set at 1.5.

**7-33 June’s pattern generally follows a classic slight decline across the month, with a sharp increase at month-end.**

June’s pattern looks fairly typical – a slight decline across the start of the month, and a slight incline as month end approaches.

**7-34 July’s pattern is remarkably flat across the month. The flat start may be due in part to the July 4th holiday; there’s a slight bump going into the 2nd half.**

July’s pattern has a couple of observations worth noting. 1st, the start of the month is unusually slow. We’ve typically seen the 1st few days above 1, and sometimes approaching 1.1, but here its right around 1.0. The bump for the 4th business day in July should probably be ignored and is presumably due to the vagaries of the July 4 holiday and its holiday factors. We know there can be pretty low values the day preceding and/or following the 4th. So that the numbers here are slightly up may simply be a function of the holiday factors. As it happens, were you to eliminate the two highlighted values, the average would drop to just below 1, and end up being completely consistent with the other days before & after.

The other point to mention here is that the 2nd half of the month does appear to be distinctly higher than the 1st half. And this would be consistent with the observation we made of April – that the 2nd half of the 1st month of each quarter will be slightly higher due to quarterly report releases.

**7-35 August is the quietist month of the year, with sales declining strongly through the entire month.**

You might recall that August had the lowest seasonal factor of the year, down significantly from July, and September. A look at the month’s pattern reveals why it can be so helpful to estimate the DOMs on a calendar month basis, for one would not have likely thought the month would see such a distinctive downward slope across the entire month, as we see here. The DOM factors are well above 1.1 the first few days of the month, well above the levels we’ve seen other months. But this does not reflect that the early part of the month is busy, so much as the fact that it is the latter half of the month that is especially slow. Even the last day of the month sees activity well below that of the entire 1st week.

You may also notice there is considerably more volatility with the data, that the band of acceptable values is very wide, so wide that volumes are often off the chart. Of course, I could have set the Y-axis wider to accommodate, but then you would have a false sense of how this month compares with all the others. (Of course, I could have broadened all the other months but then the charts would be more difficult to read; there’s always compromises). Note too that the maximum standard deviations was bumped up to 1.7, not because there was an odd outlier, but to have the weighted averages come out more smoothly across the month.

**7-36 In contrast to August, if anything September shows slight growth across the month.**

We saw how August declined quite steeply across the entire month. In stark contrast, September shows a slight incline across the month. As ever, the pattern is not totally smooth, but it clearly has an upward slope to it. Imagine how inaccurate the DOM factor for this month, and August, would be if we relied on just one average DOM pattern representing all 12 months.

**7-37 October has its own unique pattern – almost completely flat all month. And October is the busiest month of the year.**

October has its own unique pattern; it is the flattest month of the year. What’s striking about October is that it’s also the busiest month of the year, with a seasonal factor almost 10 basis points above September’s. Evidently, to get to that higher level, there will be a jump right at the start of the month.

**7-38 November’s pattern is much more similar to those from earlier months in the year.**

With November we see a pattern much more similar to that we saw with most months over the 1st half of the year. There’s a relatively strong start, slight decline across the month, and a strong last day. Given November’s seasonal factor is only around 1.0, it looks like there will be a bit of a drop going from October to November, as well as a decline across the month that gets its seasonal factor about 7 basis points below that of October.

**7-39 December is broadly flat, with a pattern in the second half of the month that is likely somewhat influenced by the 3rd Friday & Christmas holidays.**

December is a bit of a mess. It’s pretty flat the first half of the month, but then sees sharp lift for a few days, and a broad decline through month-end. The last day of the month is by far the quietist last day. What’s difficult to sort out here is how much this latter half is influenced by the holiday factors. The slow final day of the year is picked up in part by the holiday factor. The holiday factors average around 0.71 for the last calendar day, while the second to last day averages 0.64, and the 3 days preceding are lower still. So there is a 7-10 plus basis point bump at month end that is missed here because the data here has been adjusted for the holidays. Nonetheless, it clearly is pretty quiet as the year draws to a close. As for the spike between the 13th & 15th business days, this may be partly due to this falling around the time of the 3rd Friday, where activity is much increased. “Holiday” factors for the 3rd Friday adjust for this, but perhaps those factors slightly understate the degree to which December’s 3rd Friday is busy. The bump here isn’t much, but it’s good to have an understanding of what may be causing it.

**7-40 When the DOM Factors for all 12 months are plotted together, we can more clearly see a distinct pattern across the month.**

It’s interesting to see what all 12 months of DOM Factors look like when brought together. That’s what we see here. And again we see the distinct pattern we’ve frequently been describing across the month: a somewhat strong start, a decline across the 1st half of the month, a rather flat 2nd half, with a strong close on the final day. It’s a bit surprising to see quite how flat it is the 2nd half of the month, particularly those last few days immediately preceding the final day. And it’s quite noticeable that there is something around a 3-4% bump that occurs when transitioning from the 1st half of a given calendar month to the 2nd half.

While we may have a distinct average pattern across the month, nonetheless, we can also see that some calendar months behave quite differently from this average. August is obviously quite unique. April is clearly much quieter during its 2nd week, and we earlier commented on how September & October differed. So we would want to be careful about just applying one overall average to every month. It can be quite informative to see the result when we take a look at all these months side-by-side.

**7-41 The DOM Factors are set up side by side across the year.**

To see what all these DOM factors look like, a chart was put together that confines each month length to the maximum 23 days. As we previously saw, most months have a day or two missing in the middle, and virtually all months have a very strong last day. The day to day volatility makes this chart a bit difficult to read.

**7-42 The daily factors are smoothed by applying a 5-day moving average, and dropping the last day spike.**

To make the chart an easier read, two changes were made. First, a 5-day moving average was applied, to help smooth out much of the day-to-day noise. 2nd, the last day of the month was dropped, as that day behaves so differently from all the others.

**7-43 The problem with even the smoothed line is clarity on how one month transitions into the next.**

There is a central problem that remains with this chart: it is the behavior as one calendar month transitions into the next. I’ve highlighted the transition into August and October here because these are months that saw the biggest shift in their seasonal factors (from their prior month). Notice how strong August starts. Is that really the case, that daily activity suddenly soars as the month begins? On the other hand, notice how October appears to pick right up from the end of September; is this right?

**7-44 When we apply each month’s seasonal factors to their DOM Factors, we arrive at a much clearer & “truer” picture of daily activity across the year.**

We address this transition by bringing in the final piece – each month’s seasonal factor. Here, the DOM Factors were multiplied by the monthly seasonal factor. August was the quietist month, so all the DOM factors drop sharply (down about 7%), while October’s factors bump up (by about +7%). With this revised look, we get a truer sense of the day-to-day behavior across the year. We can see that August now basically picks up from where July ended. At the end of August, however, there is a big jump at the start of September. Meanwhile, October experiences a sharp jump at the start of the month, while the drop off into November appears to be more gradual; i.e. the start of November appears to pick up from the end of October, and to then drop quite sharply.

**7-45 While there is always a certain degree of volatility, even after smoothing the data, there are only a few occasions when a distinct shift occurs.**

Data is messy. Daily data is very messy. Which is why a fair amount of work needs to be done to data to help clarify the underlying picture. When we look at our smoothed daily activity, after applying the monthly seasonal factors, we can see that for the most part there is really not much shifting that occurs in the data. It bumps around quite a bit across the months, and year, to be sure. But there are really only 3 distinct occasions that stand out to this viewer. Mid-April sees a definite bump up, perhaps due to tax day and quarter end, as earlier discussed. We can see there is some increase occurring in the middle of January and July as well, though not as much, while October sees little mid-month increase at all.

There is a huge jump that occurs at the start of September. Activity plunged throughout August, and the start of September sees it come back to a level close to where it was in the middle of August. September then continues to climb across the month, ending the month near where activity was in July.

And then October sees a distinct jump at the start of the month, and stays high the entire month, only to see activity fall off quite rapidly the 1st couple of weeks in November.

**7-46 Using the same set of DOM Factors every month fails to capture the distinct patterns some months have, and month-to-month transitions are unclear.**

One final point about the DOM Factors – was it worthwhile to develop 12 different sets, one for each month? Did it really make that much of a difference? In the instance of the NYSE sales, it certainly appears to be the case that different patterns each month appear to do a better job of capturing daily behavior. For one, it better reflects how some months can behave quite differently from the others – certainly that was the case for August (as well as for March & April, & July & September). And for another, there is the huge challenge of capturing how daily activity transitions from one month into the next: should there be a smooth transition or a jump, when and how do you apply the smoothing or jumping? These are very difficult to determine when you rely on just one set of DOM factors to reflect behavior across an entire year. Which is why you will almost always want to try to develop DOM factors for each of the 12 months. In effect, of course, you’re developing factors for each day of the year; you’re just breaking them up between the 12 months, and separately capturing seasonality by month, for generally you will always want to be tracking and analyzing monthly data.

Of course, if you have little historical data to draw upon you may be forced to rely on just one set of DOM Factors. If that is the case, do not be surprised to find the adjusted data will prove more difficult to work with.

**7-47 To conclude, we review a few key rules of thumb about developing DOM Factors.**

To conclude, let’s review a few key rules of thumb that you will want to follow in developing DOM Factors.

First, if at all possible, you will want to develop a different set of DOM Factors for each calendar month. If you find that every month has virtually the same pattern, then you’ll probably be fine with just one set. But that is very unlikely. Developing a different set for each month will help you get closer to the day-to-day pattern that occurs across the year, and by so doing, you will be able to do a better job of seasonally-adjusting your daily data so that you can more clearly determine how your data is behaving over time. This will be especially important when you try to measure & analyze events.

Second, be sure to manipulate your data so that “missing” days of short months are always found mid-month, (or at some other point during the month when activity is relatively flat). This manipulation will make it much easier to ferret out the patterns across the month.

Third, as with all such analysis, you’ll want to remove outliers so a clearer picture is revealed of more typical data behavior.

Finally, be patient. Daily data is inherently messy. Your results will likely prove difficult to read, even after removing outliers, and looking at each month separately, and smoothing the data, and so on. This is the nature of the beast. Nevertheless, please recognize that although your data will not be very accommodating, it will likely still be well worthwhile to go through all this effort as you will be able to better understand how your data is behaving – on a daily basis. And that’s no mean feat.

**Chapter 8: Estimating Trend Daily**

Now that we have the DOM factors, we’re ready to move on to estimating trend on daily data. A warning up front. This is a challenging process. Daily data will inevitably display much volatility, even after adjusting the data for numerous predictable day-to-day variations. Inserting moving averages is going to help with this process, as will being able to observe month-to-month differences between actuals and your estimate of trend.

Throughout the process of estimating trend daily, keep in mind its purpose: to inform & explain data behavior, to provide timely awareness & accurate measures of sudden data shifts, and to learn from past experience so that future performance may be improved.

**8-2 This chapter will examine how you go about estimating trend on a daily basis.**

This chapter is roughly divided into five parts. We start with an introductory overview, then describe in some detail, how the model is set up. We then walk through the process of estimating trend; follow that with a discussion of “Low Factor Adjustments” that are a part of the model to help avoid the undue and disproportionate impact of seasonally-adjusting data when certain days find themselves with unusually low factors, such as may occur immediately following a major holiday. And we conclude with a summary review of key points from this chapter.

**8-3 Perhaps the greatest challenge with estimating daily trend is the substantial volatility in the data, even after normalizing.**

One of the greatest challenges of dealing with daily data is its inescapable volatility. There’s always going to be substantial fluctuations due to all manner of causes.

**8-4 Swings in the data occur both day-to-day, and week-to-week.**

While normalizing the data – for the day of the week, holidays, day of the month, etc. – can help reduce it, there will still remain a substantial amount of volatility. Daily fluctuations occur for a vast amount of reasons, some of which may be explicable, and some that are just simply the inevitable “noise” of any given data set. Even the explicable reasons for daily fluctuations may not merit trying to track and measure, simply because they may not provide much helpful insight and the effort just isn’t worth the payback. But if you’re in a highly competitive industry, and you’re accustomed to tracking & reporting data daily, then you still need to grapple with this volatility. Hopefully the levers and elements provided by this model, and described in this chapter, will help you better cope with this frustrating aspect of the data.

**8-5 Many variable characteristics of the calendar need to be considered in order to correctly normalize the daily data.**

It’s easy to think of daily data as simply a collection of 365 daily data points over the course of the year. But to try to work with so many individual data points, in just one year alone, would be overwhelming. Inevitably, we will want to break the data up into pieces to some extent. In particular, we will want to think & work in terms of the month of the year: the months are generally the most important time period that is reported on and acted upon. Daily data is noisy, and rounding it into its monthly components helps summarize the data.

But the calendar is fickle, and it’s helpful to be reminded of what makes it so difficult to work with. The calendar months are of different lengths, varying from 28 to 31 days. If a business is closed on weekends, month lengths vary even more, from between 18 to 23 business days. Holidays interrupt the flow of data, causing various temporary slowdowns (or upticks) across the year. Easter can be rather nettlesome because its timing changes each year, and crosses two calendar months. Unplanned holidays can occur, where business suddenly and dramatically changes for a few days due to events that are outside our ability to control – natural or manmade disasters, for example. Sometimes month lengths can be especially short, such as, for example, when businesses like the NYSE are closed for Good Friday, and in a month when there are already 5 full weekends accounted for. Leap years make for a longer year obviously with the extra day added, a day which makes February a little less reliable a unit of time to examine. And year lengths vary not just because of the leap year, but because having a year of 365 (or 366) days length means that one (or two) days of the week occur 53 times, not just 52.

In short, the design of our calendar is part and parcel of why trend series data can be difficult to analyze – there are just so many aspects of the calendar itself that automatically builds in a certain level of irregularity to the data. So be it; we’ll do our best regardless.

**8-6 The “Daily Trend Model” has three sections:**

We start with the development of the Daily Trend Model. This model can be found on the website; it is labeled “DailyTrendModel”. The model here tracks the NYSE. You can use this model for your own data, by simply overlaying your own data onto it: replace the daily data with your own, and replace the various daily factors with your own. If you have fewer years of data – this model goes back to 2011 – delete or erase the extra unneeded years. If you have more years, add those on, using the existing model here as guidance.

The DailyTrend Model has three tabs, which we will walk through individually. The “Inputs” tab brings in all the key factors we’ve developed – EDFs, DOMs, etc. The “Calc” tab handles virtually all the calculations. The “Trend” tabs are where the trend is estimated. But because we are dealing with so much data, the tabs are divided by year, with one tab for each year. So this model has 10 different “Trend” tabs, one for each year between 2011 and 2020.

**8-7 The “Inputs” tab brings in all the factors that influence the relative level of activity across the days of the week, month, & year.**

As mentioned, the “Inputs” tab brings in all the various factors we’ve developed for describing the data. Let’s walk through it.

**8-8 The top of the page imports all the previously developed daily and holiday and monthly seasonal factors.**

The top of the “Inputs” page has all our previously developed factors related to the daily data. These factors were previously developed to determine more precisely the “true” length of each month. Now they will be used to adjust each day’s volumes to account for the day of the week, the influence of holidays, as well as the relative busyness of their calendar month. The Daily Trend Model we’re using here will only be for the period 2011-forward, so arguably the factors shown here for earlier periods will of course not be used. But they’re left here, partly as it is easier to just bring in the whole set as is, but also because when it comes time for you to insert your own data and results, you may well want to have the additional placeholders to work with.

**8-9 We will want the Day of Month (DOM) Factors to be available in a format that meets two criteria.**

The DOM Factors, that we developed in the last chapter, measure the activity on a given day of the month relative to a daily average across the month. They tell us how much more busy or quiet each day is simply because of where it is placed during the month; and we noted how there can often be an uptick in activity at month’s end, especially when dealing with sales. We are now going to want to *apply* those factors, and in doing so, we want to do it in a way that’s easy to use. In other words, we’re going to want to be able to apply them all by formula alone, without having to insert or move factors around manually. (Not only would manual shifts be cumbersome and time-consuming, they can easily be done incorrectly.) We also want to make sure that the factors are applied appropriately, given the changing month lengths. For example, while March can be as long as 23 days (assuming business is closed at the weekends), it can have only 22 days or 21 days if there are 5 Saturdays or 5 Sundays or both, in the month. And then, because we’re using the NYSE sales volumes example, we find that occasionally there are only 20 business days in the month – this occurs when there are 5 full weekends, and when the Good Friday holiday falls during the month of March. We want to set up the DOM Factors in the “Inputs” tab in such a way that these two criteria are met – that they are easy to use, and that they appropriately handle the changing month lengths.

**8-10 The monthly DOM Factors developed in the last chapter are brought in, sorted *without* the missing days in the middle.**

We start out with the monthly DOM factors we developed in the last chapter. Notice that the “missing” days in the shorter months, like January, are now placed at the bottom. We had previously placed the missing days in the middle to ensure every month appropriately ended strong on the final day. Now we have the missing days back at the bottom because it will be much easier to pick them up when we come to apply them in the Calculations tab. Note that the 11th day in January is shaded pink. This is to draw attention to the fact that the formula skips here. The 10th picks up the value from the 10th day in the DOM Factor file, but the 11th picks up the formula from the very next day of activity, which for January happens to be from the 13th row in the summary DOM Factors table.

**8-11 Additional source tables need to be developed for when months have a shorter length.**

Next we begin to address the varying lengths of the months. Here we have added in a table for what are called the “Medium Length Months”; these are the months where there is one extra Saturday or one extra Sunday. Note that we’ve hidden the months of April thru October in the tables here in order to be better able to see what’s going on.

**8-12 The different month lengths are identified in the “Inputs” tab.**

The tables are labeled to identify the length of the month. The 1st table has the “long” months; the 2nd table has the “Medium Length Months”. They are so labeled at the top. And, in Row 127, they are given the number “1” for “Long” months and the number “2” for “Medium” months. These values will later be used in array formulas to enable picking up the appropriate set of factors.

**8-13 The shorter months pick up the values from the original “Long Month” table, dropping the day(s) in the middle.**

Next we shorten the month length by picking up the values from the original “Long” month table. Note however that the 2nd table drops the day in the middle. Thus, for March, it picks up the 1st 11 days of the month, and the last 11 days of the month; the value for the 12th day is left untouched.

**8-14 Because we are dealing with indices, we want to make sure that the total for the factors always exactly equals the number of days.**

The DOM Factors are factors, indices that measure the relative length of each day. Because these are indices, we want to make sure they always add up to exactly the number of days that we are dealing with. Thus, March’s factors total exactly 23.00, as highlighted here in Cell E125. We want to make sure that the shorter “medium” length month totals exactly 22.00, as you can see is the case in highlighted Cell T125.

But how do we get that total, given the dropped day was 1.02 days in length (Cell E112)? Dropping that day’s 1.02 value from the month’s total of 23.00 should give us a total of 21.98. We need to make an adjustment to ensure the month total reaches the desired precise 22.00 days.

**8-15 In order to ensure each month’s length exactly matches its number of days, we pick up the dropped value, and calculate the required adjustment.**

Continuing with our March example, given that we are dropping 1.02 days from the month of March, we’re going to need to slightly increase every other day’s values in order to have the month total the desired exact 22.00 days. To do this, in the table at the bottom we pick up the length of the middle days that are being dropped. Thus, Cell E130 picks up the 1.02 value from the 12th, the day that’s being dropped. Then, in Cell E134, a formula calculates the ratio between the desired exact month length, and the month length without the dropped day. In this example, that equates to 22.00 days divided by 21.98 days, which works out to be the 1.001 found in Cell E134.

**8-16 Applying the calculated “Factor Change” to all the factors ensures that the month factors total will exactly match the number of days.**

Now that we’ve calculated the Factor Change we need to apply, we do so by taking each day’s original value and multiplying it by the “Factor Change”. It’s hard to tell, due to rounding, but every day’s value is slightly higher. However, if you look closely, you’ll find the rounded value for Day 19 was “0.95” in the “Long Month” table (Cell E119), while it bumps up to “0.96” on Day 18 in the “Medium Length Month” table (Cell T118).

**8-17 The same procedure for adjusting month lengths is applied to the “Short” months and the “Extra Short” months.**

We just laid out the procedure for adjusting each day’s value so that the factors for the shorter months will always exactly total the number of days. Thus, we can see that the factors for March exactly total 22 days when there is one day dropped, 21 when two days are dropped, and 20 on those rare occasions where 3 days are dropped. These rare occasions only apply to the months of March & April, due to the roving Good Friday holiday that can sometimes land in March and sometimes in April.

It’s a bit involved, but we now have all the DOM factors laid out so they can be easily picked up when it comes time to adjust the actual daily data for the day of the month.

**8-18 The “Inputs” tab is also used to pick up the Growth Rates & Event estimates that were developed using the *Monthly* trend model.**

Before leaving the “Inputs” tab, I want to point out one other set of data that has been brought in. This is the estimates that were made of Growth Rates & Events using the MONTHLY trend model. We’re going to find a lot more volatility when we work with the Daily data. We will find that it is helpful to have our previously developed monthly estimates as a guide, to help us make out some of the forest for the trees. As we’ll see, these estimates will be picked up as reference in the “Trend” tab.

But first, let’s turn to the “Calc” tab.

**8-19 The “Calc” tab performs all the adjustments to the daily data to arrive at seasonally-adjusted data, which is then compared with trend estimates.**

The “Calc” tab is where we will perform all the necessary calculations to seasonally-adjust the daily data, adjusting it for day of week, holidays, day of month, monthly seasonality, and as we shall later see, an additional adjustment for days with very low net factors. The seasonally-adjusted data will be smoothed by calculating a moving 1-week and 2-week average. Estimated trends will be calculated here, and then compared against original actuals. Let’s walk through these different parts of the calculations.

**8-20 The “Calc” tab starts off by identifying the date – the date itself, the month, the day of week, and the week of date.**

We start by showing the dates in various forms. The individual dates are in Column A. The month is shown separately in Column B in order to later expedite totaling values by month. The day of the week is displayed for reference. The “Week of” date identifies the Monday of the week the date falls in; it is calculated separately so charts can easily show the individual weeks.

**8-21 The daily volumes are brought in, as well as the factors used to normalize the data for day of week & holidays.**

We next bring in the daily data, expressed here in billions – using the “Units” identified at the top of the page in Cell H1. Then the Equated Day Factors & Holiday Factors are brought in, and the Normalization Factor calculated as the product of the two.

**8-22 Next the Day of the Month Factors are brought in, with appropriate calculations made to ensure the correct set of factors is correctly applied.**

Next we bring in the Day of the Month (DOM) Factors, but to do so correctly requires some additional calculations first. We begin by identifying the month in numerical form – so January is Month 1, February is 2, and so on. Then we determine whether or not the day is to be counted, done by simply applying a “0” if there is no volume that day, otherwise it is considered an active day and gets a “1” in Column J. We next calculate a cumulative count of the days of activity in the month. Again, if there is no activity on the given day, it gets a zero, otherwise, there is a SUM formula that adds up the month-to-date counts from Column J. The formula starts again with each calendar month, so February 1 is back to a “1”.

The Month Length column identifies whether the given month is long – in which case it gets a “1”, medium length gets a “2”, short gets a “3”, and extra short a “4”. How is this calculated?

**8-23 The Month Length is calculated at the bottom of the worksheet, where each month’s day count is compared with the longest length the month can be.**

The Month Length is determined at the foot of the worksheet. In Column I, the longest length each month can be is inserted manually. January can have at most a total of 21 days, thanks to the New Years and ML King holidays. February can be as long as 20 days, with the extra day in leap year offsetting the Presidents’ Day holiday, and so on. These numbers need only be inserted once, for every year that number will represent the maximum possible month length. In Column J, the Totals have array formulas that sum the total number of active days for the given month; thus, January had a total of 20 active days. Column K then identifies the month length, by simply subtracting the actual length from the “Long” length, and adding 1. So for January, the 20 days is subtracted from the maximum 21, to get 1; then 1 is added to that to arrive at the “2” found in Cell K3701. Now we can go back up to the DOM Factor calculation.

**8-24 The appropriate Day of the Month Factors is pulled in from the “Inputs” tab, based on the calendar month, the day of activity, and the month length.**

Now that we know the calendar month, the month length (long, medium, or short), and the day of activity, we can use an array formula to pick up the appropriate DOM Factor from the “Inputs” tab. For January 5th (Row 15), which is the 3rd active day in the month, the DOM Factor comes from the “Medium Length Month” table on the Inputs tab; Row 99 of the “Inputs” tab identifies the month of the calendar year, Row 127 of the “Inputs” tab identifies the table as the “Medium Length” table, and the activity day identifies that Row 103 has the 3rd activity day (Cell K15 of the “Calc” tab) for the medium length January.

It certainly was a bit involved to get at this correct set of DOM Factors. I invite you to find a better and easier way. I did it like this as it involved fairly simple straight-forward formulas, a minimum of manual intervention, and the ability to make all the calculations automatically. This setup does the trick.

**8-25 Each days’ seasonally-adjusted volume is calculated, using the normalization Factor, DOM Factor, & monthly Seasonal Factor.**

We next bring in the monthly Seasonal Factor. For January, the seasonal factor was 1.02, and is placed in Column N. Combining the Normalization Factor (Col H), the DOM Factor (Col M), & the Seasonal Factor (Col N), by multiplying the 3 together, gives us the Combined Factor in Column O. When the original sales for January 5th (1.631 B, per Cell E15) are divided by the Combined Factor (1.07, per Cell O15), we get the seasonally-adjusted volume for January 5, 2011: the 1.521 billion found in Cell Q15. (Note Column P, the “Low Factor Adjustment”. We’re going to examine what this adjustment is about later on when we encounter some striking results as we walk through the “Trend” tabs.)

So what does Column Q represent? It gives you the original day’s volume adjusted for: the month of the year it falls in, the day of the month, the day of the week, and the influence of holidays. By adjusting for all these factors you are getting closer to a reading of the “true” level of activity for any given day. And you are better able to compare any day of the year with any other day of the year. In so doing, you get a better read on how to answer the question, “how are you doing?”.

**8-26 Moving Averages are calculated to help smooth the data.**

As we saw at the beginning of this chapter, even after seasonally-adjusting the data to take out some of the predictable “noise”, there still remains a lot of volatility. To decrease these fluctuations, a one-week moving average is calculated, which simply calculates the average over a one-week period.

**8-27 In this model, moving averages collect data from before and after the calculation date.**

However, in this model, (& with every model I’ve ever built), my strong preference is for calculating moving averages using the MIDDLE of the period, NOT the end. As an example, the highlighted area in the table on the right shows the calculation for January 11 as being the average for the period from January 5 thru Jan 11. My issue with moving averages being calculated using prior period alone is that the result is always a bit behind. If a measure is steadily growing, such an average will always be lower than the actuals, while it won’t be if the average is calculated by picking up an equal length period both before and after the date of the calculation.

Here we are trying to capture a 1-week average, or more specifically, a 5-day moving average since the exchange is closed weekends. The moving average should therefore pick up two business days before and two business days after, which is what’s shown here on the left, with the Jan 11th average picking up the period from Jan 7th thru the 13th. The average does not *look* like it’s centered because of the inactive weekend, but it is reflecting what we want to capture – an equal number of active days before and after.

I also need to point out that the formula for computing the average is not an AVERAGE formula in Excel; rather, it is computed as the sum of the highlighted 7-day range of volumes (i.e. Range Q17:Q23) divided by the sum of the Counted days (Range J17:J23). It is computed this way to ensure the zeroes at the weekends are not included in the calculation, and so holidays are also excluded from the average calculation.

**8-28 With a 5-day (or 6-day) workweek, it will be necessary to have moving average formulas skip to capture the desired day range.**

On the last page we saw how the range in the formula for Tuesday the 11th started the prior Friday. Here we continue to Wednesday, with the formula simply copied down, so Wednesday’s average encompasses the prior Saturday through the coming Friday. But when we next turn to Thursday, we see the formula skips, and starts on Tuesday and ends with the coming Monday (which is “0” here because it happens to be a holiday). The skip is necessary in order to ensure two active days precede the moving average and two days come after. Of course, if I was content to just use the past week there wouldn’t be this issue, so that’s the drawback. Obviously it’s up to you the reader to decide how you want to do it. But once the formulas are set up properly the one time, it can easily be copied the rest of the way down the worksheet. (By the way, to do so, you get the formulas right for 7 separate days, then you highlight all 7 days, and with your mouse click on the bottom right “handle” of the range and drag down through the entire worksheet, then just let go and your formulas are done.)

**8-29 A moving 2-week average is also computed to provide an additional option for observing smoothed data.**

I’ve also included a 2-week moving average. As we shall see, a longer period helps further smooth the data and assists with understanding how the data is trending & performing over time.

**8-30 The final calculations on the “Calc” tab address the estimation of trend, and comparisons of it with actuals.**

The trend is estimated in this model (& file) the same way as it was done in the Monthly Trend Model, only on a daily basis. Namely, we bring in our estimates of the annual Growth Rate & Events, estimates that will be made on the Trend tabs. A starting value is inserted on the 1st trend tab, and is picked up here in Cell V2. the starting value is then “grown” using the growth rate & events, with literally 1/365th of the growth rate applied to the prior day’s estimated trend, while event values are applied in full. The Combined Factor (from Column O) is applied to the Estimated Trend to arrive at daily estimated trend values (Col W). These estimated values are then compared against the actuals (from Column E) to arrive at the daily differences shown in Col X. Because of the volatility of the data, there will always be differences, sometimes substantial, but we will always be endeavoring to develop the trend estimates so that the *total* differences, each month, are minimal.

**8-31 Weekends and holidays are hidden on the “Calc” tab, so that charts do not pick up “0” values.**

Once all the formulas are in place, we want to hide any days that have “0” values. We do this so that charts showing data performance are not filled with erratic drops to zero every week. The calculations are not changed by hiding this data, we’re just making it easier to read the charts.

**8-32 As new data comes in, you’ll need to update certain formulas on the “Calc” tab.**

A quick note on updating this file as new data comes in. The figure here shows the NYSE data through the end of 2016. January 2017 data is still absent. What’s notable however are the blank spaces over in Columns Q, S, V, & W. The reason for the blank cells is so that the chart does not have a plunging line as the data drops to zero with the 1st day having no data.

**8-33 Updating the “Calc” tab for the new data requires copying in data at the start of the “old” month, and deleting formulas at the start of the coming month.**

The data update involved copying the formulas into the blank cells – in columns Q, S, V, & W. You will also need to delete the 1st week or so of formulas for the coming month – shown here as blank cells from Feb 1 thru Feb 7. That’s it, though you will also likely want to then turn to the appropriate “Trend” tab to modify, if necessary, the trending, given the latest data.

**8-34 There is a separate “Trend” tab for each year, to accommodate all the days where potential growth & event estimates may be entered.**

We now turn to the “Trend” tabs. You may recall the monthly trend model had the one “Trend” tab. For the daily trend model however, a separate tab is created to handle each calendar year. As you can see from the figure here, there is a place for Growth Rate & Event entries for each day of the year. With such a large quantity of dates to work with, it seems best to simply have a separate tab for each year. The illustration is for 2011, the first year we’re choosing to look at. The “DailyTrendModel” file has tabs for each year from 2011 thru 2020.

The figure here shows the entire tab, so while the design at the top of the page is very similar to that employed in the monthly model, all the calculations of seasonality are absent, making this a simpler tab to work with. Let’s quickly walk through the set-up of this tab.

**8-35 The Growth Rate section of the Trend tab, has a place to enter growth rate changes on any day of the year.**

The Growth Rate section of the model is similar to that we saw with the monthly trend model, with the key obvious difference being that the rows are the days of the month, rather than the months of the year, and the columns are the months of the year, rather than the separate years. Conditional formatting has again been applied to spotlight whenever the growth rate estimate is changed – such as was done here on Dec 27. Light grey borders have been added every 7 days, to help distinguish the days across the month.

The date of the 1st Monday of each month has also been inserted at the bottom, in Row 42. This was added to assist with modifying the chart range so that it might always start on a Monday. So, for example, a chart starting at the beginning of the year should be set to begin on January 3, given that there is no activity over the weekends.

**8-36 The next section compares monthly totals for actuals & estimates. It also brings in growth & event estimates from the Monthly Trend model.**

The “Actuals vs Estimates” section of the “Trend” tabs brings in monthly totals for Actuals and Estimates, for comparison. As before, when we walk through estimating trend, we want to make sure the estimates compare closely with the actuals. Here monthly total actuals - Col E in the “Calc” tab - are compared with the estimated trend value totals, from Col W of the “Calc” tab. The goal here will be to try to keep the monthly difference to less than 2%, and the difference for the year overall to less than 0.25%. Certainly the monthly differences should generally be under 1%, but you will likely find that daily data can be quite volatile and “uncooperative”, so that occasionally wider differences will occur. These differences could be bridged by inserting even more changes in growth rate & event estimates, but then you risk following the data swings with such frequency that it becomes more difficult to get a read on how you’re really doing.

This section also displays the monthly Growth Rate & Event estimates that were made in the Monthly Trend model. I would certainly recommend you build and use that model before you work up the Daily Trend model. It’s just so much more difficult to have a clear sense of where you are and how you are doing using the daily model alone. The monthly model helps provide much needed perspective & reference that assists while estimating trend here in the daily model. These estimates were placed in the “Inputs” tab.

**8-37 Estimates of “Events” are manually inserted in the next section.**

Estimates of events are inserted in the next section below the Actuals vs Estimates area. Again, the estimates can be inserted for any given day of the year.

**8-38 The chart completes the “picture” of the estimation of trend. The volatile daily seasonally-adjusted volumes are shown, in light blue.**

We now turn to the chart, which as usual, is found to the right of the growth & event estimates. We’ll walk thru each of the lines displayed on this chart, starting with the daily seasonally-adjusted volumes. As you can see, they’re pretty volatile; they’d be even worse if we took out the seasonalizing. Because the line tends to be so erratic, it’s downplayed on this chart by displaying it thinly, in a fairly faint shade of blue.

**8-39 Note that the X-axis is formatted to show the starting date for each week.**

The x-axis of the chart is set to clarify the dates displayed for the daily data. This weekly setting is obtained by right-clicking on the x-axis, clicking on “Format Axis”, and then under Axis Options, setting “Major Unit” to “Fixed” (not “Auto”) and to a count of “7” and to “days” (as opposed to “Months” or “Years”). What should the “Minimum” figure be? The “Minimum”, which essentially determines the starting date for the chart, is set to 1/3/2011. Why “1/3”? Because that is the date of the Monday that is associated with the 1st data point in January. The date was picked up from what I’d earlier inserted in Row 42 of the “2011” tab. Note that if I wanted the chart to start with February, I would use the date “1/31” rather than “2/1” as the starting date. Why? Because I want it to be clear that the vertical gridlines on the chart are always associated with a Monday, the 1st day of the week. It gets very confusing if you simply use the 1st of a given month as the starting point and then have a chart’s vertical lines be associated with any and all of the 7 days of the week. It’s a bit of work making sure you start with a Monday, but it’s a lot more work trying to comprehend a confusing chart.

Finally, note the format of the x-axis does NOT include the year. I set it this way because I’ve inserted the year as a heading for the chart instead, so displaying it on the x-axis as well is superfluous. However, if I ever wanted the chart to display more than one year, then the years would probably need to be displayed. To set the date format, again right-click on the x-axis, click on “Format Axis”, then click on “Number”, and then choose “Date” as the Category, and select the desired type. I chose what Excel shows as “3/14”, the most minimal date display they have.

**8-40 And note that the Y-axis will always need to be played with in order to have the chart better focus on where the action is with the data.**

Meanwhile, the y-axis of the chart is also adjusted to have it better display the data in full. A small chart has been inset onto this page to give you an idea of what it would look like if you did NOT manipulate the y-axis. It would be extremely difficult to make out the data behavior without going thru the invaluable effort of adjusting the “Minimum” & “Maximum” values for the y-axis.

**8-41 Due to the quirks of working with Excel, be aware that lines “skip” the weekend on the chart.**

It may have been hard to notice on the previous page, so I’ve zeroed in on the 1st week of January. As you can see, the line plots the points for Monday thru Friday, and then the following Monday. But obviously, while the chart does not show any data points for the weekend, it does display a significant gap. I’m not aware of any method for eliminating this, but I’m not sure that’s necessarily a problem. While it’s a bit irritating to have the long line going from Friday to the subsequent Monday, it’s arguably appropriate since there are days there being “missed”.

**8-42 A 1-week moving average is displayed …**

We next show the 1-week moving average, which comes from Col R of the “Calc” tab.

**8-43 … as well as a 2-week average.**

And then add the 2-week moving average. Each one is useful, but when both are there, it can be a bit confusing. My preference is to have just one moving average line. And between the two, I think the 2-week average line is better. Certainly that seems the case here, for the 2-week average is similar to, while less erratic than, the 1-week average. Both are calculated and displayed so that one can choose which is better. And of course, you could easily change the 1-week line to simply calculate a 3-day moving average or any other length you prefer to use that helps elucidate how the data is performing.

**8-44 By hiding the column of an unwanted 1-week average, it will disappear from the chart, but can be easily re-displayed by simply unhiding that column.**

In order to eliminate one of the lines from the chart, I choose to simply hide the data on the “Calc” tab. The benefit of doing it this way is that if one later wants to see the other line, all that needs to be done is to unhide the column. But if you delete the line on the chart, you have to go back in and add it to the chart, which is a far more laborious procedure. Much simpler to just hide or unhide the column of the line you wish to hide or display. Here, the 1-week average line is removed from the chart by simply hiding its column, column R. Remember, because this model is on Manual Calc mode, when you go back to the chart, you’ll need to hit the F9 Calc button at the top of the keyboard in order to update the chart display to conceal the line.

**8-45 Hiding the 1-week average line makes for an easier read of how the daily data is trending over time.**

Hiding the 1-week line appears to really help clarify the display, and make it easier to determine the estimated trend line.

**8-46 With the adjusted daily volumes & 2-week averages as guides, the chart is ready for determining the estimated trend line.**

The estimated trend line completes the picture of the daily trend model.

**8-47 If you’re wondering whether all this seasonalizing is worth it, here’s what 2014 looks like using the unadjusted actuals.**

It’s easy to think that once you apply something like a 2-week moving average that, of course, the data will be much smoother and easier to read. So why bother with all the seasonalizing, the adjusting for day of the week, day of the month, holidays, and the month of the year. This chart hopefully demonstrates why it still is worthwhile. Yes, the seasonally-adjusted line still has volatility, but nothing like that of the 2-week moving average on the original unadjusted actuals. Yes, much of the spiking for the unadjusted line is due to the enormous influence of those quarterly 3rd Fridays. But notice how differently August behaves when it is not adjusted for its distinctive day of the month pattern; how volatile the data is surrounding the Christmas holiday; and how rarely, if ever, the data appears smoother than the seasonally-adjusted data. It’s a lot of effort to be sure, but it clearly makes quite a difference. Granted, I picked 2014 because the difference was so impressive, but every other year also witnessed a marked deterioration when the unadjusted line was brought out. And perhaps even more importantly, even if the unadjusted line had been smoother, it is not “truer” for it is not adjusting for predictable fluctuations. And by not adjusting for those predictable fluctuations you run the high risk of misinterpreting how the data is behaving over a given period of time.

**8-48 There are a number of guidelines to keep in mind as you walk through estimating trend on the daily data.**

As before, we will want to keep a few general guidelines in mind as we walk through the process of estimating the “Estimated Trend” line. These are guidelines, and on occasion will necessarily or unavoidably be broken. But they remain good goals to strive for along the way.

First, try to limit changes to 1 or 2 per month. Perhaps an event occurs that begins & ends within a calendar month, requiring a positive & negative event estimate – there’s two changes right there. Add another event, or change in growth and this guideline is broken. So yes, this will occasionally occur. But hopefully these are the exception, not the rule. Put another way, hopefully the total number of changes over a given year will add up to less than 20. Once you go beyond that, it really becomes increasingly difficult to LEARN from the data, to truly understand what’s going on. Of course, if you can identify virtually each and every one of those changes – not only know what’s causing it but be able to feel comfortable that the level of change makes sense given your understanding of the metric, then that’s fine. But if you can’t explain most of the changes, you have to wonder how useful the product will be, how much people will likely turn to it for understanding how you’re doing, and why.

The second point is perhaps the most important, and speaks to the heart of what this entire “book” is about. It’s the matter of distinguishing “growth” from “events”. When numbers increase or decrease over time, we quickly call it growth or decline, when so often it’s really a “lift” or a “drop”. Change over time usually occurs as a result of specific events that suddenly cause a given metric to jump up or down, and not due to some steady “growth”. To be sure, there may be an underlying growth that accounts for some of the change, but data rarely behaves smoothly over time, and the changes are usually explained as shifts that can be otherwise called “events”, rather than simply ascribing it to a change in the slope of the metric. The fundamental importance of this distinction will become apparent when we try to estimate the trend line, and to decide whether a change of course should be captured as an event that brings the metric up or down overnight, or “growth” that sees the metric change its slope dramatically. And you will hopefully come to agree, that far more often than not, change should be captured by the “event” (lever) rather than as “growth”.

Third, we want to be sure to capture a change on the correct calendar day. With the gyrations in the data, and the tendency to focus on the easier-to-read 2-week moving average, we nonetheless want to be sure changes are implemented on the appropriate day.

Fourth, as we proceed, we want to make sure the estimated line is close to the actuals. This is tracked by the Actuals vs Estimate comparison section. And here we will want to try to keep the difference between the two to less than 1% each month, and to accept perhaps up to a 2% difference as an exception. Now, your data may be far more volatile than the NYSE sales volumes, and even 2% differences will be elusive. If so, try to establish a fair maximum and stick to it. This will still help guide your estimations so that they still align fairly well with what the “true” trend of the data may be.

Finally, while 1-2% may be the goal for the differences on a monthly basis, try to have the difference for the year as a whole contained to perhaps less than 0.2%.

**8-49 The Trend tab will be formatted in these pages to help maximize the size of the chart, while being able to clearly see the changes & entries made.**

To begin, I want to point out the formatting changes I’ve made to the worksheet so you are better able to see changes and entries. I’ve hidden the columns for the growth & event entries from May thru November. Also, I’ve severely narrowed the rows in the “Events” section by setting the row height at “1” for every row except the rows where event entries are made (so you can anticipate entries on the 2nd, 18th, & 22nd days); the 1st & 31st will always be shown fully as they bookend the month.

**8-50 We start by setting the approximate level of where the first examined year begins, set here at around 1.6 billion.**

So to start out, we need to set the beginning level, which of course would correspond to the approximate level activity is at when the period begins. This is manually input in Cell N9 near the top of the page. I’ve used 1.6 billion as starter, very close to the approximate level it appears to average over the 1st week or so.

**8-51 We then set the growth rate, initially using the growth rate established in the Monthly Trend model.**

Next I need to insert a growth rate. I recommend initially using the growth rate established in the Monthly Trend model. Arguably, the growth rate used in that model should probably be what’s used in this model as much as possible. Certainly doesn’t have to be an identical rate, or applied over an identical time period, but generally speaking you will likely do well to try to come close to matching the growth rate used in the monthly trend model. This rate was based on the “higher perspective” offered by looking at monthly data.

We earlier input the Monthly Trend Estimates in Row 50, as reference. We can see from the highlighted row that a growth rate of -10% was used for the entire year. So that certainly seems like a good rate to set out with.

**8-52 One of the great temptations, usually best avoided, is to chase apparent steady changes in growth rates, such as the one highlighted here.**

I mentioned earlier that generally speaking there will be few changes in growth rate applied to the estimation line, and that usually these should correspond with changes inserted in the Monthly Trend model. Looking at the behavior of the 2-week average line in late January to early February, it appears to show a steady decline occurring. However, the reason for this decline is precisely because we are looking at a moving 2-week average. When a metric suddenly jumps or drops, the moving average will take much longer to accomplish the same level of rise or fall. So the steady drop here really is the result of a fairly sharp drop occurring right at the end of January, & at the very start of February.

**8-53 It can require some breathtaking growth rate values to accurately capture a significant change in growth. An annual rate of -250% works here.**

Just for the sake of argument, let’s try to create a trend line that replicates the apparent decline in sales. As you can see, an annual rate of -250%, applied on January 26th, seems to match almost perfectly the general slope of the 2-week average line. How can the rate be less than -100%? Because the rate is an annual rate, so each day 1/365th of that rate is applied. So actually each day is seeing a drop of something slightly under 0.7%. The growth rate then is returned to the former rate of -10% the following February 10th, just as the 2-week average line stops declining. Aside from the obvious fact that the 2-week average takes two weeks to capture a sudden shift in the metric, one also must question the usefulness of an annual rate like “-250%”. Imagine the kind of alarm that would stir in the executive suite. And how would you answer the question, “what does that rate mean”?

**8-54 While it may be clear that inserting an “event” is the appropriate approach, you then have to consider the target level to obtain.**

All right, so hopefully you can now agree that usually you will want to apply an “Event” rather than a change in the growth rate to capture how performance is trending over time, day-by-day. The next question then becomes, to what level do you raise or drop your trendline, when an event occurs? This question was relatively more straight-forward when we were looking at monthly data for there were so few data points to consider, and we were always trying to avoid applying too many shifts and changes to trend. But with daily data we find ourselves going back and forth between the daily data itself, and the moving average on that data. A perfect example of this dilemma concerns where to shift the trendline to at the start of February. If we look at the daily data, we see that most of the data points coalesce around the solid pink line, just below the 1.4 billion sales level. On the other hand, the 2-week average never dips below the line described by the dotted line on this chart. Of the two, which is the more appropriate or correct line? I would suggest the dotted line is the more appropriate, for it better aligns with the general *trend* in the data. While much of the daily data is lower over those 1st two weeks or so of the month, the 2-week average line indicates that the average never drops below where the dotted line has been inserted. Indeed, we are talking about a period of approximately two weeks, and the midpoint of that period is right around that minimum level; accordingly, the dotted line definitely seems an appropriate target level.

Now this is just one example, and every set of data will behave differently. The point I wish to make here is that setting the trendline involves looking at both daily behavior and the 2-week (or 1-week moving average), that generally the daily data is most valuable for informing when to apply an event, while the moving average is key for determining target levels.

**8-55 The timing of events corresponds with the day best reflecting when a shift occurs, while the level of change aligns with the 2-week moving average.**

This charts illustrates the point we were just making, that the timing of the event here is the 2nd of February, for that is when sales volumes realize a markedly lower level. Yes, one could time one drop for Jan 31st, with a second drop on Feb 2; but then you run the risk of having too many data shifts to deal with, and more importantly, too many to explain. An event of -9% (Cell D112) brings us to the approximate level we desire; playing with -8% and -10% left us at too high and too low a level, respectively. Before leaving this change, notice where the trend line now continues. Obviously there is a big shift up we’ll need to apply around Feb 18th, but it’s interesting to observe how the trend line aligns with the daily data from March 22nd on – perhaps a bit high, but not by much.

**8-56 As there are no changes occurring in January, we should be able to have an actual vs estimate difference for that month that is very close to 0.**

While the drop we’ve inserted for the start of February looks about right, we need to go back and take another look at January. By having no event or growth rate change entered for January, we’re essentially committing ourselves to a trend line that should be at a level that almost exactly matches the actuals in that month. But we can see that the estimate is almost 2% higher than actuals. We want to reduce this gap, and we do so by reducing our starting point in Cell N9. However, when we drop the starting level, we’ll also then need to make sure to also reduce the -9% drop for Feb 2, in order to ensure the trend line continues to hover around the level it’s at here during February.

**8-57 “Playing” with the starting level gets us to a much reduced gap for January, while tweaking the February event obtains the right level to start the month.**

To get to the right starting level in January, I simply “played” with the amount in Cell N9. I even changed the unit display so I could get the difference down to zero at the 1/100th of a billion level (which meant the month totals would differ by less than 5 million). This will probably be the only time I try to obtain this level of precision. Normally it is just not possible or appropriate to try to get it so exact. But here we are dealing with the starting level, and it seems just fine to at least try to set out at the right level, down to the nearest 1 million. And this was accomplished by dropping the starting level initially down to 1.58 billion, and then ultimately down to 1.576 billion. That gets us to an almost exactly 0 difference for January. I then needed to bump the event amount up to -8%, from -9%.

**8-58 Much of the trend estimation process consists of working with the timing and values for events, making sure each month’s actual vs estimate gap is minimal.**

And so it goes. The trend estimation process is a constant interplay between setting events, and the occasional growth rate changes, such that the difference between the actuals vs estimates is minimized. Again, we’d like to try to contain each month’s difference to 1-2% at most, and to keep an eye that the differences aren’t always positive or always negative, that they swing back & forth so that the year total difference should be very close to zero.

In this example with the NYSE sales, I looked ahead and could see that I wanted an event on March 21st that would drop the level to around 1.4 billion, taking the trend line about to where it is on the chart here. I also needed to enter a positive event on Feb 18, corresponding with the date the jump in sales occurs. I found a “+17%” event for Feb 18 and a “-16%” event for March 21st appeared to give me a line that fit very well with the running 2-week average, the timing of the daily jumps, and which gave me differences for February and March of around 1% and 0% respectively. That seemed pretty good, and with that the 1st quarter of 2011 was complete.

**8-59 Once a given period is done, reset both chart axes to reveal the next period. Show the prior period’s last month - for reference, & potential modification.**

We now shift to the next few months. The chart needs to be reset. You will likely find it helps to still show the last month of the prior period. So we had earlier worked on Jan thru March. I’ve set the next chart to still show March, as well as April & May, & part of June. As we’ll soon see, maintaining that visibility is helpful as you may find it necessary to go back and modify the prior period.

Of course, you’ll also probably need to reset the y-axis range; here the minimum was dropped from 1.2 billion to 1.1, while the maximum was dropped from 2.2 to 1.8 billion.

**8-60 Sometimes the actual vs estimate gap is caused because of an estimate used in a prior month.**

As has hopefully been made clear by now, this approach to estimating trend requires a constant back-and-forth review and modification of the estimates inserted for growth rate & events. Some choose to have software do this task for them – fine, but do you fully understand what that software is doing? And does the output of that software make sense, and how useable is it? With this approach, you’re essentially always asking yourself: what is the data doing, why does it change as it does, and do those changes & explanations make sense? If you can answer those questions, you’ve built a solid foundation for answering the question, “how are you doing?”.

Looking at the next NYSE period, we see an example of one of the situations you are likely to encounter: the gap for a given month is created by estimates made for a prior period, even though you may have liked the estimate you’d previously made. So here we had used “-16%” as the event estimate for March 22. But we can see that this estimate leads to a sizeable gap of 3% for April (Cell F47). We’re going to want to increase the March 22 drop (in Cell E132).

**8-61 Modifying a prior estimate can help reduce the gap for a given month, though it can then lead to larger gaps in subsequent months.**

As noted, we want to increase the amount of the drop for March 22, changed here from “-16%” to “-17%” (Cell E132). I tried “-18%”, and while that made for a smaller gap in April (of just -0.1), it also led to the estimate being low by 1.0, or “-3%”, in May. Consequently, I’ve compromised here by using “-17%”. While a +2% actual vs estimate gap remains for April (Cell F47), it is offset by a -2% gap for May (Cell G47).

Of course, I could reduce the gaps in both months by making additional changes. But I chose not to as the 2-week moving average appears relatively smooth over this period. While there is some slight shifting up or down every week or two, none of it seems very significant, so I think it warrants having the one simple line track across the whole two month period. You may disagree, in which case you may make additional modifications, or perhaps slightly alter the settings I’ve employed. There really is no one right answer. However, if you knew what was going on, if you had the inside track on the behavior of the time, then you would be in a much better position to determine if and when more modifications were appropriate – simply because you knew it made sense and you knew their application better fit the behavior of the data.

**8-62 The estimation process continues, running 3-4 months at a time.**

The estimation process continues into the next period. Ordinarily, I would have had the next period – on the chart – run thru Aug 15, or even Sep 15. But I cut off the end at 8/1 here because there is a huge increase coming, and the chart scale will have to widen substantially to accommodate it. By confining the period to May thru July, I’m able to maintain a better look at the daily cycling of the data on the chart. And here I’ve chosen to insert a 9% bump at the end of May (Cell G141), an 11% drop on June 20 (Cell H130), & a 20% bump on July 27 (Cell I 137). These changes appear to fit the data well. Also, there are small differences in June & July. As for the April & May gaps, while they are a bit larger than I’d like, they nicely offset one another, so that combining the two months results in almost zero difference. Now on to the August sales surge…

**8-63 When a prolonged spike occurs, the lift should equate with the total net increase over the period.**

Inevitably, there will be occasions where a metric experiences a dramatic rise (or fall) for a prolonged period of time. When these occur, if you can, and it’s appropriate, you want to insert a single lift (or drop) that captures the net change for the entire period. The early August sales surge shown here provides a perfect example. From Aug 4 thru Aug 12, sales were over 2 billion shares each day, reaching a peak of almost 4 billion on Aug 8. Rather than trying to capture the increases or decreases each day, or every other day, you want to set the lift for the period such that the estimated trend values sum to a total that is very close to the actuals. That is what has been accomplished here by inserting a lift of 105% on Aug 4 (cell J114); a drop of 37% was entered on Aug 12 (Cell J122), designed to capture the new level of sales that occurs over much of the rest of August. How did I come up with 105%?

**8-64 To verify a prolonged lift is set correctly, make sure that the total Estimate vs Actuals Difference is close to 0.**

The way to make sure your estimate of a prolonged lift has been set correctly is to toggle back to the “Calc” tab and verify that the total Estimate vs Actual Difference is close to 0. The figure here shows the detail from the “Calc” tab for this period. You can see the event estimates in Col U, the 105% lift on the 4th (Cell U226) & the 37% drop on the 12th (Cell U234). And you can see the estimate vs actual differences highlighted in Col X. The goal here is to have the sum of these cells be close to 0. You can see that the estimate for the 4th is high by over 700 million shares, and that it is low by over 700 million shares on the 8th, but the key is that the total difference for these seven days combined is just 21 million. Not bad given that the total share sales for the period was over 20 billion. Getting this difference close to 0, while setting the drop at the end of the period so that it matches the approximate level sales are at going forward, is precisely what we want to achieve.

**8-65 If a 1-day spike is large enough (a change of perhaps 50% or more), it may be appropriate to insert an “event reversal”.**

The rest of the summer period is estimated above. A lift - of 21% (Cell L113) - was inserted for the Oct 3-11 period, again verifying on the “Calc” tab that the total difference for those days was minimal. Drops were added in late Aug & at the start of September. As you can see highlighted, a 1-day spike was inserted for Sep 22. You’ll also notice that for virtually the entire month of September, the estimated trend line is slightly below the 2-week moving average. Should it be raised? No, for the Estimate vs Actual difference for September is close to 0, & there’s a slight positive difference at that. I certainly could have used a lower decline on September 1: using “-10%” instead of “-12%”, the difference is again a positive 0.1 billion; but that difference is obtained by leaving out the Sep 22 spike. That’s a judgement call. What would the latter half of Sep look like if there hadn’t been that 1-day spike? Well, the 2-week moving average line would have likely ended up almost exactly where I have traced the estimated trend line here. I’m content to run it like I have here, for I want to capture the substantial 1-day spike in my estimated trend, which I’ve done with the 50% lift estimate (Cell K130). The Event Reversal Formula (see Chapter 6) was applied to derive the -33% drop the following day (Cell K131).

**8-66 A series of step functions captures the decline over the rest of the year.**

As you can see, I’ve finished off the year by inserting a few more step functions. These are inserted at the approximate dates that drops occurred, and are of a magnitude that the monthly estimate vs actual differences are close to 0.

**8-67 One might argue there appears to be a steady decline occurring through much of the last few months – it’s usually better to ignore these.**

If you look again at the last few months of 2011, it is quite striking how there appears to be a fairly steady downward trend over this period. The heavy dashed pink line attempts to approximate this decline. What does it look like if we have the model capture this alternative perspective on sales behavior over the last quarter?

**8-68 Even over extended periods of time, severe annual growth rates are required to capture significant declines.**

The figure here displays what it takes to have the model trace the kind of decline suggested by the slope of the 2-week moving average over the last few months of 2011. We saw this earlier, when we looked at the decline over the late January-early February period. Again we see that it requires extraordinary values to capture the rate of decline. Here an annual growth rate of “-170%”, applied on Oct 19, thru Dec 29, seems to fit the data.

There are three major problems with this alternative vision. First, there is the difficulty of getting your head around a “growth rate” of “-170%”. Of course, when applied on a daily basis, this actually equates to just under ½ of 1%; (170% divided by 365 days in the year). But how meaningful is such a rate? It seems unlikely that senior management would well understand the meaning of this, and much more likely that it would either prompt enormous alarm, or a concern by them that perhaps they need to look elsewhere than you for an understanding of how their business is faring.

Secondly, you have to think again about the distinction between the words “growth” & “event”. “Growth” implies a steady state, that one’s business is headed on a particular trajectory. And because “growth” is perceived as occurring over time, as opposed to at just one or two specific moments in time, one would understandably conclude that the situation is continually getting worse and worse each day. Meanwhile, when a drop is assigned to “events”, the business decline is perceived as being due to a one or more major negative occurrences. There’s a very different mindset about what is going on, depending on whether you describe a decline (or big rise for that matter) as being driven by “growth” or “events”.

Finally, the severe downward slope here really does not describe how the data is behaving. If you look closely at the daily line, you can see that it really does tend to coalesce around 1.5 billion during much of October, around 1.35-1.40 billion shares during November, around 1.25 billion shares in early December, and to settle out at around 1.15 billion shares at yearend. Applying a series of “events” better fits the data behavior, and certainly provides a much better understanding of how the data behaves over the last few months of 2011.

**8-69 Sometimes spikes can occur in and around holidays. These spikes can be magnified by their holiday factors.**

I want to now address an issue that can sometimes arise with the daily data, an issue that is driven because sometimes the daily combined factors can be very low. Here is an example of this problem. In 2013, the seasonally-adjusted sales for July 5, the Friday following the July 4 holiday, was almost 2 billion. Such a huge number suggests that the actual sales that day were enormous. But that was not the case.

**8-70 Sometimes very high seasonally-adjusted daily values can be driven by very low Combined Factors.**

The driver of this issue can be better recognized by looking at the calculations behind it. We can see from this detail of the 1st weeks of July 2013, that there is a very low Holiday Factor for the Friday following the July 4th holiday, (a holiday factor of 0.41, per Cell G927). This is not surprising, as one can well imagine many taking off that day to enjoy a 4-day weekend. This low holiday factor leads to a similarly low Combined Factor of 0.43 (Cell O927). Actual sales that day totaled 850 million (Cell E927). We can see from the other days surrounding that, by itself, the volume was a bit below the other days (by around 10-20%). But when each day’s Actual value is divided by its Combined Factor, suddenly the 5th emerges as a huge spike, with its seasonally-adjusted volume of 1.957 billion (Cell Q927) coming in at almost double all the other days.

How shall we address this? With what I call a “Low Factor Adjustment”.

**8-71 Developing a “Low Factor Adjustment” starts with determining what *actual* sales would “normally” be.**

The development of a “Low Factor Adjustment” is going to entail determining the gap between what sales normally would have been, and their actual value, then adding that gap to the trend. To see what that looks like, we start with determining what sales would normally have been on the day with the low Combined Factor. We have determined the Estimated Trend amount on July 5th is a little over 1.1 billion shares. July 5th has a combined factor of 0.43, implying that the “expected value” on the 5th would be 43% of the estimated trend, or the 0.481 billion shares marked near the bottom of the figure.

**8-72 Next, we measure the gap between the normal sales and the actual sales that occurred.**

Next we mark what the Actual sales were on July 5th, and measure the gap between the “Normal Actual” amount and the “Actual” amount. Here we see that gap was about 0.369 billion shares.

**8-73 A proposed seasonally-adjusted sales is the sum of the Trend amount plus the Gap.**

Then we add that Gap amount to the Trend amount. This gives us a Proposed Seasonally-Adjusted Sales for July 5th of 1.477 billion shares We’ve asked what sales would have been if the day had behaved “normally”. We then determined the difference between the actual sales and that normal amount, and then added (or subtracted) that difference onto the Trend.

**8-74 The “Low Factor Adjustment” is the amount required to get from our original seasonally-adjusted sales to the Proposed Adjusted Sales.**

Finally, we can determine the “Low Factor Adjustment”. It is going to be the amount by which we need to change the Original Seasonally-adjusted Sales in order to get to the Proposed Adjusted Sales. In our running example, originally the seasonally-adjusted sales was calculated as a whopping 1.957 billion shares. When we reduce this by a “Low Factor Adjustment” of 0.480 billion, we arrive at our Proposed Adjusted Sales of 1.477 billion.

**8-75 The formula in the “Calc” tab calculates this “Low Factor Adjustment”.**

Now that we’ve seen on the chart what this adjustment looks like, we return to the “Calc” tab to see what it looks like there. We can see that the formula in Col P has arrived at the desired Low Factor Adjustment of a negative 0.480 billion (Cell P927). We can also see, in the adjacent Cell Q927, that the new seasonally-adjusted volume is now the desired 1.477 billion shares. I refer you to the worksheet on this website to see the formula itself.

**8-76 Note that the Low Adjustment Factor formula uses two key variables: a “Maximum Low Factor” and an “Adjustment Level”.**

The calculation of the “Low Factor Adjustment” relies on two key variables that are placed at the top of the column: a Maximum Low Factor and an Adjustment Level.

**8-77 The “Maximum Low Factor” determines when a “Low Factor Adjustment” is to be applied.**

The “Maximum Low Factor” determines when a “Low Factor Adjustment” is going to be calculated. The point of the Low Factor Adjustment is to ensure that we avoid unusually high (or low) adjusted values being calculated, simply because the math results in a magnification of the amount. When the Combined Factor is 0.5, adjusted sales will double the actuals; if the Combined Factor is 0.1, adjusted sales will multiply the actuals by 10. But if the Combined Factor is 0.9, say, then adjusted sales will be about 11% higher than the actuals. We don’t want to be adjusting every calculation, we just want to focus on those where the adjustment is large. I’ve applied here a Maximum Low Factor of 0.5, meaning that anytime the Combined Factor is above 0.5, no additional adjustment will be made, but when the Combined Factor falls below the 0.5 maximum, then a Low Factor Adjustment will be inserted. That’s why you can see here that the only day that was adjusted is July 5th. Note that in the Excel worksheet, I’ve applied conditional formatting to the Combined Factors in Column O so that the cell will be highlighted whenever the Combined Factor falls below the Maximum Low Factor inserted in Cell P1.

Now why a maximum of 0.5? Well, this value is certainly somewhat arbitrary on my part. I wanted to make sure no day sees an overly large adjustment being made due to a low Combined Factor, and I felt a factor that doubles the actual value or more, is sufficiently impactful enough to merit some kind of adjustment. By all means, you should play with this and see whether you think a lower, or higher, maximum feels appropriate to you.

**8-78 The “Adjustment Level” indicates the percentage of the Low Factor Adjustment that is to be applied.**

The “Adjustment Level” in Cell P2 indicates the percentage of the “Low Factor Adjustment” that is to be applied. The input value here of “100%” implies we will take the full adjustment. A smaller (or larger) amount would imply the adjustment would be accordingly reduced (or enlarged).

**8-79 How much the final seasonally-adjusted sales is effected depends on the “Adjustment Level” setting.**

The impact of the “Adjustment Level” setting inserted in Cell P2 of the “Calc” tab can be observed in this figure. If the Adjustment Level is set at 0%, we return to the original high level of nearly 2 billion shares. Set at 50% or 100%, we end up at the lower amounts shown on the chart. This is a subjective decision; for now, simply using “100%” seems a good Adjustment Level to apply. But you may find that your own data warrants modifying this Level, which is why this flexibility has been added to this model.

**8-80 While adjusted sales are still well above trend, they are not nearly as much so as they are without the application of the “Low Adjustment Factor.**

In conclusion, we can clearly see that applying a “Low Adjustment Factor” can make a substantial impact on seasonally-adjusted sales for a day where there is a low Combined Factor. The “Proposed Adjusted Sales” are still well above trend, but we can see that there are a couple of days in mid-June that come in slightly higher. So we still see a large value for the day, but not nearly as great as originally.

It should be noted that this adjustment works equally in both directions. If the Actuals for July 5th were very low, the adjusted sales for the day would also be extremely low. Applying the “Low Adjustment Factor” will raise the adjusted sales amount, leaving it still below trend but not nearly as much as it would be without the additional adjustment.

**8-81 Let’s summarize the key points about estimating daily trend.**

We’ve now completed describing the approach for estimating trend daily. Let’s summarize the key points about this approach:

1. The approach is very similar to that used in the Monthly Trend model. As with the Monthly model, it entails “playing” with estimates of the annual growth rate and events to track how the data behaves over time. Similarly, it relies on seasonally-adjusted data, and it displays the running differences between actuals and estimates. On the other hand, there is no “live” tracking and changing of the monthly seasonality. And the tracking of trend relies heavily on a 2-week moving average, (though, depending on how cooperative your data, it may be more appropriate to use a moving average covering a shorter – or possibly longer – length of time).
2. I recommend most of the adjustments to trend be captured as “events” rather than growth rate changes. You will frequently find that attempts to change growth rates require using values that are incredibly high and often three figures in length (i.e. over 100%). The problem with these high growth rates is that they can be difficult to interpret and understand; what does a 50-100% growth rate mean to a company whose annual sales typically change by 10% or less from one year to another? Also, whenever any change to trend is input, it begs the question, what caused it? How do you explain this x% rise or fall in sales?
3. Related to the events is the caution that you try to avoid inserting too many changes to the trend line. Every change input calls for an explanation: what caused this change, is the amount of change reasonable, etc. The daily metrics, even after being adjusted for the calendar and seasonality & so on, will inevitably behave with notable & unavoidable volatility. You’ll want to try and resist the desire to pursue some of that volatility.
4. You will likely find it helpful to rely on the previously done Monthly Trending for guidance. This is definitely a case of “seeing the forest for the trees”, and daily data is definitely about being in the trees. Nonetheless, you need to go there, and track that daily data if you are to understand what is driving your performance. This need especially applies to those organizations in the habit of following daily data, organizations where there truly are major things going on almost every day as is the case in highly competitive industries like telecommunications, for example.
5. As ever, as you derive your estimated trend line, you want to try to minimize the differences between actuals and estimates. However, do not let this goal force you to input lots of changes to trend so that you can keep the differences as close as possible to zero. Sometimes a small difference of 1-2% (or more) may be a good thing if you are unable to explain the changes it would take to have the trend line fit more closely. If you can’t avoid larger differences however, do try to have the differences offset quickly. So, for example, if you’re high by 2-3% in August, hopefully you find yourself low by a similar amount in September, or by October at the latest.
6. Be aware of the “Low Factor Adjustment”, an adjustment built into this model to avoid having relatively quiet holidays cause a disproportionately large impact on seasonally-adjusted values. You certainly should consider raising or lowering the 0.50 maximum low factor amount that is in use in this model.
7. Finally, be patient. I’ve said this before, but this really does apply to the effort to estimate trend on a daily basis. There’s a lot of noise, and there’s a lot of values you’re trying to follow. If you’re open seven days a week, that’s over 1,000 data points in less than 3 years. Always keep in mind the purpose of trending your daily data. You want to keep on top of what’s going on with your performance. You want to learn from your data so you may perform better going forward. And you want to be able to explain and illuminate your performance to senior management. This model can certainly do much to help you with these goals. But you will need to resist the frustration that comes with following “ill-behaved” data. Good luck!

**Part IV: Applications**

Thus far, we have examined in-depth how to go about seasonally-adjusting your data, for both monthly data & daily data. The focus has been on the process of developing seasonally-adjusted numbers that will allow you to learn more from your data. Well, now that you can seasonally-adjust your data, what do you do with it? That’s where we turn now in Part 2. We will look at how to apply the information gained from the seasonal adjustment process. We’ll start by taking a quick look at reporting, then examine how to use the results to go about estimating your price elasticity. We will also look at retention, and of course, at forecasting. For the most part, the applications will focus on monthly performance, rather than daily. But most of the guidance here can equally well be applied to analysis of the daily data.

**Chapter 9: Estimating Trend**

In this chapter, we look at monthly reporting.

**9-3 This chapter will briefly review what to present in reports, rather than on how to present it.**

The focus of this chapter will be on what to present in your reports, not how to present it.

**9-4 There are numerous excellent resources for guidance on how reports & charts should look, and not look.**

There are numerous excellent resources to refer to for ideas on how to present your data, and I highly recommend you’re looking at them. Stephen Few is a master at focusing on what to present and how to present it, with emphasis on only showing what is necessary and excluding everything else. Cole Knaflic has an excellent book on Storytelling with Data that helps show how to walk through the presentation of your data so that your audience isn’t overwhelmed, and instead well understands your message. Chandoo.org is a website with all manner of info on using Excel. His site, and audience, have numerous helpful suggestions on keeping reports clear and simple, as well as advice on how to perform the appropriate analysis in Excel. Edward Tufte of course is often referenced as a classic guide on presenting data.

**9-5 Answering the question, “how are you doing?”, requires three fundamental elements:**

What I want to focus on here is what you include in your report. Here I go back to where we began: what is the purpose of a report? I would argue that ultimately the primary purpose of a report is to answer the question, “how are you doing?”. That answer not only gets at where your business is at now, but would also speak to how well a given recent initiative is performing, or where your business appears to be headed. The answer to the question, ”how are you doing?”, involves three critical elements.

1st, it requires being able to quantify the approximate level you are at today, for your given metric. 2nd, it involves providing what your approximate growth rate is, or what the current slope is for your data set; this can be very difficult to determine, even when looking at the past, but this trend model will include your current growth rate estimate. 3rd, it entails identifying and quantifying any recent events, occasions where a step function up or down has occurred. Combine these three elements, and you have provided a clear picture of how you are doing. And by knowing how you’ve done in past, and where you’re at today, you are in a much better position to postulate on where you’re headed.

**9-6 All three key reporting elements can be found on the “Trend” tab, with the current level being read from the chart’s Estimated Trend line.**

All three of these key reporting elements are found on your “Trend” tab of the Trend Model file. The current level you’re at can be read from the chart. It is quite simply the value of the estimated trend for the current month. Here, as of January 2017, the current level is approximately 23.5 billion shares per month. Your current growth rate is the Growth Rate you’re using in your model for the current month, so in this example taken at the end of January 2017, the current growth rate is the 1.0% we find located in Cell S13. Finally, recent events are of course, those you’ve input in the Events section of the worksheet. Here we have placed a 1-time spike (& reversal) in Nov 2016.

It could be that you’d like to have this “Trend” tab explicitly display all these 3 elements. While I haven’t done so here, it would be easy for you to use a given set of cells or columns to display this vital info. Such info could be displayed at the top of the worksheet, perhaps immediately above the chart.

**9-7 The “Trend” tab’s chart can very effectively allow your audience to “see” how you’re doing.**

The chart on the “Trend” tab certainly offers a very effective way of allowing your audience to “see” how you are doing. When first presenting this information, I would recommend you slowly walk through each line, just as I did back in Chapter 6. So, show them a blank chart with just the two axes, then insert the actuals, then the seasonally-adjusted line, then its 3-month moving average, and then finally your estimate of trend.

**9-8 A summary chart might exclude one or two of the lines used in the “Trend” tab, and would likely show some forecast period.**

You’ll want to give careful thought to whether all these lines are necessary. It could well be that your audience is better served to have one of the seasonally-adjusted data lines removed, and perhaps the actuals aren’t necessary. However, keep in mind that your forecasts are only going to have two versions – the trend, as indicated by the red estimated trend line in the example here, and the forecast actuals, indicated by the light grey line. There is no difference between your forecast trend line and a forecast of the seasonally-adjusted data. We’ll be getting into forecasts much more in a later chapter.

**9-9 The trend chart can be supplemented with info describing past, and future, events.**

Your trend chart can also be used to call out specific events – what they are, and your estimate of how large they were, or will be. In this example, I’ve chosen to describe certain events that resulted in significant rises or drops in sales volumes. Unfortunately, my lack of inside knowledge of this industry leaves me poorly prepared to give better explanations for the data behavior. I would trust an industry analyst would offer better explanations, and provide better forecasts. The forecast here is obviously all about demonstrating how a forecast trend might be presented.

**9-10 One of the most valuable outputs of this seasonally-adjusted trend model is the identification and measurement of events.**

I’ve mentioned it frequently before, but I think it still bears repeating, one of the most valuable aspects of this trend model is its identification and quantification of the events, or step functions, that occur with virtually any given metric. Most of the time, these events are the drivers of what makes your performance change over time – for better or worse. When you see and quantify an event, you must ask yourself why?, what is causing this sudden change to occur. You may know from what you’ve heard around your organization, or you may need to do some research to ferret out the answer. But by using this model, you should be the very first person aware that an event has occurred, and have in hand a good estimate of the amount of impact.

**9-11 Your independent measurement of events can provide a helpful validation – or perhaps an alternative result – to estimates others have made.**

Of course, there will no doubt be occasions when you come up with estimates for an event that differ from estimates developed by another department. With all due respect to Marketing departments, I’ve never known one to come up with an estimate for the impact of a marketing campaign that is smaller than the estimate I’ve derived. On the other hand, there will also be occasions where your estimates closely match those from others, in which case it provides a validation of your estimate, and theirs.

**9-12 Event measures can be made on numerous types of drivers of step functions.**

There are all sorts of events that you’ll find need measurement. Internally, these can be anything from ad campaigns & promotions to new products or processes. External factors can cause a sudden shift, be it from new regulations or the activity of a major competitor, or natural disasters or good or bad news in the press. Hopefully you design your reporting in a way that you clearly show when the event took place, what the event was, and your estimate of its amount.

One special kind of event merits attention all its own: the impact of price changes, to which we turn next….

**Chapter 10: Estimating Price Elasticity**

In this chapter, we look at how you can use the trend model to derive one of the most important metrics every company should know: the price elasticity of its products.

**10-2 The price elasticity measures how much demand changes given a change in price.**

The price elasticity is one of the most fundamental concepts in economics. The price elasticity measures the percentage change in demand given a change in price. For example, if you increase your price by 10%, by how much would the volume of sales for your product change, all else being equal. A price elasticity of -0.6 implies that the 10% price increase would result in sales volumes dropping by 6%.

I have been surprised to find in my consulting practice that many organizations do not know their price elasticity. Wouldn’t you want to know what to expect if you were contemplating a price change? This chapter will walk through the steps of how you can use this model to derive an estimate of your price elasticity. The example I will use in this chapter will be completely hypothetical. You will probably find your own data will not reveal such clear and “clean” results as those you find here, but I’ve intentionally had the data behave well here to help clarify the process.

**10-3 Estimating your price elasticity involves a number of steps.**

Estimating the price elasticity involves a number of steps. We start out by bringing in the price change data and identifying occasions when other major events took place at or about the same time; we’re going to want to eliminate those occasions due to the difficulty of isolating the change in sales due to the price change versus the change due to the other simultaneous event. Next we’ll revisit the original event estimates to get more comfortable as to their timing and measured impact. Then we’ll plot the results and identify any outliers to remove from the final price elasticity estimate. We’ll end by looking at determining the net revenue impact of the price change. We’ll see that for a product the impact is fairly straightforward, but when dealing with customers we’ll see it becomes much more involved.

**10-4 As an example for demonstrating how to estimate price elasticity, we use a hypothetical firm and trend its sales history.**

As mentioned, we’re going to use a hypothetical firm and sales history to demonstrate how a price elasticity can be estimated. Data was made up, and then trended using the same technique we described in earlier chapters. The complete 2001 to early 2017 history is shown here. Looking closely at the light gray line, you can see there is a fairly pronounced spike during the summer months. And notice that the seasonally-adjusted lines and the estimated trend lines are very close to one another – there isn’t as much volatility in this data as you might typically encounter.

**10-5 To minimize bias, it is preferable to complete the data trending before you start the process of estimating the price elasticity.**

I highly recommend that before you endeavor to estimate your price elasticity, that you first go through the exercise of trending your data history, complete with your estimates of Growth Rates & Events. Doing this beforehand helps reduce the potential for bias that can creep in if you are “looking for” an event impact consistent with a significant price change.

**10-6 The process begins with bringing in the data points for past price changes.**

We begin with the obvious – bringing in the price change information. As one might expect, the more price points the better. Here you just want to identify the month of the price change, and its percentage amount.

The calculations shown here take place in the “Calc” tab of a file called “XYZ Model”. The trending and seasonalizing of the data took place in the hidden columns to the left of the Price Change, column AA.

**10-7 Next you need to identify outliers – occasions when other significant events were occurring at around the same time as the price changes.**

When you are estimating your price elasticity, you want to be sure to isolate those occasions when no other significant event was occurring at or about the same time. For example, in an effort to bump up sales, say you ran a major ad campaign at the same time as you rolled out a price drop of 10%. How would you know how much of the resulting sales lift was attributable to the price change versus the ad campaign? If you can’t isolate how much of the sales lift is specifically do to the price change alone, you really must exclude that data point from your analysis.

I would suggest you review the dates of past price changes and determine if you’re aware of other major events that were going on at the time. These events could be internal – such as marketing campaigns, product introductions, etc. They could also be from competitors, where you’re aware of a major competitor introducing their own significant marketing campaign or price change. Now, competitors will always be introducing major roll-outs – that’s the nature of business. You want to try to identify those instances where you’re pretty confident that your own business was impacted as a result. Again, we are trying to make sure that our price impacts are not tainted by other events occurring simultaneously.

You’ll also want to exclude external events – events that were outside of your industry’s control that may have significantly impacted your sales. Examples here would be major weather events in a specific state, or something like 9/11 or the late 2008 financial crisis.

I recommend you describe these outliers – just a couple-few words should suffice, but will be helpful as reminder to yourself and/or your audience of what it is that may have affected your sales.

You’ll notice in the example above I also have a “Use?” column (Col. AF). This is going to identify, by formula, those instances where a price change took place and there were no other competing events. A “1” in this column will mean that the price change is to be used for the price elasticity calculation; otherwise, all other months will get a “0”.

**10-8 The next section picks up the price change amount, with separate columns for whether the price change is used or not, and for positive or negative changes.**

The next section picks up the price change amount, but inserts the change in the appropriate column. If the price change is to be used, it gets picked up in Columns AH or AI, with AH picking up price increases while column AI picks up price cuts. Why have a separate column for price increases vs decreases? Because I have found that usually the price elasticity is different for price cuts versus price increases. A 10% price increase might lead to a 6% sales drop, while a 10% price cut might lead to only a 4% sales lift (on average).

Technically, you probably do not need to pick up the price changes that are not being used. I like to have it here as a way to make sure that every price change gets a home, so to speak, even if it’s not being used. It can also be useful to make one more aware of the price changes that are not being utilized.

**10-9 For the price changes being used for the price elasticity calculation, separate columns pick up the estimated impact of the price change.**

Columns AL & AM pick up the impact of the price changes that are going to be used, depositing them in the appropriate column, into column AL for a price increase, or column AM for a price cut. Separating the two ensures separate elasticity calculations are made for price increases & decreases.

**10-10 The final column is simply an identifier of a price change; it will later be used in the trend chart to highlight when price changes occurred.**

The last column here, column AO, is a placeholder that marks when a valid useable price change occurred. As we will later see when we work with the “Trend” tab, this column will highlight when a price change occurred. The large number, “999999” in Cell AO1, is much larger than any other value on the Trend chart, thus making sure the price change is visible on the trend chart.

**10-11 The first price change is an outlier as a competitor was running a major ad campaign; this change will not be used.**

So let’s see how these columns work in practice. We begin with the 3% price cut in March 2015. This point is being treated as an outlier due to the major marketing campaign run by the “ABC” company. Inserting a “1” in the Competitor column ensures that this change will not be used; the “0” in the “Use?” column indicates its exclusion. The price change is not being used, and it’s a price cut, so the -3% amount is dropped into Col. AK. As it is not a valid price change, the value in Col AO is “0”.

Note that the Competitor column also has a “1” in Feb & April 2015. This is because it’s being assumed that even though the campaign was launched in March, a price change in either February or April would also be subject to a distortion as far as getting a read on the pure impact of a price change alone. Indeed, you may want to eliminate additional adjacent months surrounding this competitor event. However, you really have to be careful about this or else you may end up having all your price changes eliminated because you’re so concerned about even a small amount of impact from a concurrent event.

**10-12 The next sample price change does not have any concurrent events, so it can be used for the elasticity calculation.**

We now go to the next price change, a 7% price hike in July 2015. This price change does not have any concurrent events associated with it, so it can be used, as noted by the “1” in the column AF “Use?” column. It’s a price increase and the change is being used, so the 7% gets picked up in Col AH. Earlier, we noted that the sales history has already been trended, following the technique described in prior chapters. Note that Cell AL 185 picks up a “-5.5%” event for the same month as the price change. The formulas in the two “Estimated Impact” columns only have events picked up for those months where a useable price change has occurred. Because we have a valid price change, the last column here picks up the “999999” from Cell AO1 so that the chart is sure to show that a price change took place.

**10-13 The final price change does not have any concurrent events, so it too can be used for the elasticity calculation.**

The third and final price change over this 2+ year period is highlighted here. Again there are no concurrent events so the change can be used. However, note that there is nothing in Col. AL; the first time around evidently the data showed no sign of any kind of impact from this 5% price hike; Cell AL194 is “0%”. Despite this, there is still a “999999” in Cell AO194 because this price change is still considered valid and useable as there is, as yet, no known concurrent event.

**10-14 Price changes are reviewed for concurrent events over the entire period.**

We have shown price changes for 2015-forward. The same procedure applies to the entire 16+ year time frame. The only manual intervention required here is to review whether or not a major concurrent event occurred at the time of the price changes. Of course, memories will struggle recalling the history 5-10+ years back. Ideally you have people in house that do remember major events. And certainly the internal events – marketing campaigns and such - should not only have to rely on people’s memories; hopefully there exists a record somewhere that can be drawn upon.

Might you still miss recollection of a significant concurrent event many years ago? Certainly. But hopefully it will pop up as an outlier when you review your price elasticity data points and you then ask yourselves again whether something might have occurred at the time. Otherwise, you can also put greater weight on more recent events, or simply restrict your analysis to a more recent time frame.

**10-15 We now turn to the “Trend” tab to determine if any events need to be modified, given an awareness that a price change occurred.**

We now return to our original trend estimates, which were developed strictly based on how the data was behaving. We now want to review those original trend estimates, given the knowledge of when past price changes occurred. The chart here displays the original trend estimates.

**10-16 Blue bars are added to the Trend chart to mark when price changes occurred, but not how large they were.**

We now add blue bars to the chart that highlight when price changes occurred. The bars are inserted by a two-step process: first, the Price Change column with the “999999”s is added as another “line” to the chart; then the line is changed to a bar (by right-clicking on the line, selecting “Design” under “Chart Tools”, then “Change Chart Type” and then select the “Column” style of chart). Having used such a high number as “999999” we’re ensured the column bar displays in the chart.

Most importantly, note that while we now can see when price changes took place, we do not know their amount – we’d have to go back to the “Calc” tab to see that. I would urge you to resist that temptation, for you want to make sure your event estimates are not biased by knowledge of the amount of a given price change.

**10-17 Event estimates are reviewed to determine if any merit modification given the knowledge of when the price changes took place.**

I’ve highlighted in the “Events” section the timing of when the price changes occurred. As you can see, their timing corresponds with the insertion of events, or at least they do for the first three price changes. The fourth price change, in April 2016, does not have an event, though there is a small “0.5%” lift added for the following month of May 2016. The estimates that were previously there seem just fine, though I would now modify them if it seemed appropriate. For example, if I did not have a drop of “-1.3%” in Jan 2014 (Cell P31), I would now insert it (or at least some estimate around that -1.3% amount). Of course, I don’t know what the price change amount is, or even if it was a negative price change as opposed to a positive one. We’ll see how it looks when we turn to the price elasticity chart. Once all the history is reviewed, we are ready to do just that.

**10-18 A Price Elasticity chart was created that plots all the price changes against the estimates of how much sales changed.**

Once all the past price change data points have been reviewed, we’re ready to examine the Price Elasticity Estimate chart. This chart plots the percentage price changes against the estimates of sales volume change that occurred. We’ll quickly walk through how it is created.

**10-19 The price elasticity chart starts by comparing price increase amounts against their estimated impact.**

As suggested by the model structure here, we’re going to have two sets of price changes to evaluate: price increases & price decreases. We’ll start off with the more common price increase. Column AH contains the useable price increase amounts. Column AL picks up from the “SalesTrend” tab the impact of the price increases.

**10-20 A “Scatter (XY)” chart is created that plots the price increases alone – their amounts against their estimated impact.**

The first step in creating the price elasticity chart is to “insert” a chart picking up our price increase data points. To do this, the useable price increases (Col AH) and their estimated impact (Col AL) are highlighted, from top to bottom (Rows 10 thru 251) and a “Chart” of the “Scatter” type is inserted, giving you a chart similar to what is displayed here. You want to make sure the chart is showing the price increase amount (Col. AH) as the X-axis and the impact (Col. AL) as the Y-axis.

**10-21 A second set of data is added to the XY chart, picking up the price cuts and their estimated impact.**

The next step is to add a 2nd set of data to the chart. This set picks up the price cuts and their estimated impact, from Columns AI & AM.

**10-22 Next, we add a “Trendline” for each of the two sets of data.**

We next add a “trendline” to the chart. For the price increases, this is accomplished by right-clicking on one of the price increase data points, and selecting “Add trendline”. The same is done for the price cuts.

**10-23 Each “Trendline” is formatted to display their equation, and to set the Y-axis intercept at 0.**

We then have Excel display the equation for each of these trendlines. This is accomplished by right clicking on each trendline, selecting “Format trendline…”, then clicking the squares at the bottom of the “Format Trendline” table. The first square to check is labelled: “Set intercept = ”, with a box in which you make sure the value is “0”. It is likely the box will already be pre-filled with “0” as there are so many data points that are at “0” for both axes – these, of course, are all the occasions where there are no price changes or the price changes are invalid for estimating purposes; that they are all there at 0 is not a problem or an influencer for the chart, especially as the Y-intercept is set at 0”.

You should also check the box for “Display Equation on chart” and for “Display R-squared value on chart”. By clicking on these three boxes we get a display of the trendline equation.

**10-24 The trendline equation tells us how much the y-axis changes for a given change in the x-axis; it also provides a “goodness of fit”.**

What do these trendline equations mean? Let’s use the price increase example, highlighted here in blue. The 1st part of the equation tells us the “slope” of our line. “y=-0.7262x” means that for every 1% change in the x-axis, the y-axis goes down by 0.7262. The x-axis in this chart measures the price change amounts; the y-axis measures the estimated impact of the price change. So this equation tells us that for each 1% increase in price, sales fall by 0.7262%. The 2nd part of the equation is the R-squared value, which is a measure of goodness of fit, of how closely the trendline corresponds with its data points. Here the R-squared value is 0.9445, which is a very good fit. If the data points were more scattered, this value would be much lower. I should point out that the R-squared is made artificially higher by including all the “0” data points. These data points should NOT alter the x value, which is of greatest interest to us; but they will raise the R-squared value.

**10-25 The chart is then “cleaned up” to clarify and highlight the read-out of the price elasticities.**

The chart is then cleaned up so it’s much easier to read. The axes are labelled to clarify the x-axis tracks the price change while the y-axis the impact on sales. The trendline equation is formatted to be easier to read, and notably, the values are rounded to 2 decimal points. Price elasticity measures tend to be quite messy, and it will rarely be the case that more than two decimal points of accuracy is appropriate. In other words, using the price increase measure as an example, it is just not appropriate to say that a 1% increase will result in a 0.7262% drop in sales. There aren’t that many points; there is some variability in the data; it just isn’t that accurate. Here I’ve rounded it to two decimal pints so the read out suggests a 0.73% drop in sales. Arguably it would be even more appropriate to use just one decimal point of accuracy and state there is a 0.7% drop in sales; however, you’ll generally want to be able to see that second decimal point of value so I’d recommend displaying it but then interpreting it to either a one decimal point or at most two decimal point level of accuracy.

**10-26 The chart is examined to identify potential outliers.**

Once the price elasticity estimate chart is prepared, we can examine it to see if there are any apparent outliers. Looking at our chart, for example, it seems evident that the highlighted data point is very different to all the others. Given how consistently the other price changes impact behavior, it seems reasonable & appropriate not to include this data point with our price elasticity estimate. Presumably something else was going on about this time that might explain the apparent lack of response to the price change. Notably, there was no “event” associated with this 5% price increase. Leaving this data point in skews the line, leading to an estimated value that is lower than it “should” be.

**10-27 The final price change does not have any concurrent events, so it too can be used for the elasticity calculation.**

In order to remove the outlier from the calculation, we return to the “Calc” tab and insert a “1” in one of the “Concurrent Events” columns. Here I’ve pretended that staff now recall or discover that a major competitor had a 12% price hike in the same month, thereby negating the expected negative impact on sales that would normally come from a price increase by XYZ. Inserting the “1” leads the data point to be shifted over into the “Don’t Use” category, and the “999999”s in Col AO also go away.

**10-28 With the outlier removed, the slope of the trendline changes, and we have our final price elasticity.**

When we return to the Price Elasticity Estimate chart, we see that the outlier has now gone away, and that the equation has changed. Before the estimate was -0.73, now it has gone up slightly, to -0.75. This is our final price elasticity estimate. It says that, all else being equal, we should expect a drop in sales of 0.75% for each 1% increase in price. One might consider rounding this still to -0.7%, but I’ll leave it like this for now, especially as it happens to “round” to “-3/4”.

**10-29 Choosing outliers can be somewhat subjective, and sometimes a potential outlier might be better left in.**

Choosing outliers is certainly somewhat subjective. When you get down to it though, there is a certain level of subjectivity with all statistical analysis. The equations and formulas and methodology are designed to help limit the subjectivity and get us as close to the “truth” in the numbers as we can. Notice the potential outlier highlighted here. I’ve chosen to leave it alone – for two reasons. 1. While the data point received no event estimate, the “0” value is still not that far from the trendline, so leaving it in doesn’t skew our trendline much at all; (the slope increases from -0.89 to -0.94 when the point is removed). 2. We have very few data points to work with here – just two others – so I was reluctant to remove it. However, this is clearly an instance where further rounding is called for. I would definitely round the price elasticity on price cuts: from “-0.89” to a simple “-0.9”.

**10-30 The price elasticity is a critical metric that every company should know for its products.**

As we said at the outset, the price elasticity is a crucial metric that every company should know about its products. If you don’t know your price elasticity, or are unsure of the figure you’ve been using, this chapter has outlined a way for you to determine it. Knowing your elasticity should lead to better planning and greater confidence in anticipating the impact of a proposed price change. Obviously you’ll need to keep in mind and modify your plans to the degree that you can anticipate other factors intervening, such as a major competitor opting to match your price change.

**10-31 The net revenue impact of a price change is not the simple sum (or product) of the price change and the price elasticity.**

So what is the net revenue impact of a price increase? When a price change goes into effect, your revenues increase (or decrease) with the price increase (or decrease). You can think of the price elasticity as measuring the offset to your price change. While revenues per product go up, your sales fall. However, you have to be careful not to think that it is a simple formula that derives the impact on revenue. People can easily make the mistake that it is simply the price change less the price elasticity. For example, you might think that if your price elasticity is -1.00, and you implement a 10% price increase, your revenue impact will be 0% (or the price increase of 10% less the expected 10% drop in sales).

But as you can see in our simple example, when sales drop by 10% while price increases by 10%, net revenue ends up dropping: the higher price is being applied to less sales and you net less revenue than previously.

**10-32 In fact, the net revenue impact of a price change is a little more nuanced.**

The net revenue impact of a price change is the COMBINATION of the price change and the price elasticity. The price elasticity offsets the price change, but the offset is applied to a higher number, (or lower if the price is reduced). In order to correctly calculate the expected revenue impact of the price change, you need to be sure to apply the right formula. The right formula, shown here, requires adding “1” to both the price change percentage and the price change times the price elasticity, and then subtracting 1. For our example, the expected revenue impact of the proposed price hike falls from 0% to -1.0%.

**10-33 Be careful not to confuse the price elasticity of a product with the price elasticity of your product.**

It may seem obvious to state, but it can easily be forgotten: the price elasticity for your product is not the same as the price elasticity for the product in general. A classic example of this would be gasoline. It’s well known that gasoline is price inelastic, meaning that when its price is raised its demand goes down by far less, so that a 50% national price increase may only result in a 5% reduction in gas purchases. However, if one gas station raised their price 50% while all the other stations held their price constant, obviously the one station would see a huge drop in demand for its product.

This is why it’s so important you separate the price elasticity for your product generally (when you and all your major competitors raise your prices at roughly the same time) from the price elasticity when you alone raise your price (while your competition holds theirs constant).

**10-34 For companies with recurring or subscription billing, there are actually three different price elasticities to measure.**

Before we leave price elasticity, I want to address the special circumstances faced by companies with recurring subscription billing. These would be firms that have customers that typically pay monthly, on an ongoing basis. Examples would be telecom firms, cable TV providers, property & casualty insurers, and so on. For these firms, there are actually three different price elasticities they need to consider when it comes to determining the full impact of a price change.

First, there is new business, which is essentially what we looked at in this chapter. By how much do new business sales drop, given a 10% price hike?

The second metric is attrition, which measures the rate at which customers cancel their coverage. The price elasticity here would measure by how much you’d expect attrition to increase given a 10% price hike.

The third metric is average price, or what I like to refer to as “the price elasticity of price”. This gets at the degree to which your customers reduce the total collection of products they purchase from you, in the wake of a price increase. For insurance companies this manifests itself as customers opting for higher deductibles or reducing coverage as a means for offsetting the impact on their total premium. Thus, in the wake of a 10% price increase, an insurance company might find the average premium only goes up 9%, as customers reduce their coverage slightly to mitigate the impact on their checkbook.

For a company doing recurring billing, you can see that all three pieces need to be determined in order to obtain the net impact on revenue of a proposed price change. Insurance, and other such businesses, can be much more complicated than people might think.

**10-35 For companies with recurring accounts, the net revenue impact of a price change will depend on many factors.**

The table here displays how one can go about calculating the net impact of a price change for a business doing recurring billing. The “Inputs” at the top include the proposed price change, as well as a break out of the different price elasticities. The impact on sales calculates the expected decrease in monthly sales due to the price increase. Applying the -0.8 price elasticity to the 10% price increase results in a net sales impact of -8%. This is applied to the current sales level. For current sales, you don’t want to pick up actual sales, you want to look at seasonally-adjusted sales. And perhaps the best figure for that would be what your current estimated trend level is. The “XYZ Model” file tracks sales (& attrition); I picked up the current sales trend level directly from the file. When this level is applied to the 8% sales impact, we find monthly sales expected to drop by the 2,902 amount in Line 7.

We’ll be looking much more closely at attrition, or churn, in the next chapter. But I bring it in here for the purpose of calculating the true net impact of the price change. Applying the price elasticity of +0.6 to the price change results in an expected 6% increase in attrition. Given a current monthly attrition rate level of 17.6%, which translates to a monthly rate of 1.46%, the net impact of the price increase works out to be the 0.09% in Line 11. That may not seem like much, but when that is applied to the 2 million total accounts, it works out as the 1,754 lost accounts monthly that we see in Line 13.

So, if we have 2,900 fewer sales each month, and 1,700 more accounts leaving, the total impact of the price increase is to drop total accounts by over 4,600 each month, or by almost 56,000 over the course of a year. That 56,000 annual account loss represents 2.8% of the total accounts.

The price elasticity of price for XYZ we had at 0.9, implying that there will be a net 9% increase in the average price. Line 18 then calculates the net impact on revenue, combining the price increase and the offsetting decrease in accounts; the net impact works out at about 6%, or almost 60% of the original proposed 10% price increase (Line 19.).

It’s a bit of work to do this, but again it is well worth it. Don’t you want to be able to calculate, correctly, the net impact of a price change? If you know you all-important price elasticities, something like this set of calculations here will enable you to determine the net impact – a valuable contribution for any firm.

**10-36 High-growth regions will usually find sales as a percentage of total accounts is higher than slow-growth regions, which leads to a smaller net revenue impact.**

One interesting observation to make about the net revenue impact of price changes. You will likely find that high-growth regions will experience a smaller net revenue impact than their older, more established counterparts, all else being equal. The simple example here demonstrates the point. Both states have the same assumptions about price changes and elasticities. But the high-growth state has a higher value for sales compared with total accounts. That higher ratio leads to a lower net revenue impact. Essentially, the price change impacts the total accounts more, thereby increasing the price change offset, and leaving you with a smaller net change in revenue as a percent of the proposed price change.

**Chapter 11: Monthly Retention / Attrition**

This chapter is for any company that has ongoing paying customers, where one of the key metrics is their ability to retain those customers. My focus here will be on monthly retention. Not annual retention, not daily, but the month-to-month tracking of retention. In particular, the approach will focus on a monthly measure of retention that’s updated monthly, not an annual one that’s updated monthly. The distinction is key, for I’ve often encountered annual measures that provide little useful insight from an analytical perspective. By tracking monthly retention, and seasonally-adjusting it, you’ll be able to much more quickly identify and quantify significant positive or negative changes in the metric. All too often, emphasis is applied much more to sales than to retention, but given the enormous level of effort, and cost, required to obtain new customers, more and more businesses are recognizing the value of paying greater attention to retaining their customers. But to do that properly, you’ll need good measures. And that’s what we’ll look at.

**11-2 Our review of retention – and its counterpart, attrition or churn – will modify the traditional measure and address how to trend it.**

The chapter starts with a review and revision of the typical formula used for measuring retention. We’ll also look at how to modify the formula to capture attrition (or churn), the flip side of retention. We’ll further modify the formula to focus on monthly attrition, and explain why the monthly metric is much more timely and informative than the traditional annual metric. We’ll then quickly walk through trending attrition, using the same approach we’ve encountered all along the way. However, the trending will be of the attrition rate, not the amount, as the rate is generally more insightful than amounts which are prone to fluctuate with total customer count. We’ll finish by combining sales & attrition trends to arrive at a clearer picture of net growth.

**11-3 Retention measures the percentage of customers retained over a given period, usually expressed as an annual rate. Attrition measures those who leave.**

Let’s be clear on the definitions being employed here. Retention generally refers to the percentage of existing customers that remain over a given period. Typically the given period is one year, though certainly for many organizations that “typical” period may be a month or less, or perhaps two years or more. We’ll stick with the annual measure for purposes of this discussion. So, if you start the year with a hundred customers, retention would measure the percentage of those 100 that remain one year later. Your retention measure might also include new customers that come in during the year. As we’ll see, this makes it more complicated to track, but we’ll arrive at a useful and accurate measure.

Attrition is the flip side of retention. It measures the percentage of customers that leave. It may also refer to the count of customers that leave. But as the number of leavers will likely increase as customer count increases, a percentage measure is generally more useful and appropriate. Another common term used for “attrition” is “churn”.

**11-4 The most challenging part of measuring retention is the treatment of new customers.**

The tricky part with any retention rate formula is how to deal with new customers who come in during the year. Ideally you separate them out, and indeed some companies separately retention on existing customers versus new. But there are some definite challenges with that approach. For one, the minutiae of tracking all the different types and timings of attrition can often lead to numbers that regrettably “don’t add up”. For another, there is the complexity of providing two different measures of retention, leading to the inevitable request for a third measure that combines existing & new customers. And finally, there is the question of whether the time and cost are worth the effort; are the separate measures truly managed, not just tracked?

**11-5 Perhaps the most common formula used for measuring retention is the following:**

Let’s presume your measure of retention is for new & existing customers combined; what formula do you use to measure retention? I did a Google search on measuring retention, and by far the most common formula I came across was the formula shown here, that the retention rate, for any given period, is obtained by subtracting new customers from the count of ending customers, and dividing that by the starting customer count. In the equation, the starting point is 1 year prior to the end point, and new customers are the count of sales during those intervening 12 months.

**11-6 To demonstrate this first traditional retention formula, we’ll follow a simple example, where we start and end with 100 customers, and bring in 20 new.**

We’ll use a simple example to demonstrate this formula, a formula I’ll call “Formula #1” (as we’ll be looking at two others shortly). In this example, we start the year with 100 customers, and see 20 new customers come in during the year. And we end with 100 customers. If we subtract out the new customers, we see that we’ve retained 80 customers, giving us a retention rate of 80 over 100, or 80%.

**11-7 Another formula that some firms include new customers as part of the starting base.**

A second formula that I’ve seen used has one important variation. Instead of subtracting the new customers from the ending account total, it includes the new customers with the starting. You end with 100 customers; you started with 100 customers and added 20 along the way. Of the 120 total customers, you’ve “retained” 100 of them, for a retention rate of 83.3%. One of the nice attributes of this approach is you end up with a higher figure. And this approach, like #1, is easy to calculate. But which of these, if either, is more appropriate?

**11-8 The central challenge with any retention rate formula is how to treat new customers, or more precisely, how to treat when they are new.**

The central problem with any retention rate formula that looks at all customers is how to treat the new customers. We saw with the two retention rate formulas that they handled new customers quite differently. So ask yourself, what is retention, given that we are looking at a world where existing and new customers are intermingled? I will submit to you that the key to answering this question is to know WHEN the new customers come on board. It’s time to put on my economist’s hat.

**11-9 Economists love to make assumptions. Let’s assume that ALL new customers come in on the 2nd day of the year.**

As you probably know, economists love to make assumptions. They do that to help illuminate a problem. Allow me to do that here with our problem concerning how to handle new customers when it comes to measuring customer retention. Let’s start by assuming that it’s the end of 2017, and we’re looking back and measuring the retention rate for 2017, and let’s further assume that all the new customers came in on the 2nd day of the year.

**11-10 Given that we essentially start the year with the new customers, as well as the existing, then arguably we can include the new in our baseline.**

If all the customers came in on the 2nd day of the year, then essentially, by the end of the year they have had virtually an entire year to leave, while the existing customers had the full year to leave. So we can treat the new customers as though they were existing; after all, they have had virtually the same amount of time in which to leave as the existing customers did. So the formula for measuring retention, for 2017, is to take our ending customers and divide them by the starting and new. We’ll call this the “NewYear” formula, since all the new customers arrive at the new year.

**11-11 Let’s next assume that all the new customers come in on the last day of the year.**

Now let’s assume that all the new customers arrive on the 2nd to last day of the year. Are we going to want to include them in our measurement of retention? Absolutely not; they haven’t had any time to leave yet. So here we’ll want to subtract our new customers from the count of ending customers, while our baseline will be our starting customers. We’ll call this the “YearEnd” formula, since all the new customers arrive at year end.

**11-12 Our “YearEnd” and “NewYear” formulas are identical to the “Formula #1” and “Formula #2” that are used by almost everyone.**

You may have noticed as we were walking through our “unreasonable” assumptions that the formulas being used were identical to the formulas used traditionally. The primary formula, “Formula #1”, is identical to the YearEnd formula. It treats all new customers as though they came in at the very end of the year. Our second formula treats all new customers as though they come in at the start of the year. Clearly, neither of these is a reasonable assumption. So what should it be?

**11-13 We want the “correct” formula to treat new customers as though they come on board around mid-year: *on average*.**

So what does the correct formula look like? Well, we’re going to want a formula that treats new customers as though they arrive mid-year. Obviously they won’t all do that at the same time. But, *on average*, they will. Some will arrive at the start of the year, some will arrive year end, others will arrive at all manner of days in between; on average, they’ll arrive around midyear. So we want a formula for midyear arrival of new customers.

**11-14 A MidYear retention rate formula is a “blend”, an average, of the two traditional formulas.**

The “MidYear” formula is quite easy to construct, if we just re-state our traditional formulas a little differently. The “NewYear” formula has been revised to show that there are zero new customers being subtracted from the ending customers in the numerator. And the denominator re-states the new customers as “1 New”, meaning that it picks up all 100% of the new customers. Similarly, the “YearEnd” formula shows all new customers subtracted from the ending in the numerator, and that there are zero new customers added to the denominator. Well, midyear should be the average of these two formulas, right? In the numerator, we’ll subtract the average of 0 and 1, or ½, of the new customers. And in the denominator, we’ll add ½ of the new customers to the beginning customers.

And voila, we have a retention rate formula that remarkably well reflects the “true” retention rate. Our example was for all of 2017, but it obviously can be applied anytime during the year. Though new customers will seldom exactly come in at midyear, they will probably usually come in very close to that.

**11-15 Applying the MidYear formula to our simple example, we arrive at a retention rate that’s between the 80% and 83.3% rates we previously obtained.**

We can now try out our new formula on the simple example we used previously. We subtract half the new customers, or 10, from the numerator, and add half of the new customers to the denominator. And we arrive at a retention rate of 81.8%, roughly halfway between the 80% we got using the “YearEnd” formula and 83.3% obtained using the “NewYear” formula.

**11-16 Attrition is simply the flip side of retention. A bit of algebra gets us to the appropriate “MidYear Formula” for calculating the attrition rate.**

Now that we have the “right” formula for retention, we can shift over to its counterpart, attrition. The attrition rate is simply 1 minus the retention rate. Here we walk through the math, picking up the midyear formula for the retention rate. A small use of algebra gets us to restate the value “1” with the numerator matching the denominator, and the denominator matching the denominator for the retention rate formula. With the denominators the same, we can string together the numerators so we have “Start + ½ New” minus “End – ½ New”. When we sum the two numerators we have “Start – End” and ½ New minus a minus ½ New, which gives us ½ New + ½ New which is equal to one New.

Our final formula makes implicit sense. The numerator is simply “Start – End + New” which is precisely how one goes about calculating the attrition for any period. And the numerator picks up our starting point and adds half the new business to it, which again is precisely correct for some of the “New” will be almost a year old and some will be almost brand new and on average the new will be a half-year old so we would want to count half of the total amount.

**11-17 Walking through our simple example of how the formula works, we arrive at an attrition rate that is equivalent to 1 minus the Retention Rate.**

Finally, let’s walk through our simple example of how the formula works. We apply the “MidYear” attrition formula to our assumptions on starting and ending and new customer counts, and we get exactly what we would expect, the counterpart to the retention rate. The retention rate example worked out as 90/110; the attrition example works out as 20/110. It works, it makes sense, it’s simple, and it’s a more accurate & appropriate formula for calculating the “true” attrition rate.

**11-18 The Retention Rate and Attrition Rate formulas can apply to any period of time.**

The retention & attrition rate formulas we’ve developed can work for any period of time. The “Start” can refer to where you were one year ago, or one month ago, or even yesterday. The “New”, in turn, can refer to new business over the past year, past month, or past day.

**11-19 Many organizations, though by no means all, generally track retention (& attrition) as an annual rate.**

Different companies and industries focus on different periods of time when they examine attrition (or retention), much depending on the level of attrition that occurs. If you’re losing 5-10% or more of your customer base each month, you’re probably looking at it in monthly terms, or even daily. But most firms fortunately do not suffer such inordinate rates, and as a consequence, they typically track & express attrition as an annual rate. I’m going to choose to examine attrition as an annual rate. While annual formulas are a little different than monthly, it’s not much effort to make the translation between the two. Just to be clear, we’ll be looking at attrition on a monthly basis, but it will be the measure of the annual attrition rate that we’ll be tracking on a monthly basis.

**11-20 For analytical purposes, in order to be able to trend attrition (or retention), and fully capture major shifts when they occur, we’ll want to track it monthly.**

We now turn our attention to tracking attrition on a monthly basis. While it may be the case that annual retention rates are reported to senior management and the outside world, when it comes to analyzing attrition, the focus should be on monthly activity. By focusing on monthly rates we can capture when significant shifts occur. Just as with sales, most changes in attrition rates over time are not generally due to slow or steady growth or decline, but are instead the result of step functions. In order to observe and measure these critical step functions, we will want to track attrition as a monthly rate. We’ll want to convert the monthly rate to an annual equivalent so that the measure has better meaning & context. And we’ll want to seasonally-adjust the rates over time. Seasonally-adjusting the rates will also include making adjustments for the calendar – or “normalizing” the rate.

While attrition is the natural counterpart to retention, the discussion here will focus on attrition. Why? For one, attrition rates are easier to comprehend on a monthly basis. An attrition rate of 1% or 2% for a given month is probably more meaningful than a retention for a given month of 99% or 98%. Also, as we’ll later see, it can be quite compelling to match seasonally-adjusted sales against seasonally-adjusted attrition – the comparison will help flush out your “true” rate of customer growth at any given point in time.

**11-21 Out with the old…**

Perhaps the single biggest mistake many companies make is in their choice of metric for tracking retention, or attrition. So many still insist on using an annual measure, one that tracks the volume of customers leaving over the prior 12 months. Regardless of the exact formula being used, this time period has a number of serious shortcomings: it is virtually incapable of signaling when a sudden shift in performance has occurred; it can send the wrong signal as to how retention / attrition is performing at any given point in time; it can be unclear whether a change from one month to the next is driven by the activity of the prior month or from a shift that occurred 12-13 months before that “drops” from inclusion in the rate. To be sure, for some purposes, such as an annual report or industry overview, the annual metric is fine. But for analytical purposes, it is time to leave this old measure behind.

Instead, the way to track and analyze retention & attrition performance is to use a monthly measure that tracks the level of leaving customers for each prior month. Ideally, of course, this monthly figure should be seasonally-adjusted, enabling you to be much clearer about how you’re doing now, and to be able to much more quickly identify & quantify when a sudden shift takes place.

**11-22 …in with the new…**

We now turn our attention to tracking attrition on a monthly basis. While it may be the case that annual retention rates are reported to senior management and the outside world, when it comes to analyzing attrition, the focus should be on monthly activity. By focusing on monthly rates we can capture when significant shifts occur. Just as with sales, most changes in attrition rates over time are not generally due to slow or steady growth or decline, but are instead the result of step functions. In order to observe and measure these critical step functions, we will want to track attrition as a monthly rate. We’ll want to convert the monthly rate to an annual equivalent so that the measure has better meaning & context. And we’ll want to seasonally-adjust the rates over time. Seasonally-adjusting the rates will also include making adjustments for the calendar – or “normalizing” the rate.

While attrition is the natural counterpart to retention, the discussion here will focus on attrition. Why? For one, attrition rates are easier to comprehend on a monthly basis. An attrition rate of 1% or 2% for a given month is probably more meaningful than a retention for a given month of 99% or 98%. Also, as we’ll later see, it can be quite compelling to match seasonally-adjusted sales against seasonally-adjusted attrition – the comparison will help flush out your “true” rate of customer growth at any given point in time.

**11-23 To express monthly attrition as an annualized rate, we multiply each month’s attrition by 12, and add 6 months of new business to the denominator.**

Now let’s quickly walk through the formulas. On the left side we have the attrition formula we developed earlier. “Start – End + New” equals the “attrition”, in this instance for an entire year. The denominator for the annual formula is the start plus half of the new business for the year, (which is approximately equal to 6 months of new business).

Now we look at the monthly attrition rate, expressed in annualized form. The numerator multiplies the attrition for the month by 12, essentially giving us the approximate annual level of attrition. In the denominator, the “Start” amount is the customer count at the end of the prior month; added to that is 6 times the New Business for the month. Why do we multiply by 6? We want to capture the equivalent of a half-year of new business; multiplying that month’s attrition by 6 does that. If you compare these two formulas, you can see they are actually very similar. The key difference is simply the period of time that’s being referenced. The formula on the left looks at a whole year of activity: 12 months of new business, and using the level from 12 months before as the “Start”. The formula on the right uses the level at the end of the prior month as the “Start”, and the “New” refers to the current month’s new business alone.

**11-24 As reference, here are the formulas for calculating monthly and annual retention & attrition rates.**

Here, for reference, are all the formulas for calculating both retention and attrition, as an annual or monthly annualized rate. Note that either for the annual rate, or the monthly, the denominators are identical. Which makes sense as the only way you’re going to be able to get “1 - attrition/retention” to work right formulaically is if you make sure you are using the exact same denominator.

Let’s now move on to seeing how these work out when applied to a hypothetical example. For this, we return to the example we previously developed.

**11-25 With agreement on the appropriate formulas, we’re ready to walk through how to trend the attrition rate.**

Well, we’ve now sorted out the appropriate formulas for measuring monthly attrition, on an annualized basis. We’re ready to walk through an example of how the attrition rate can be trended over time. To do this, we’ll pick up use of the XYZ hypothetical data earlier employed to demonstrate estimating price elasticity. Here we can see the entire calculation and trending of attrition rates laid out. The calculations are again done in the “Calc” tab of the “XYZ Model” file. Let’s walk through this calculation process step-by-step.

**11-26 Monthly attrition is calculated using the Actual values for New Business and Total Customers.**

To start, we need to calculate attrition. Or rather, I will choose to calculate attrition. It could be that you have attrition or churn values already reported. And perhaps such data is reliably “clean”. It’s just that I’ve often encountered that it’s easy to get lost in the details of what exactly gets represented as attrition. Do you count lost renewals, unpaid new business, reversals, reversals of reversals, etc.? It can get pretty messy. I’m going to choose to keep it simple instead, and just calculate attrition, using the data on new business and customer count.

Here we see displayed the “Actual and Forecast Values” for Sales, in Col S; this data was earlier referenced with the price elasticity calculation. Column BB brings in the “Customer” count; this data would typically be imported into the “Input” tab, and then brought in here for use in calculating attrition. If we know sales, and the beginning and ending customer count, we can calculate attrition: it’s equal to the starting count minus the ending count plus new business. So for Feb 2001, we started with 1,800,000 customers, ended with 1,796,000 odd customers, and added 24,343 new customers: 1.8 million – 1.796 million plus 24,000 gets us at just under 28,000, or the “27,927” figure we find in Cell BC 12.

**11-27 The monthly annualized attrition rate is calculated using the calculated attrition and customer counts.**

Next we calculate the attrition rate, expressing it in a monthly annualized form. Using the formula developed earlier, the attrition rates in Col BD are obtained by multiplying the attrition in Col. BC by 12, and dividing it by the sum of the prior month’s customer count plus the current month’s sales multiplied by 6, to arrive at the value in Cell BD8. So for Feb 2001, the attrition rate equals: (12 x 27,927) / (1.8 million + 6 x 24,343) = 335,124 / 1,946,058 = 17.2%.

**11-28 The attrition rate formula can be easily modified for those firms choosing to still apply traditional NewYear or YearEnd formulas for calculating retention.**

While I would not recommend it, you may choose to continue to apply a traditional formula for calculating retention, and attrition. Perhaps you’d like to change, but management is reluctant to use a formula they don’t understand, or that isn’t consistent with history. After all, your retention rate may go up or down a percentage point or two by modifying it.

If that’s the case, you can easily adapt this model to calculate the attrition rate to be consistent with your formula usage. All you have to do is to modify the value in the highlighted “NewBusMonths” cell (Cell BD8), based on the formula you use for calculating retention. If you are using the common YearEnd formula, i.e., End – New / Start, you would put a “0” in the New Business Months cell. And you’d place a “12” there if you employ the NewYear formula for calculating retention. These different values will modify the number of new business months applied in the denominator of the attrition rate formula. Note that in the attrition rate formula, the numerator is always the same: Start - End + New, which is equal to the “Calculated Attrition”.

**11-29 Normalization factors are brought in for normalizing attrition, adjusting the data so every month is of approximate equal length.**

When we trended sales, we went about seasonally-adjusting the data. Part of the process of seasonally-adjusting the data was to first normalize the data, to adjust it so every month is of approximately equal length. We now need to do the same with the attrition rate. Because it is a rate, we will need to normalize both parts, the attrition in the numerator, and the sales in the denominator. We’ve already normalized the sales, so we only need to normalize attrition. To do that, we need to bring in a set of normalization factors, such as those shown here in Col. BE. Where do these come from?

**11-30 If daily attrition (or churn) data is available, follow the same procedure as with sales for developing normalization factors.**

In order to develop normalization factors for attrition, you will need to follow the same procedure described in Chapters 1, 2, & 3. You need to develop factors for the day of the week (the Equated Day Factors), and holiday factors for fixed-day and fixed-date holidays.

**11-31 The key question with normalizing attrition is: do you have the data needed to develop these factors?**

develop holiday factors, you’ll want several years’ worth for fixed-day holidays, and for the fixed-date holidays you’ll want 2-3 decades worth if you can get your hands on it.

My guess is that will rarely, if ever, be the case. Particularly for attrition data that is usually not as carefully tracked as sales; and that can be subject to much more nuance and complexity because of all the different ways accounts can flow.

**11-32 If your data is limited, do the best you can with what you’ve got.**

At the risk of sounding trite, you’ve got to do the best you can with what you’ve got. Perhaps you can develop EDFs for attrition, and then can extrapolate to the holidays using the EDF factors for attrition, as compared with sales, as a guide. If the attrition EDFs vary less for attrition than for sales, have the attrition holiday factors vary less as well.

Generally, I would leave the holiday factors as they are in sales, and modify them to the extent you have data that allows. I’m not going to try to go into the minutiae of how you go about that for there are just too many different kinds of potential types of patterns in the data that you may encounter that it renders discussion almost meaningless. I can’t address them all, and if I tried, I’d probably lose you, my audience. So I shall simply counsel that you do your best, that you keep it simple, and above all, be comfortable that what you use makes sense.

I confess that when I created all the data for this hypothetical example, I never attempted to develop daily attrition data subject to different equated day and holiday factors. So I’ve chosen to take the cheap way out, and have simply picked up the normalization factors developed for sales and applied them here to attrition. It may often be the case that they are similar, though in practice you will no doubt find significant differences as well.

**11-33 The 1st step in normalizing the attrition *rate* is to normalize the monthly attrition amount.**

So let’s get back to the process for normalizing the attrition rate. We start with normalizing the attrition rate’s numerator: the attrition amount. As displayed here, to do that we simply divide the calculated attrition, from Col. BC, by the normalization factor in Col BE. For Feb 2001, a short month, the attrition of 27,927 is increased to 30,252 (Cell BF12).

**11-34 We then bring in normalized sales, which was earlier developed when sales were trended.**

Next we bring in normalized sales, which had previously been calculated when we walked through trending sales. As you can see, all the values in Col. BG are picked up from Col. F. (Of course, you could dispense with this restatement; I’ve inserted it here so its clearer to see what goes into the determination of the normalized attrition rate.)

**11-35 With the necessary components in place, the normalized attrition rate can be calculated.**

We now have the necessary components in place to be able to calculate the normalized attrition rate. Recall that the formula is 12 times attrition divided by the sum of the starting accounts plus 6 times the current month’s sales. So for the Feb 2001 example, it will be equal to 12 times the normalized attrition of 30,252, divided by the 1.8 million starting accounts plus 6 times the normalized sales of 26,370, which gives us a normalized attrition rate of 18.5%. What does this mean? It’s telling us that we’ve adjusted the attrition rate to account for the length of the month. For February, attrition was dampened somewhat simply because it’s a short month – normalizing the attrition brought the rate up, to a rate more similar to those in the surrounding months, as it happens.

**11-36 Next, the attrition rates are seasonally-adjusted using seasonal factors that will be developed in the “Trend” tab for the Attrition Rate.**

The rest of the procedure for trending the attrition rate is virtually identical to that which we followed when we trended sales. First we estimate seasonal factors, by trending the data in a “AttrRateTrend” tab set up for the attrition rate. We’ll be getting to that as soon as we’re done walking through this portion of the “Calc” tab. As you can see, there are three different sets of seasonal factors to choose from. You may well want to start off with no seasonality, as I did here with the “Orig” set of factors being all set at 1.00. I then walked through one set of developing the seasonal factors in the attrition rate Trend tab; that got me to the “Revised” set displayed here in Col. BJ. I brought those rates into the “Inputs” tab and changed them all to values so they wouldn’t keep getting changed as I ran through another round of trending the attrition rate data; that results in the “Final” set of seasonal factors shown here in Col. BK. The choice of which set of seasonal factors to employ is made over in the attrition rate trend tab; that choice is brought in here, and determines which set of factors is dropped into the “In Use” column, Col. BL. The seasonally-adjusted monthly attrition rate is simply the normalized attrition rate divided by the seasonal factor in use. For Feb 2001, that means dividing the normalized rate of 18.5% by the 1.06 seasonal factor, to arrive at a seasonally-adjusted rate of 17.5%, in Cell BM12. The monthly rates are then smoothed by calculating a running 3-month average in column BN.

**11-37 The “Estimated Trend” for the attrition rate is derived by applying estimates for growth & events to a running trend percentage.**

The next section is again handled exactly as it was for sales. We begin with an estimated level for attrition, a level that’s set when we look at the chart in the attrition rate trend tab. From that, we then adjust the estimated trend for the monthly estimates of growth rate & events. Again, the growth rate and event estimates are developed in the trend tab; this is where those values are picked up and applied to a running account of the attrition rate. You can see here that until March 2002, the trended rate stays steady at around 17.5%. But in March, a -3.5% event is applied, which brings the rate down to 16.9% (in Cell BQ25).

The estimated trend tracks how the rate behaves over time, with the noise of seasonality and the calendar effect removed. The “Estimated Trend Values” calculated in column BR put that noise back in. So for March 2002, the estimated trend of 16.9% is multiplied by the seasonal factor for March, of 1.06 (from Col. BL), and a normalization factor of 1.04 (from Col. BE), to arrive at the value of 18.6% in Cell BR25. Note that the normalization factor used is the set developed for attrition. While the normalized sales were *also* used when we calculated the normalized attrition rate, it makes such a small difference that it is quite safe here to simply pick up the attrition’s normalization factors alone. We’ll re-visit these estimated trend values when we later compare the estimated trend with the actuals.

**11-38 The “Forecasts” section picks up from where the estimated trend left off, and re-applies calendar effect & seasonality to derive extrapolated trend values.**

The next section is also handled exactly as it was for sales. The “forecast” is treated as an extrapolation of trend. In other words, you are forecasting to continue at the same pace & place as you left off at. Your forecast entails modifying that extrapolation to capture any projected change in growth rate or upcoming “events”. In our example here, you can see the extrapolated trend in April 2017 (17.7%, in Cell BS206) is the same as the estimated trend in March 2017 (Cell BQ205). They’re identical because the growth rate is at 0% (in Cell BO206). It’s not until October 2017 that the trend changes, due to a projected 2% lift occurring that month.

Meanwhile, the extrapolated trend values take the extrapolated trend and multiply it by the normalization and seasonal factors.

**11-39 The “Other” section picks up past & present trend and actuals, and compares the two. A 12-month annual attrition rate is also calculated here.**

The final section in the “Calc” tab for attrition is “Other”. The first two columns (Col. BU & BV) pick up the estimated & forecast trend, and the actual & forecast values. This is done so the chart, in the “AttrRateTrend” tab, can show trend and actuals as one line each instead of two lines each. Col. BW compares our estimate of actual monthly values with the actuals. Thus, for March 2017, the actual rate was 19.5% (see Cell BD205 or Cell BV205), while the estimated trend value was 19.7% (Cell BR205); the 19.7% estimate is 0.2% higher (Cell BW205).

Finally, column BX205 performs a calculation of the 12-month annual attrition rate. Everything else here is for the 1-month attrition rate, expressed as an annualized rate. Here we show what the rate would be using the more traditional 12-month formula. When we walk through the chart on the “AttrRateTrend” tab, we’ll see how much less useful this traditional approach to measuring attrition is by comparing our monthly attrition with this 12-month annual rate.

**11-40 The “AttrRateTrend” tab operates the same as the “SalesTrend” tab, with manual estimates of Growth Rate & Events inserted.**

We now turn to the “Trend” tab for the attrition rate, which is called here “AttrRateTrend”; (I keep trend tab names shorter so I am able to see all the tabs across the bottom of my computer screen). While the units are different – percentages as opposed to numerical values, everything else on this tab operates the same as it did for Sales. At the top are the tables for inserting your estimates of trend, as determined by the “Growth Rates” & the “Events”. Actuals vs Estimates are then compared, with an effort made to ensure their differences – in Row 53 - are small. Here you can see the difference never exceeded 0.5% in any one year (Cell L53; the difference in 2010).

While not shown here, if you look at the original file (“XYZ Model”), you’ll see that the estimation of the seasonal factors is performed the same as with Sales; the calculations are again done below the area shown here. The output of the Seasonal Factors calculation can be seen with the small “Seasonal Factors” chart in the lower right-hand corner of the figure displayed here. You can see there is a distinct pattern of a quieter level of activity during the summer months.

**11-41 For the XYZ example, the Attrition Rate chart shows several significant shifts, thanks to its focus on MONTHLY attrition.**

When we zero in on the Attrition Rate chart, we can see that the rate experiences significant gyrations over the 2014 to early 2017 period displayed here. The attrition rate changes by more than half a percentage point no fewer than 4 times (May & Sep 2014, Mar 2015, & July 2016). These sudden shifts are made apparent because we’re looking at MONTHLY attrition, how much the rate changes due to the number leaving each individual month.

**11-42 The “Annual Attrition Rate” is added to the chart, showing the 12-month attrition rate over time.**

You’ll recall we calculated the 12-month annual attrition rate, in Col. BX of the “Calc” tab. This was the rate that captured the approximate percentage of customers that left over the prior 12 months. That line has been added here.

**11-43 The “Estimated Trend” line is removed to help highlight how the ANNUAL rate measure compares with the MONTHLY rate.**

Next I have removed the Estimated Trend line from the chart. I want to make it easier to observe what goes on with the MONTHLY annualized attrition rate, as compared with the ANNUAL attrition rate. In particular, I want to walk through the various ways that the traditional annual measure of attrition is inferior to the monthly rate – a monthly rate that is annualized and seasonally-adjusted.

**11-44 The MONTHLY rate responds far more quickly when a significant shift occurs in the rate.**

The 1st deficiency with the ANNUAL rate is that it responds much more slowly when significant shifts occur. Or phrased differently, if you don’t track attrition on a monthly & seasonally-adjusted basis, you will be unable to see when substantial shifts have occurred. Highlighted in the chart here is a jump in attrition from around 17.5% in Feb 2015, to almost 19.0% in March. Note too that the rate was around 17.7% for the 6 months prior to the jump, & remains above 18.5% for the 10 months following. In one month, the attrition jumps almost a full percentage point, remaining above 18.5% through the end of 2015. But the ANNUAL rate doesn’t reach 18.5% until Dec 2015. The rate of increase in the ANNUAL rate does bump up in the March-May time frame, but not nearly as much as the MONTHLY.

**11-45 Because the Annual rate captures attrition over the entire year, it can give a totally false picture directionally.**

Another issue with the annual attrition rate is that it can give you an utterly false picture of how you’re doing NOW. An example of this is highlighted in this chart. As the end of 2015 approaches, the monthly attrition shows improvement, falling from around 18.7% at the end of 2015 to about 18.3% during the first few months of 2016. But the annual rate continues to climb during this period, a climb that is primarily capturing the huge increase in attrition that took place at the start of 2015, nearly a full year before. The annual rate took about a year to catch up to the 1-month rate, which is what you would expect, because essentially the annual rate is roughly an average of the prior 12 monthly rates.

I have to ask, what good is a measure that is providing you the wrong message on how you’re doing? Imagine you work at this XYZ firm, and at the start of 2016 you implement some major program to bring down attrition. The monthly rate shows you had some success, for the rate drops almost 0.3% and stays low the next 4 months, before dropping even more. But the Annual rate keeps climbing in Jan 2016, as well as Feb & Mar. The Annual rate doesn’t start dropping until May. And real improvement in the Annual rate doesn’t start being reflected in the Annual rate until the *summer* of 2016. Are you kidding me? Again, how useful is that? And no, this is not some drummed up example that would never happen to you. This is a problem that comes from using a metric that is incapable of giving clear, or consistently correct info on how you’re doing, NOW.

**11-46 Because the Annual rate captures attrition over the entire year, it’s direction can be governed as much by what occurred 13 months ago, as by last month.**

Another problem with the Annual rate is that when a marked change in the rate occurs, you can’t be sure whether it’s due to what happened last month as opposed to what occurred 13 months ago. It’s an annual rate, with all prior 12 months effectively having equal weight. So when a given month “drops off”, because it was 13 months ago and is now no longer part of the running Annual rate, the impact of its “removal” will be captured in the current month’s annual rate. In the example highlighted here, the 12-month rate starts to climb in the summer of 2015, a climb that is partly due to dropping off the high values from the prior summer: in the summer of 2014, the rate drops from about 18.4% down to the neighborhood of 17.5%; one year later that drop manifests itself by bumping up the 12-month rate.

Again, how useful is that? How would you know how much of the change in this month’s reported annual rate was due to what occurred 12-13 months ago, rather than from what occurred last month to this?

**11-47 To summarize, there are a number of key reasons that a seasonally-adjusted Monthly attrition rate is a much more valuable metric than an Annual figure.**

Hopefully you are now convinced a monthly attrition rate is much more useful measure than an annual 12-month figure. If you’re already using a monthly figure, forgive me for preaching to the choir. Though I would be quick to point out that a monthly figure is most valuable when it is seasonally-adjusted; if it isn’t, you’ll still have the challenge of obtaining an accurate read of how it’s performing now.

A review then of the benefits of employing a seasonally-adjusted monthly attrition rate for tracking attrition (or retention).

1. You’ll quickly see when a sudden shift occurs.
2. You obtain a much more accurate read of how you are doing now, as opposed to a read of how you’ve done on average over the last 12 months. Perhaps even more importantly, it will help ensure you don’t get false reads of how you’re doing, be it that you’re actually doing worse than you think you are, or better.
3. You won’t be fooled by shifts in attrition that may have occurred 12-13 months ago.
4. You’ll have a very powerful tool for monitoring any activity you undertake that is designed (or not) to impact attrition.

**11-48 By clarifying month-to-month behavior for Attrition, as well as Sales, we help clarify Net Growth.**

Before leaving attrition, I would like to demonstrate how seasonally-adjusting your attrition, as well as your sales, can lead to greater clarification of the net growth of your customer base. Net growth, of course, is simply sales less attrition. Once both those elements have been seasonally-adjusted, you’ll be better able to see whether Net growth is positive or negative in a given month, after adjusting for seasonality, and you’ll be able to quickly see which of those two elements is more responsible for net growth’s direction.

**11-49 For attrition, we can develop a trending of the attrition amount, in addition to the attrition rate already performed.**

So far, we have walked through estimating trend for the attrition rate. I focused on that because generally the behavior of the rate is more informative for assessing the performance of attrition. Here I’ve run trend on the attrition amount itself, switching to the amount so we can more easily compare it with sales. Not surprisingly, while the pattern is very similar to the attrition rate, it will not be identical.

**11-50 Displaying seasonally-adjusted sales & attrition on the same chart quickly shows when one metric or the other is higher.**

In order to get a clearer picture of how the firm is doing in terms of net growth, we place both sales & churn lines on the same chart. Of course, the key here is that we’re displaying seasonally-adjusted sales & churn, in order to have a truer picture of how they are each performing.

**11-51 Charting techniques can provide a nice visualization of how sales & attrition compare.**

In this chart, I’ve used the same data as on the prior page, only I’ve displayed it in a way to highlight when net growth is positive (green) or negative (red). It’s a little tricky setting this up. I relied on some helpful guidance from the Chandoo.org website, which is an excellent place for ideas on what you can do in Excel. If you study the XYZ Model Excel file, you should be able to ferret out how this chart was put together.

**11-52 A Total Accounts line can be added to the chart, using a 2nd Y-axis.**

A comparison of sales & attrition is made as the difference speaks to net growth in accounts. Here we have brought in the Total Accounts as a dark pink, dashed line. Note that because the numbers are so much higher, the values for the line are read off the 2nd y-axis on the right. And note that this 2nd axis is NOT proportional to the primary y-axis on the left. Ordinarily, a 2nd y-axis should always be proportional to the 1st, but that isn’t really appropriate here because the total accounts line changes with the difference between sales & attrition, not with the change in either sales or attrition by themselves.

**11-53 Notice that often the accounts total and the sales vs attrition gap can run in opposite directions.**

What’s interesting to note is how the behavior of the total accounts does not correspond with the gap between sales & attrition. Highlighted here is just one example among many of where sales exceed attrition yet the total accounts are in decline. By now, it should be quickly apparent how this is the case: the sales & attrition lines are seasonally-adjusted, while the total accounts line is not. If we were looking at unadjusted sales & attrition, we’d find attrition on top. But obviously, this seasonally-adjusted picture is a more accurate reflection of “true” performance.

**11-54 Seasonality explains the apparent aberration: near year-end, sales are typically lower & attrition higher.**

While total accounts may be declining during the highlighted period, if you look closely you can see that they are declining at a slower pace than typical for this time of year, near yearend. XYZ should be pleased by this slowdown, brought about by strong sales growth & a marked decline in attrition toward the end of 2014, while the performance in both reversed as 2015 ended, resulting in a steeper decline in net accounts.

**11-55 2017’s flat growth in sales & churn helps clarify the net account seasonal pattern: growth early in the year, decline near yearend.**

This chart highlights 2017, when both sales & churn are flat, on a seasonally-adjusted basis. Of course, they’re flat here because we’ve assumed no expectation of growth in either for the entire year. As you can see, during the 1st half of the year, net accounts grow well. Meanwhile, growth is flat the 2nd half of the year, even though sales exceeds attrition enormously on a seasonally-adjusted basis. This is yet another example of the value of presenting key measures on a seasonally-adjusted basis. Of course, if net accounts were also presented on a seasonally-adjusted basis, we would expect to see them growing steadily throughout 2017.

**11-56 If you prefer, you could display the comparison of trend between sales & attrition.**

There are lots of choices of what you can display, (and how you display it), when comparing sales with attrition. Here the trend lines for them are shown. The advantage here is that it focuses on approximately when major shifts occur in how each metric behaves.

**Chapter 12: Forecasting**

We finally turn to the future: forecasting. This chapter will end up being a bit shorter than the others, and perhaps shorter than you might have expected for a topic of this importance. But that’s because, to a large degree, the heavy lifting has already been done.

**12-2 The key to successful forecasting can be summed up in eight words:**

It’s no exaggeration to state that “your forecasts are as good as your history”. What I mean by that is that so much of your forecasting is dependent on how well you’ve identified where you are “truly” at today, and how much you’ve been able to learn from your past. There is so much critical information to be gathered from your history – insights into what level of impact you might expect from a proposed new product launch, marketing campaign, process change, or whatever. Comparing your proposals with what you’ve done before and the measures you’ve been able to make of your past performance will go far in informing what impact to expect from any projected activity you’re contemplating. Obviously, we don’t know what will happen in future, and history never does precisely repeat itself, but the more knowledge you’ve gained from your past, and the better your defining how you’re doing today, the more likely you’ll succeed at putting together a solid & accurate forecast. And, perhaps more importantly, the better the understanding of your past, present, & future performance, the better position you’ll be in to explain why future results do inevitably deviate from your projections.

**12-3 The forecasting process relies heavily on what has already been learned from trending the history.**

The steps for the forecasting process are quite simple:

1st, you want to re-examine your trendline with a particular focus on ensuring you’re comfortable with the exact current level you are at, and with the growth rate that is currently in use. After making any modifications to the recent past, you simply extrapolate your current trend, a step that is already automatically performed by the trend model. Next, you modify this projection. You start by examining future growth rates, a step that primarily considers external factors. Then you add in projected events, which will usually be driven by internal activity. Finally, you review the forecast for reasonableness, and consideration of any further tweaking.

**12-4 Determining your current level and growth requires close review of your current trendline for potential modification.**

To start the review of your current position, we return to the “Trend” tab. And here we will pick up on our earlier example of trending sales for the “XYZ” company, carried out in the “XYZ Model” file.

The review of the current position involves particular focus on where the trendline is at the end of the Actuals period. In this model, March 2017 is the last month for which we have data; the following months are labeled the “Forecast Period” in the chart, and are highlighted with a soft yellow background to emphasize the separation to the forecast period. (And you may recall, the yellow background is used as a way to avoid having even more lines on the chart; i.e. having one line for actuals and another for forecast actuals, one line for past trend and another line for future trend, etc.). The text box highlighted in the figure asks whether the trendline is at the appropriate level and growth rate. Do we want to modify the growth rate of the trendline so that its extrapolation into the future period is at a different slope?

**12-5 When considering the current position, particular focus should be on the Growth Rate, for initially this is the rate applied to the entire future period.**

When considering the current position, one of the most important questions to consider is what you consider the most reasonable estimate of the growth rate for today. Initially, this rate is going to be applied to the entire forecast period. Previously we had a -1.0% growth rate starting in Sep 2016, and continuing through the end of 2018. But given the slight uptick in March, it’s reasonable to modify the growth rate so that is flat (0%) from Feb 2017 forward. Arguably, the growth rate could be made flat from Sep 2016 forward; this is a judgement call. And judgement calls will be inevitably faced in virtually any situation. You’ll never get it absolutely “right”, because you’ll never know exactly what it “should” be. If it gives you greater comfort to apply sophisticated statistical techniques, go to it. Just don’t fool yourself into thinking that those techniques will get it “right”, or even necessarily, more right. What you want to give the most consideration to is that growth rate as you head off into the future. So in this example, does a flat rate seem more appropriate going forward than a slight decline of say -1%?

**12-6 The forecast trendline simply extrapolates trend through the entire forecast period; forecast actuals add on the calendar effect & seasonality.**

With the starting point for the forecast established by tweaking the trendline through the most recent actuals available, we now turn our attention to the forecast itself. The chart displayed here has widened the Y-axis so you can see how high & low the forecast (& historic) actuals run. As you can see, the trendline is extrapolated, and runs at the same pace through the entire Forecast Period. Initially, no events are inserted; we’re just looking at how it runs if the trend is simply extrapolated. The forecast actuals, in grey, have had the calendar effect & seasonality applied. Note that the total forecasts for 2017 & 2018, of approximately 437,000 each year, are highlighted in Cells S52 & T52.

If one wasn’t anticipating any change in the growth rate, and any future events, the forecast would be done! But that’s not likely to be the case. Let’s see about the business of modifying growth & inserting events.

**12-7 Changes in the forecast growth rate will often be due to external forces.**

The extrapolated trend has growth set at the rate established at the end of the actuals period. The question then becomes when and to what degree does the growth rate need to be modified over the future period. While certainly not absolute, I’ve found that most of the time adjustments to growth rates are caused by largely external factors. Why? Because we’re talking about factors that prompt an adjustment at a steady ongoing pace, and most of the actions organizations take result in shifts of growth rather than a change of its pace.

I’ve outlined here some of the major factors to consider that might change the future growth rate. The 1st is a change in the industry trend. Perhaps you belong to an industry trade group or service that is predicting a general slowdown or uptick in growth for next year. Usually these predictions do not involve shifts in growth, and they are unable to be specific about the timeline. But using their projection as a baseline, you may choose to adjust your growth rate accordingly.

Another potential adjustment to your growth rate may be due to what is occurring in an industry on which you’re dependent. Auto insurance companies should anticipate a slowdown in their sales if the auto industry is predicted to slow down. Ideally, you have calculated a correlation coefficient that measures the degree to which your sales are impacted when the dependent industry sales change – for example, a correlation coefficient of 0.2 would imply that a 10% slowdown in auto sales would lead to a 2% slowdown in your auto insurance sales.

A 3rd factor to consider is the outlook for the national economy. Here again you’re asking yourself the degree to which projected changes in key (to you) economic measures will impact growth of your sales, or any other metric you may be forecasting.

For all three of these factors, it is vital that you avoid double-counting. A slowdown in the national economy is likely to cause some slowdown in your dependent industry which in turn leads to a slowdown in your industry’s trend. You want to capture the net effect of all three. Perhaps the industry trend change is sufficient to capture the net effect to your organization. The point here is to evaluate, as best you can, what is going on in the outside world and to what degree that will likely impact your business. Of course, you also need to consider when to apply the change. To a large extent, the amount and timing of growth rate changes are judgement calls. You should be able to capture and explain your proposed growth rate change.

**12-8 Forecast growth rate changes can also be driven by internal factors.**

Obviously, internal factors can also prompt changes in growth rate, even if not as frequently as the external. Rolling out program or process changes can adjust your growth rate, particularly when these changes take many months to execute. Perhaps you’re introducing a training program that will take a year to execute, and you’re expecting your given metric (say sales) will be 3% higher as a result: you could increase your growth rate by 3% starting with the month that the rollout begins; just be sure to drop it back down 3% at the end of the program.

Another source of a growth rate change is when marketing campaigns prompt an immediate increase in sales, but then see sales drop back down in time, unless new marketing efforts are undertaken to sustain the gain. Lacking the additional efforts, you will want to have growth slow down so that sales return to their previous level, or as close to it as judged appropriate.

Inevitably, there may be other causes for expecting growth to change in future. Perhaps you want to capture expected competitive activity of an unknown or difficult to quantify impact, with an unknown date of arrival – adjusting the growth rate accordingly may be an effective way to address that, and explicitly measure and identify. Or perhaps there’s concern about corporate capacity or any of a number of possibilities that lead you to expect some change in the pace of growth for some specific duration of time.

To sum up, changes in growth can be difficult to estimate because they tend to be more subtle and take place over a longer period of time, so your history may prove less informative when it comes to estimating. The important thing is to consider the growth change possibilities, and to explicitly measure and identify them so that when actual results come in you can review them for potential explanation of differences between actuals and forecast.

**12-9 Anticipated changes are inserted in the Growth Rate section of the “Trend” tab.**

The figure here displays a change of the growth rate from flat to 1.0%, in July 2017. Perhaps the industry trade group is anticipating a 2% improvement in the industry starting in the summer. You’re not sure that you agree with all of that, nor of how much it will impact your own firm’s sales, but here you decide to at least include half that increase in your own sales, and to apply it beginning in July.

**12-10 Some “events” may need to be followed up with a growth rate change.**

We earlier talked about how you following a major one-time marketing campaign, you may need to insert a growth slowdown get sales back to their former level. Let’s walk through this with an example. Here I’ve posed that you plan some big marketing campaign for October that you anticipate will lift sales by 2.5%. The trendline pops up by the 2.5% inserted in the “events” section (Cell S40).

**12-11 Growth may need to be temporarily slowed to return sales back to their prior level, or near their prior level.**

Let’s say the campaign is something of a one-time effort. You always have marketing efforts going on, to be sure, but perhaps this particular lift is due to a special TV ad campaign that will run for a couple months; it’s out of the ordinary and costs much more than most of your efforts. In a situation such as this, typically companies will indeed see a big lift when the campaign rolls out, and then there’s a slow but steady return to the level they were at previously. Typically the reversal takes 3-6 months to play out. That’s what we’ve inserted here, where we have growth decline starting in January, about 3 months after the campaign begins, and we have set the rate of decline such that sales are back to their approximate prior level about 6 months later, at which point the original 1% growth rate is re-inserted (Cell T17). Note that the decline is to the level it would have been at had the original 1% growth rate just held steady, without the ad campaign. And note further that I still have the level slightly higher than it would have been without the campaign; it seems reasonable that even after the impact of the campaign has fully played out, sales would be at a slightly higher level.

**12-12 Perhaps the most significant modifications to your forecast will be for expected future events.**

The main driver of changes to your forecasts are likely to be “events” that cause a quick & significant shift in projected volumes. These shifts can be the result of a wide assortment of drivers: Price changes will almost inevitably cause an overnight shift in sales, and attrition. If you’ve estimated your price elasticity by following the procedure described in Chapter 10, or otherwise have determined the elasticity, you can apply it to your projected price change to arrive at an estimate of its impact. Certainly marketing campaigns will shift sales. Here your history should prove very helpful in offering assistance for estimating the likely impact. Product introductions may well prompt a shift: again, reliance on your history will provide useful guidance for an impact estimate. New regulations, from the government, or industry, or initiated yourselves, may well shift some of your metrics – costs or sales or both; here too history will hopefully provide some assistance with your estimate. And there are no doubt numerous other possible events that may cause a shift in your projected metric.

**12-13 Review your history for guidance on what impact to apply to a coming event.**

As was stated up front, your forecasts are as good as your history. And if you have successfully analyzed your history, and made appropriate estimates & identifications of past events, you should be well armed to make your forecasts for new projects you may be planning in the coming months. Perhaps your firm is planning on launching a new product line in the middle of next year, and you’re wondering what lift is appropriate to apply. Your history is a good place to start. This figure highlights past product launches for the hypothetical XYZ firm. Unfortunately, they have only introduced major new products on three occasions throughout the 16+ year history. Furthermore, the May 2006 launch coincided with a marketing campaign & a price cut – with these additional activities occurring simultaneously, we will be unable to use this event for isolating the impact of a new product launch alone (though we may be able to roughly guesstimate the distribution of the impact across all 3 activities). For the 2017 product introduction, we really only have these two examples to draw upon. And, as the 1st instance occurred so many years ago, even that will carry less weight. So one would rely heavily on the Aug ’16 launch for guidance on the upcoming one. Hopefully, your history will have more examples than this to draw upon. But even if there is only one clear past event, you still can use it for calibrating an estimate for a future instance. Here you could speak with the department(s) responsible for the product launch to ask them how much more (or less) significant do they think this new launch will be. What factors should enter into their comparison? Your Sales, (& perhaps Marketing) department should be able to provide some guidance as they’ll be more familiar with how the new product in Aug ‘16 influenced the efforts & motivations of the sales force.

**12-14 After discussions within your organization, and your examination of past like events, you’re ready to add an estimate of the coming event to your forecast.**

Once you’ve interviewed the appropriate departments, and given your own estimates of the impact of past events of a similar nature, you should be ready to come up with an estimate of the planned activity. I emphasize “your own” estimates of past events, because often other departments have their own unique perspective, and yes often, that perspective may draw a more positive conclusion regarding the impact of their efforts – quite human. This is where forecasting becomes as much an art as a science. For it is up to you to do your best to gauge what the “true” impact of past events were, as well as to judge how a given future activity will compare. Of course, I’m presuming here that “you” work in the financial analysis area and not in the subject department. You may actually be in the product development area and have been asked to estimate the impact of what your area will be putting together. Hopefully, management will permit you complete objectivity, and you’ll be able to produce your estimate unpressured. Regardless of where in the organization you reside, just do your best to derive an objective figure that you can support, based on your review of the history and your analysis of how the future effort compares.

In our running example, we’ve determined that the new product should have an impact of about 3-4%; not being able to land on one figure or the other, we decide to input an event estimate of 3.5% for August 2018, inserting it in the appropriate cell (Cell T38) in the Events section of the “Trend” tab. The running trend line in the chart jumps accordingly.

**12-15 Once your forecast is complete, you can display the coming events & growth rates on a chart to bring clarity to the driving assumptions.**

Once you’ve completed your forecast, you can put together a chart that displays all the key drivers of growth (or decline) over the forecast period. The figure here is a simple example of how one might highlight the major events over the forecast period, as well as from the recent past. The trend line is labeled because its behavior is what effectively determines the entire forecast. Underneath the trend line is a text box describing the growth rate assumption over most of the period.

Obviously your forecast may have more elements to it than what are displayed here; I’ve kept this simple for demonstration purposes. Essentially all I’m trying to accomplish here is to display the key drivers of growth over the forecast period, and the assumptions about the change. Management can then take a look at this and decide to what degree they may desire to change those assumptions – and consider if other actions may be warranted to improve the outlook.

**12-16 Displaying a good amount of history prior to the forecast helps provide context on recent performance, & info on specific results.**

I’ve observed that frequently forecasts only focus on the projected period. They might provide a prior year total figure, but otherwise all tables & charts focus entirely on the forecast itself. I think this is often a mistake, because a full display of recent (& not so recent) past can be extremely helpful for providing context. When management can actually see the slope of growth over the prior two years, see the lifts & drops that occurred as a result of recent identified initiatives, and see by comparison the assumptions about growth and events going forward, they will be in a much better position to make good decisions about the reasonableness of the proposed forecast. It is unavoidable to apply a certain level of intuition and instinct to a future in which no one *knows* what will happen. A good amount of history will go far to help inform that intuition.

As a rule, I would recommend that the amount of history displayed be as much or more as the length of the forecast period. As example, the forecast period here is for 1 year 9 months (from April 2017 thru Dec 2018), while the past displayed is for 2 years 3 months (from Jan 2015 thru March 2017).

**12-17 Annual forecast totals are displayed on the Trend page; consideration of these values is a critical part of the forecast review.**

Another key element in any review of your forecast’s reasonableness is to examine what the totals look like for the period. The “Trend” tab displays the annual totals. I leave it to the reader to determine how they wish to display this information; it will likely be consistent with an appearance your organization prefers to use, though again I would consider doing some research on how to optimize the presentation of your results. Such a display will likely include annual percentage changes – for example, that the annual increases here for 2016, 2017, & 2018, are approximately 4.8%, 4.5%, & 2.7%, respectively.

**12-18 Another check on the forecast’s reasonability is a comparison of the original extrapolated trend with the final forecast.**

You might recall we started with a simple extrapolated trend, a line that took the level and the growth rate at the last month for which we had actual data, and extended it throughout the entire forecast period. The figure on the left displays that original extrapolation, and has enlarged the estimate totals for 2017 & 2018. The figure on the right displays the final forecast. That final forecast is less than 4,000 higher than the original extrapolation, while the 2018 forecast is about 15,000 higher. These may or may not seem reasonable, depending on the entire situation – how aggressive or modest the extrapolation and the final forecast each seem, the uncertainty surrounding some of the key elements of the forecast, the personality, if you will, of the key decision makers at your organization as to how “optimistic” or “pessimistic” they tend to be, and so on.

It’s useful to bear in mind that the original extrapolated trend is where you started, it’s roughly what you’d anticipate if “nothing happened” over the coming forecast period. Given that as your starting point, how reasonable does it seem to you that your final forecast ends up as it does? If you can feel comfortable with that comparison, you’ve probably got a pretty good forecast.

**12-19 Forecasting can be a fairly simple & quick procedure, provided you’ve taken the time to prepare a clear & insightful history.**

By comparison, this chapter is quite a bit shorter than most of those that have preceded it. That might seem surprising given how complicated forecasting can be, and how important. But if you think about it, if you know your history well, then you should have a pretty good of sense of where you’re heading. And you’ll be well positioned to derive reasonable and appropriate estimates for any planned activities you may have going forward. Certainly the forecasting procedure can be much prolonged if there are a host of steps & hoops to jump through to develop a Plan or forecast within your organization. But hopefully you, and others, can come to see that clarity & knowledge about your past will immensely help inform your projections – projections developed with confidence & understanding.

**12-20 Seasonally-adjusting your history clarifies how you are doing today, and where you are headed tomorrow.**

We finish where we began. With an understanding and appreciation of the importance of learning from your history. Seasonally-adjusting that history will clarify how your performance has trended over time. It will highlight when events have occurred, events that you can review to identify the cause and to measure the impact. Understanding your history will help illuminate what actions have succeeded in past, and what have failed. Seasonally-adjusting your history will also enable you to be much more clear about how you are doing today, to identify the approximate level a given metric is at, its growth rate, and what recent events may have caused it to shift. By being clear about where you are at today, you are in a much stronger position to speculate on where you’re headed.

If you don’t know where you’ve been, how can you know where you are? And if you don’t know where you are, how can you know where you’re going? I hope that you’ve come to see that the procedure I’ve outlined in these pages provides a powerful way to clarify your past, your present, your future. Using the technique I’ve described here, you’ll be in the driver’s seat, not a sophisticated statistical software program that only a few understand. By being in control of the analysis of your performance over time, you’ll have gained greater understanding from the knowledge that history can provide. And that knowledge, combined with a powerful tool to help clarify how you’re doing, will prepare you well for envisioning & improving the road ahead.

**Overview**

**Title Page**

So, “how are you doing?”. That may be the most common question in the English language. It also gets to the heart of what much reporting and analysis aims to achieve. My name is Peter Gascoyne; I’ve been a consultant for about 30 years, working with a wide variety of industries and areas, private and public, trying to help them better understand their data and performance. That understanding can be greatly enhanced when data is seasonally-adjusted. In this presentation, after a short introduction, I’ll walk through some of the shortcomings of the more traditional ways of looking at reported data. I’ll then consider the benefits of seasonally-adjusting the data, and then finish with a brief overview of how you go about seasonally-adjusting your data. Hopefully from this you will see that a much clearer picture and greater understanding of your data can be obtained by seasonally-adjusting your data. And hopefully you’ll also see that you’ll then be able to provide a much clearer answer to that question: “How are you doing?”.

**2 Organizations produce a wide array and enormous volume of reports every month.**

With each passing month, organizations produce a thicket of reports: on sales, staffing, expenses, profitability, etc., etc. While the monthly report is the most common, periodic reports can come out as frequently as every day, or every week. And given the enormous power of “Big Data”, these reports can quickly slice and splice the data into almost infinite combinations, and drill down to the tiniest level of detail.

**3 What is the purpose of all these reports? They try to answer the question, “How are you doing?”.**

But what is the purpose of all these reports? Whether you’re the analyst producing them or the executive reading them, you need to ask yourself that question. Ultimately, the goal of most of all these periodic reports is very simple, yet very challenging: they seek to determine how you are doing. But how do you answer the question “how are you doing?”?

**4 Since data “noise” can obscure how one’s trending, many simply compare results against Plan or last year.**

Usually, when you look at monthly data over time, you’ll find there is a lot of “noise” – lots of fluctuation from one month to the next that can seem almost random. It’s hard to read trend when the data seems to fluctuate so much. In order to deal with this problem, or more accurately, to avoid dealing with it, many resort to comparisons against Plan or with how they did 12 months before. But as we will see shortly, Plans can age quickly and prove a less than useful source of comparison, while comparisons with last year can be full of hazard.

**5 Seasonally-adjusting the data can provide a clearer picture of trend, of how you are doing.**

Usually, much of what causes “noise” in data is predictable, and pertains a lot to the calendar effect and seasonality. Data shifts from one month to the next because months are relatively shorter or longer. And there can be a pattern of variability across the year – this “seasonality” might manifest as being busier in the summer and quieter in the winter. I live in Wisconsin, and as you can imagine, home sales here in the upper Midwest are off significantly when the winter snow & cold hits. Seasonally-adjusting the data makes adjustments for these predictable fluctuations. When the data is adjusted for the calendar effect and seasonality, a much clearer picture of trend can emerge.

**6 Seasonally-adjusting your data clarifies where you’ve been, where you are, & where you’re going.**

Answering the question, “how are you doing?”, is not just an exercise in quantifying your past month sales, and possibly comparing it to Plan or last year. Having a clear sense of where you are now means also being clear about where you’ve been and where you’re headed.

Seasonally-adjusting enables you to gain this perspective. It eliminates the noise of seasonality and calendar effect, and consequently clarifies your trend. Once you’ve seasonally-adjusted your data, you will be able to look at any point in time, past or present, and be able to identify: the level you are at, your approximate growth rate, and any recent event that has suddenly shifted the level for your metric. Armed with this information you gain a clearer idea of where you are now, and where you’re heading – recognizing that the future can always be full of surprises.

**7 Typical monthly reports have serious shortcomings**

Before we examine the seasonal adjustment process more closely, let’s first take a look at how many organizations typically answer the question, “how are we doing?”. Frequently, monthly reports are not much more than a basic accounting. Monthly totals are collected and compared against Plan or against last year. Often Actuals and Plan are compared on a year-to-date basis as well. And of course, all of this is done at great detail, parsing the data down to regions, and states, and sales areas, and sub-districts, and so on. All these numbers and calculations are very specific – sales were 21,418 last month, up 3.4% from last year and 1.6% above Plan. But all too often these reports only offer a patina of precision; there is a danger with just looking at a basic accounting.

**8 Comparing results against Plan can be of limited value, especially as Plans become “dated” over time.**

In many organizations, evaluation of “how you are doing” begins (& ends?) with how you’re doing against Plan. So long as you’re above Plan you’re doing fine, and preferably the gap between actuals and Plan is only expanding with the passage of time.

It is probably very likely that a positive comparison with Plan is a good thing, but be careful. And be aware there are a number of potential shortcomings associated with comparisons against Plan.

**9 Circumstances change that can instantly render a Plan obsolete and irrelevant.**

It probably goes without saying, but I’ll say it anyway, Plans can quickly become obsolete when external, or internal circumstances, suddenly and significantly change. A key person quits, a major plant is damaged, a client suddenly leaves, a competitor runs an awesome ad campaign, or perhaps you’re surprised to win a big contract, you get a great review with national coverage, a new product turns out to be even better than expected. Stuff happens, and it can instantly render even the best laid Plans obsolete and irrelevant. It’s not a failure of the Plan, it just speaks to the need to be constantly updating your Plan as circumstances change. But unless you’re doing so, you run the risk of comparing your actual results with a Plan that has become outdated, resulting in a comparison that becomes less relevant, and leaving you more unclear about how you are *really* doing.

**10 Internal politics can lead to Plans reflecting the desires of a few rather than the “reason” of most.**

Another common problem with Plans, of course, is that they can be to a greater or lesser degree the product of internal politics. Perhaps key parties are pushing for higher (or lower) numbers that seem reasonable. In the worst case, everybody is forced to work against a Plan that only a few at the top are comfortable with. Not only is buy-in lacking in these circumstances, but comparisons with Plan are less meaningful for the Plan numbers are probably unreasonable in an unpredictable way. What I mean by that is that it is not as though you can rely on the Plan being consistently too high by a certain percentage, but instead it is probably “off” to a greater or lesser degree in a fairly volatile pattern that makes comparisons with Plan that much more difficult to interpret.

**11 Plans can be prepared without understanding how you’re doing.**

Even if politics isn’t an issue, and you encounter no surprises across the year, Plans can still be incorrect if you are not clear about how you’re doing at the time you prepare your Plan. Let’s take a look at an example.

**12 Quiz: After a flat Year 1, sales show steady growth in Year 2. What is the Year 2 growth rate?**

I’m an economist, and as you may know, economists love to make assumptions. To illustrate this point that organizations may not be clear about how they’re doing when they prepare their Plans, let’s assume we’re working with data that is unusually cooperative – there’s no volatility, just straight line growth like in the diagram here.

In year 1 our hypothetical firm has no growth, selling 100 units each and every month, with a total of 1,200 units sold for the year. In year 2, it has steady growth, starting at 100 units sold and ending with 110 units sold in December – for an average of 105 units sold per month, and a total of 1,260 units sold for the year. Year 2’s sales of 1,260 units are exactly 60 units or 5% more than Year 1.

The growth rate in Year 1 is 0%; sales are flat. What is the growth rate in Year 2? (Pause here to consider your answer.)

**13 The correct answer for Year 2’s growth rate is 10%, NOT 5%.**

The correct answer for Year 2’s growth rate is 10%, not 5%. If you start the year at 100 and finish at 110, you gain 10 over the course of the year; you are growing at a 10% annual rate. Many answer this question with 5% - understandably: the 1st year has 1200 units sold; year 2 sees 1260 units, an increase of 5%.

Now this may all seem obvious given how clearly this example has been constructed. But obviously data is generally much more difficult to read, full of the kind of volatility we earlier noted. And if the data were more volatile, it wouldn’t be surprising at all if most were to conclude that growth was 5% in Year 2, and to go ahead and plan Year 3 accordingly.

**14 Planning for another 5% growth in Year 3 would actually mean NO growth.**

This chart reveals the problem with thinking growth was 5% in Year 2. Given 1260 units sold in Year 2, and a Plan calling for continuing that 5% growth in Year 3, you will end up with a Plan targeting for approximately 1,320 units sold in Year 3. But as we can see, to plan for 1,320 unit sales in Year 3 would imply planning for zero growth, a most disappointing outcome indeed. Instead, if you want to plan to sustain Year 2’s growth in Year 3, you would be looking at starting year 3 at 110, and ending at 120, for an average of 115 units sold monthly in Year 3, and a total of 1,380 units sold for the year – which works out as a 10% increase over Year 2.

You can see how easy it would be to not plan appropriately if it wasn’t clear how you were doing at the end of Year 2, in terms of the level you’re at, and your current growth rate. When data is noisy, as it usually is, and when we rely on typical reporting, and consider simply how one year’s total compares with another, it is very easy to make the mistake of putting together a Plan that is inappropriate.

**15 Comparisons with 12 months ago can be misleading, if not dangerous.**

Hopefully we can agree that comparisons with Plan can be somewhat hazardous if you want to know how you are *really* doing. Let’s turn our attention now to another common comparison organizations make – year over year (YOY). This seems an obvious comparison to make, and can easily be calculated. It also nicely avoids the problem of seasonality – it’s difficult to compare August sales with February if summer is always much busier, but comparing this August against last August seems fair enough.

It isn’t.

**16 Year-over-year (YOY) can provide a false picture if the growth trajectory has changed.**

One of the many problems with relying on YOY calculations is that they can give a false picture of how you’re doing when, not if, but when your growth rate changes. Let’s look at another simplistic example. Here we have steady growth of 10% in Year 1. In Year 2 we have no growth at all. What does the YOY picture look like?

**17 Year-over-year (YOY) can provide a false picture if the growth trajectory has changed.**

This chart not only shows sales, it also shows the year-over-year percentage change. At the beginning of Year 1, sales were at 100. In January of Year 2, they are 110, an increase of 10% year over year, as plotted here.

**18 Year-over-year (YOY) can provide a false picture if the growth trajectory has changed.**

Let’s continue the YOY calculation. In February of Year 2 sales remain at 110; in February of Year 1 they were slightly above 100, so that the YOY percentage change is slightly over 9%. In March, the % change is over 8%, and so on. It’s not till the start of Year 3 that we actually see YOY change of 0%.

Notice how striking it is that all the numbers at the start of Year 2 are around 10%, even though the trajectory has totally changed, and sales have changed from 10% growth in Year 1 to zero growth in Year 2. Again, this is an obvious example because I’m using simplified numbers. But the real data you use won’t behave so cooperatively, and such misreads as this will be much harder to recognize.

**19 YOY can give a false read over an extended time period if a sudden shift has taken place.**

Where YOY reporting can actually become dangerous is when there is a sudden shift in the metric being tracked. The simple example here shows sales declining at a 10% rate throughout year 1. At the end of year 1 some major change is introduced – perhaps a new product, or price cut, or whatever. Year 2 sales start at the same level as year 1, and then proceed to decline, again at a 10% annual rate. If you were tracking YOY numbers here you’d see there is no change throughout year 2; even year 2’s total sales will match those of year 1. It would be easy to think no “change” is taking place. And you would be clueless as to what is really going on – that sales are in fact declining at an alarming 10% annual pace.

Of course, such dramatic and obvious shifts seldom occur, but consider how easy it would be easy to misread almost any YOY numbers if they come in shortly after a sudden shift in your metric.

**20 A change in the YOY outlook may be due to what occurred 13 months ago, not last month.**

The example we just looked at also speaks to another issue with YOY reporting: when there is a sudden change in the YOY values – positive or negative – how can you later tell if YOY values are reflecting something that just occurred versus something that occurred 12-13 months before which is now no longer being picked up?

To illustrate, let’s now extend the sales trend. When January of year 3 comes in, “suddenly” you see sales are down 10% on a YOY basis. Think about that. Not only have you had a completely false neutral picture throughout year 2, but then when year 3 starts sales “appear” to have suddenly dropped. You’ll be right to think sales are declining, but think of the errors in judgement you might make if you thought this decline had just occurred. Perhaps you’ve introduced some new product, or your competitor does something around that time. You could be easily led to think the change(s) had a profound negative impact on your sales when, in fact, sales have been hurting for over two years.

Obviously data never behaves so cleanly and clearly as this example suggests – data will always have much more “noise” than this. But that’s the problem, because that noise can easily mask what is truly going on underneath.

**21 YOY values typically do not make appropriate & necessary calendar adjustments.**

Okay, you might think, maybe these issues occur when there are major changes in our growth, but that’s not going to be a problem if sales are growing at a much steadier pace. At least then the YOY number can give a helpful signal of how the past month performed. Well, sorry, but that often isn’t the case either. Why? Because of the calendar effect.

Let’s say your business only operates during the week and you are comparing Sep 2017 sales with Sep 2016. As the calendar here shows, Sep 2016 had 22 business days while Sep 2017 only has 21. All else being equal, sales will be down about 4-5% due to the shorter month. This problem is exacerbated in leap years, when month lengths measured in business days can change by as much as 2 days. Even if you’re open all 7 days of the week, you likely still have some days that are consistently busier or quieter than others. If you get 4 of those days one year and 5 of them the next, you still risk some degree of false reading by simply looking at unadjusted YOY comparisons. (And I have never seen an organization that endeavored to adjust for the calendar effect.)

**22 Some organizations choose to just provide the data.**

Okay, well maybe some of your numbers are reported without the comparisons shown against Plan or the prior year or whatever. Perhaps so many sets of numbers are provided that to perform these calculations would just be overwhelming. So you just report the numbers themselves. Not surprisingly, that too may be of little help.

**23 By itself, a number is meaningless; only by comparison with something else does it gain context.**

A number, any number, by itself, is meaningless. It’s only when we compare it with something else that we can have any sense of how large or small it is, whether or not it represents an improvement or decline in what’s being tracked. Now it could be that the reader already has some sense of what is a good or bad number. Thus, showing a baseball player’s batting average has a meaning because the reader may “know” that an average over 0.300 is pretty good. Though only the more avid fan may be aware that for this particular ballplayer it still represents a drop from normal.

**24 Numbers alone may only give the reader a sense of whether the metric is high or low.**

When we show the figures for a number of months in a row we get some sense of how the numbers are performing. Clearly here the numbers in March & April are significantly up from Jan & Feb.

**25 It can be very difficult to get a sense of trend from the numbers alone.**

What I am trying to get at here is that just reporting the numbers, even with a chart, can succeed in giving a reader only a very vague sense of how the metric is performing. What’s missing is a clear sense of trend, a clear sense of “how you are doing”.

**26 Typical reporting can risk providing a false picture of how you are *really* doing.**

Before leaving this section, there is one final point I need to make very clear. I’m not suggesting you not make comparisons against plan, or look at year-over-year. I’m suggesting you not ONLY do those things. Comparison against Plan is fine, but frequently it becomes less and less informative with the passage of time. Year-over-year comparisons are fine; just be very careful, for it is so easy to get an inaccurate picture. I don’t know about you, but I’d rather be given no information than be given the wrong information.

**27 Seasonally-adjusting your data clarifies where you’ve been, where you are, & where you’re going.**

Okay, we’ve dumped all sorts of criticism on what is for many the usual way of reporting data, and answering the question “how are we doing?”. What should one be doing instead?

I would recommend instead that you seasonally-adjust your data. Seasonally-adjusting your data basically puts all your history on a level playing field. When you chart your data it cycles up and down, experiencing a certain amount of “volatility”. Much of this volatility, or “noise”, is predictable. Monthly values will be higher or lower, to some degree, because the month is longer (or shorter) than normal. It may also be higher or lower because the given month takes place during a consistently busier (or quieter) time of year.

When you seasonally-adjust your data, not only do you clarify where you are today, you also illuminate what has transpired in past, and can forecast with greater confidence where you’re heading.

For the remainder of this overview, I am going to quickly walk through the process for seasonally-adjusting your data, as well as describe some of its benefits. This approach is described in great detail in the attached “book”; I put “book” in quotation marks as it comes in the form of a PowerPoint presentation.

**28 Seasonally-adjusted data helps clarify how you’re doing TODAY.**

I now want to go back to the question we started all this with: “how are you doing?”, and how one goes about answering the question “how are you doing?”?

For many the effective answer is to just give the data and perhaps compare it against Plan or the prior year. I’d like to propose you answer it instead in three ways:

1. What is the current level you are at today, AFTER you’ve adjusted for the calendar effect and seasonality?
2. What is your current approximate growth rate?
3. Have there been any recent events that have suddenly shifted your level up or down?

Estimate these three factors and you will have a very powerful description of how you are doing today. Let’s go through each of these measures.

**29 Seasonally-adjusting your data helps identify what level your given metric is at today.**

What level are you at today? Of the three measures this is perhaps the most important. This tells you, quite precisely, how high your given metric is. The measure of your current level is done AFTER you’ve adjusted for the calendar effect and seasonality. Essentially, if every month was of exactly the same “true” length, and your metric experienced no seasonality, no predictable cycling across the year, then what would the metric come in at? Your seasonally-adjusted level provides that. Another way to think of this is that if you were to experience no growth over the next 12 months, then multiplying this value by 12 would give you your total for the year, even though the metric would fluctuate across those 12 months because of the calendar & seasonality.

One other point: there will always be some level of unpredictable “noise”, over & above the calendar effect & seasonality. Identification of today’s level takes that into account by representing the level as the approximate average of the current and most recent month(s).

**30 Seasonally-adjusting your data helps identify your approximate rate of growth.**

The next measure is your current approximate growth rate. Think of the growth rate as the slope that you are on, that if you were to continue steadily at this rate, one year from now you would be higher or lower by that percentage.

I emphasize this rate is approximate for it is very difficult to estimate this figure. It’s hard to identify today because the inevitable “noise” of the recent and prior month may just be noise or it may prove the start of a change in your trajectory. But even in hindsight, after all the data is in, calculating the growth rate at any point in time can be challenging. Let’s face it, data for almost any given metric is not usually very cooperative. Lots of forces can push it slightly in one direction or another. And over time, it can be very difficult to separate the degree to which a metric is going up due to “growth” versus going up due to an “event” or events. That distinction determines what direction you’re heading. If no further “events” were anticipated, at what pace would your metric rise or fall over the coming months? – your answer is your approximate current growth rate.

**31 Seasonally-adjusting your data helps identify & quantify recent shifts, or “step functions”.**

The third and final metric is identifying & quantifying any recent “events”, something happening that has caused your metric to suddenly and significantly shift up or down – another common term is “step function”.

Say your sales are at a level of 100, and one year later they are at 110. In your mind’s eye you probably picture a straight line connecting those two dots, depicting a 10% annual growth rate. But that is rarely the case. Instead what happens far more often than not is that sales experience one or two or more steps along the way.

Of the three measures of how you are doing today, recent events is perhaps the most valuable. Take a look inside almost any organization today and you’ll find a host of people who are charged with trying to improve the business. If its sales, not only is there a sales staff and a marketing staff devoted to increasing it, but there’s probably folks in product development and process design and even in IT who measure their own success by how well sales improve through their efforts. And usually, though certainly not always, those successes are step functions, occasions where a new product or a new ad campaign or a big sales push leads to a significant increase in sales. It’s why many executives are paid the big bucks, to push the organization to a higher level. How successful are they? – that’s what the measurement of recent events provides.

Note that I refer to “recent” events rather than only those taking place in the current month. Usually it takes a couple of months or so to gather enough data to get an accurate measure of the impact of an event.

**32 Combining these elements makes for a more complete picture of “how you are doing”.**

When you combine these three elements, the level you are at, the approximate growth rate, and any recent events, you have a very powerful statement as response to the question, “how are you doing?”.

**33 Seasonally-adjusting your data enables you to learn valuable lessons & info from your past.**

We just looked at how seasonally-adjusting your data gives you a much clearer picture of how you are doing today. It can also provide valuable insight and lessons from the past. When you have seasonally-adjusted your history you will be able to look back and measure the impacts of past projects and programs. Combine all the similar past efforts together in an analysis and you can have some very useful guidance on what has and what has not succeeded in improving your business.

**34 Impacts of past product or process changes are easier to estimate.**

Seasonally-adjusting your data will enable you to look at a past program and easily measure the impact. Such a measure can be objective, with no undue influence to arrive at a “high” value. And it should be fairly easily done because the elimination of the predictable noise of the calendar effect and seasonality will render it easier to read how the data trended and approximately by how much the metric jumped when the new program was instituted.

You will have to be careful however, to consider whether there were any other significant activities going on at or around the same time that may have contributed to (or detracted from) that change.

**35 Impacts of Marketing campaigns can be easier to estimate, but watch out for back-end slide.**

Certainly another type of activity you’ll be able to go back and measure, objectively, is the impact of past Marketing campaigns. Objectivity can be important for there may be an understandable bias by your Marketing department when it comes to measuring the impact of their efforts.

But you also will need to look out for sales to slide back down after some time. It is often the case that major “one-time” marketing campaigns will see a significant initial bump, but then after perhaps a couple of months, give or take, sales start to slide back down to their previous level. You need to look for this as it is to be expected that unless the Marketing effort is sustained, eventually potential customers will again not have your product “top of mind”, and a sales slide will be an inevitable result.

**36 Events with no apparent explanation may be due to actions of your competitors.**

Sometimes you may see an “event” in the data but have no apparent explanation for it. Say sales suddenly drop 6%, even though you’ve changed nothing about your product or sales efforts. It could be that the drop is due to competitive activity – they run their own advertising push or product introduction. Seasonally-adjusting the data will help enable you to be aware more quickly that something has occurred, to quantify the damage, to do the research to try to determine what happened, & to decide upon an appropriate response.

**37 Seasonally-adjusting your data enables you to estimate the all-important price elasticity.**

Another area where seasonally-adjusting your historic data can prove so valuable is in being able to measure your price elasticity. The price elasticity measures the percentage change in demand given a percentage change in price. If you raise your price by 10%, by what percent will sales volume decline?

It is one of the most important metrics a company needs to know. When you’ve seasonally-adjusted your data you will be able to go back and estimate this figure. The website goes into much more detail on how to go about this procedure. It’s not easy, but it’s so important that it’s well worth the effort.

**38 Past pricing events can be examined and plotted to derive a price elasticity estimate.**

The chart on this page gives you an idea of how you can measure your price elasticity. Past pricing events are examined and your change in sales is plotted against your change in price. If you’re lucky, the data will behave well and a fairly confident measure can be obtained. In this chart, the price elasticity for price increases is estimated at -0.75, implying a 10% price increase would lead to a 7.5% drop in sales, all else being equal. Meanwhile, the price elasticity for price reductions is -0.9: a 10% price cut leads to a 9% increase in product sales. It is frequently the case the price elasticities for price increases and price decreases are different.

**39 Seasonally-adjusting your data will enable you to better forecast the future, and improve your future.**

When you seasonally-adjust your data, not only do you gain greater clarity and insight into where you’ve been, and how you are doing today, you’ll also get a clearer idea about where you’re headed. And the lessons learned from your past should help you to improve that future. It’s all connected – by knowing where you’ve been, you know where you are, and by knowing where you are, you know where you’re headed. Inexact though it may still be, for data is never that cooperative, you should still be able to better know, and understand, where you’re heading, and why.

**40 If the current level and approximate growth rate are known, the forecast can be a simple extrapolation.**

At its simplest level, forecasting your future simply requires taking the level you’re at today and extrapolating it using your current estimated growth rate.

**41 Modify projections for any planned activity or events you anticipate are coming.**

You can then modify your forecasts for any new activity or event that is planned or anticipated. How much of an increase or drop should you apply to these coming events? The answer should be based on how you’ve done in past – which you now should have a better handle on – and your estimate of the degree to which the coming event is likely to be higher or lower than this past record suggests. Your “fine-tune” adjustments should also take into account some read of the external environment and how that may influence the outcome. Hopefully your forecast is higher in part because your analysis of the past gives you greater confidence regarding what types of activities are more likely to be more successful.

**42 Avoid taking costly and inappropriate actions.**

Have you ever looked back and acknowledged that some activities you’ve undertaken were perhaps unnecessary and overly expensive? Bad Plans are especially susceptible to this. You put together a Plan that is unreasonably aggressive or optimistic, and at midyear when you can tell you’re not likely to meet it you sort of panic, compound the planning error, by throwing money at a costly and poorly planned marketing campaign or price change or whatever. Hopefully by better trending your data you will be more able to avoid making the bad Plans that lead to taking costly and inappropriate action.

Seasonally-adjusting your data is not easy; the book makes clear that there’s a lot involved. But arguably the effort is worth it as you are able to produce better Plans and be clearer about where you’re headed. And hopefully too your organization is more motivated as you’re able to craft more realistic yet nonetheless challenging Plans for them to pursue.

Good planning can be expensive, but bad planning can be even more expensive.

**43 But HOW do you do it? Let’s quickly walk through the seasonal adjustment process.
(Details are in the book/appendix.)**

So if seasonally-adjusting the data is the way to go, how do you go about doing that? Here I will briefly describe the steps. This process can be quite involved, for it requires carefully analyzing your history to determine the appropriate adjustments to make to the data. Elsewhere on this website you will find my book that walks you through the process. And you will find Excel templates that will enable you to make many of the calculations more quickly. I particularly recommend using the templates for normalizing the data to adjust for the calendar effect. But I would encourage users to take the trending model and recreate it themselves, using the model for guidance but ultimately creating a model that you better understand because you developed it yourself.

Seasonally-adjusting your data involves three main steps. First, you “normalize” your data to adjust for the calendar effect. Next you adjust your history for growth & events. Finally, you’ll want to plot the results and tweak your seasonality estimates.

**44 The 1st key step is to “normalize” the data, to adjust the data so every month is of equal length.**

Before we adjust for seasonality, the predictable fluctuation in measure that occurs across the year, we first want to “normalize” the data, to adjust for the calendar effect. This adjustment essentially makes every month of equal length – we’re lengthening the shorter months and shortening the longer months, so that the seasonality calculation is not biased by calendar. For example, February is the shortest month of the year (usually) and has a holiday. We don’t want seasonal factors to capture February as a quieter month because it is shorter, we want the seasonal factor to capture the degree February is quieter simply because it may be a quieter time of year.

All data has “noise”, volatile fluctuations across the months. Though you may not be aware of it, for many organizations the largest cause of those volatile month-to-month fluctuations is the calendar effect.

**45 Equated Day Factors (EDFs) measure the relative weight of activity across each day of the 7 day week.**

In order to know how long the month is, you want to start by measuring how busy each day of the 7-day week is, relative to a daily average. You’re doing this because for every month but February, there will be 2 or 3 days that occur 5 times in a given month rather than 4. How relatively busy are those days? Equated Day Factors measure that. In the chart shown here, using data I fabricated on my own for illustration, Friday is clearly the busiest day of the week, coming in more than 50% above the average day length, while Sunday is the quietist, at roughly just 60% of the length of an average day.

**46 Holiday Factors measure how holidays, and the days preceding & following, compare with normal.**

Holiday factors measure how busy a given metric is on the holiday. Also, and this is important, they capture how relatively busy the metric is the day or days immediately preceding and following the holiday. Oftentimes when businesses are closed for a holiday, the day after can be a bit busier, and perhaps the day before is busier as people rush to get their orders in beforehand. Or the opposite can occur, and the holiday can drag down the adjacent days. This can particularly occur with a holiday like July 4 when it falls during the week: the Monday before a Tuesday Independence Day may well come in much quieter than usual.

In the example shown here, factors are shown for what might occur in and around Memorial Day. Business is very quiet for the holiday itself, coming in at just 20% of normal. It’s a little busier the Friday & Saturday preceding. It’s definitely up for the Tuesday following, about 30% above average – perhaps a little “catch-up” is occurring. By the Wednesday following business is back to normal, indicated with a holiday factor of 1.00.

**47 Combining the EDFs & Holiday Factors arrives at a “truer” length for each month.**

When you combine the equated day factors and the holiday factors you arrive at a “truer” measure of how long the month really is. Note here that most days have a holiday factor of 1.00 – this simply implies that there is no increase or decrease in the day length due to holidays; it’s there so that you can simply multiply the EDF by the holiday factor to obtain each day’s “truer” length. In our example here, the “truer” length of May 2016 comes in at 30.31 days.

**48 The normalization factor captures how each month’s length compares with the average month length.**

When every month’s length has been determined you can calculate what the average month length is. Then, when you compare each month’s specific length with that average, you have a “normalization factor” which measures each month’s relative length.

In our example here, May 2016 has a normalization factor of 1.01, implying it is about 1% longer than average.

Note how the average month length is 29.94 days. You’d think it would be above 30, since there are 365 days in the 12 month calendar year. The average comes in at less than that because of the effect of the holiday factors. If business is shut or quieter during the holidays, then the year is effectively shortened by them.

**49 Adjust the original data for the calendar effect by dividing each month’s value by the normalization factor.**

Armed with normalization factors that compare each month’s length with average, we can now “normalize” our data. We earlier observed May 2016 was about 1% longer than average. Dividing that month’s original amount – 39,118 – by its normalization factor – 1.01 – we get the normalized amount – 38,667.

We now have data that is expressed in a form as though every month was of equal length. We’re now ready to estimate seasonality.

**50 The next step in seasonally-adjusting the data is to adjust for growth & events.**

In order to estimate seasonality, the next key step in the seasonal adjustment process will be to adjust the data for growth and events. This is something of an iterative process, described in detail in the “book”. But essentially what we’re trying to do in this step is to ensure seasonal factors capture how busy a month is strictly because it is that time of year. This whole procedure is undertaken using the normalized data that’s just been developed.

**51 Growth is removed so the seasonality measure only captures seasonality.**

Why do we need to adjust the data for growth and events before we calculate seasonality? The importance of adjusting growth is apparent in this simple example. Here there is some seasonality across the year, with summer busier than the rest of the year. But notice how because of growth, December (as well as October & November) is consistently busier than January (& February). If you were to estimate seasonality without adjusting for the growth you would falsely conclude December is significantly busier than January.

**52 Growth is removed so the seasonality measure only captures seasonality.**

Here you can see how the data behaves when growth is taken away. Note how January & December in Year 2 are at a much more similar level – no longer does December come in much higher than January, thanks to our adjusting the data for growth.

**53 Adjustments need to be made for events as well.**

We also need to adjust for events. Here we have a 5% bump in September that we would want to adjust for. The adjustment would reduce the last 4 months of this year by the 5% bump. Just as with growth, adjusting for events helps ensure the measured seasonal pattern captures fluctuation across the typical year that is due to seasonality alone.

**54 After adjusting for the calendar, growth & events, the “true” underlying seasonal pattern emerges.**

Once the data has been adjusted for growth & events, as well as the calendar effect, we have a clearer picture of the “true” seasonal pattern across the year. The appendix goes into greater detail on weeding out the outlier years. The highlighted line here depicts the estimated seasonal pattern. Yes, I acknowledge that rarely will the pattern each year be as clear and steady as it is here. “Noise” was much reduced for my hypothetical dataset in order to help clarify the points being made.

**55 Applying normalization and seasonal factors to history gives you the seasonally-adjusted data.**

Once you have your normalization factors to adjust for the calendar effect, and the seasonal factors to adjust for seasonality, you are ready to calculate the seasonally-adjusted history. For May 2016, for example, the original sales volume – 39,118 – is divided by the normalization factor – 1.01 – and the seasonal factor – 1.13 – to arrive at a seasonally-adjusted amount of 34,219.

**56 The seasonally-adjusted data is then plotted.**

The seasonally-adjusted data can now be plotted. The seasonally-adjusted data is clearly much smoother and easier to read than the original actuals. You can see that not only are the busier summers not evident but the greater volatility caused by the calendar effect has been reduced as well by applying the normalization factors.

**57 Adding a 3-month moving average can make past trends easier to discern.**

I usually recommend a 3-month moving average be inserted as well just to render the result a little smoother and easier to read. This helps reduce some of the inevitable remaining “noise” that exists even after the seasonal adjustment has been made. Note that in this particular example shown here, the 3-month moving average isn’t helping much, and is indeed somewhat masking the events. That’s because this dataset is so well behaved. Usually, I assure you, a 3-month moving average will further help clarify how your data is trending over time.

**58 NOW you have a much clearer answer to the question, “How are you doing?”.**

Et voila. Now you have a much clearer answer to the question, “How are you doing?”. By adjusting your data to remove the predictable noise of the calendar effect and seasonality, you will obtain a much clearer picture of how your data has trended over time.

Now the process of seasonally-adjusting your data isn’t simple, but hopefully you can see how valuable it can prove in clarifying how you’re doing. And hopefully using the templates and the “book” on this website, you won’t find it that difficult either.

You may decide to purchase software that seasonally-adjusts the data for you, or perhaps you already own & use such software. That’s fine, but I would add a note of caution. While you may understand how to handle the software – the steps to follow, the keys to press, the data to pull, and so on – do you understand what the software is doing behind the scenes? Do you know what’s happening inside that “black box”? Is it normalizing the data before it calculates seasonality, or is it somehow combining these two processes? If you don’t really understand it, and are just accepting the output with a liberal amount of faith, then arguably you are not in charge of your software, your software is in charge of you. At the very least, may I recommend feeding your software normalized data that’s already been adjusted for the calendar effect, and see what kind of difference that might make.

**59 … and where you are going.**

To conclude, if you don’t know where you’ve been, how can you know where you are? And if you don’t know where you are, how can you know where you’re going? Seasonally-adjusting your data helps clarify how your data is trending, and will enable you to answer with much greater confidence that simple yet remarkably important and challenging question: “How are you doing?”