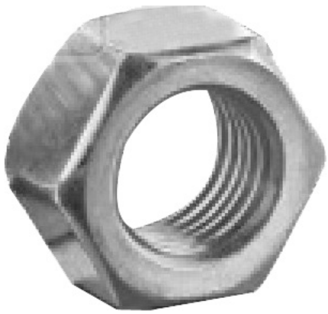


MODULE 27
Calculating Payroll

Learning Objectives:
- *Understand the elements of pay*
- *Understand the elements of benefits*
- *Estimate the future-year pay for an employee*
- *Use a simulation to estimate pay for work unit or agency*
- *Estimate turnover*
- *Estimate vacancy*

Pay for a single individual involves wages or salary and benefits. **Wages** generally refer to pay that is accumulated at an hourly rate, while **salary** refers to pay that is paid by the pay period without specific cross-reference to hourly records.

Typically **pay period** lengths are as follows:

- Weekly
- Biweekly
- Semimonthly
- Monthly

On an annual cycle, there will be 12 monthly or 24 semimonthly pay periods. Typically, there are 26 biweekly and 52 weekly pay dates, but when the first pay date is the first day of the fiscal year, there are 53 weekly pay dates or 27 biweekly pay dates.[1,2]

---

1. In a strict accrual system, the value of pay accrues in the year earned, so these dates are irrelevant. But in practice, such continuous rolling obligations may be treated as accruing on the pay date, in which case, these irregular events require careful oversight.

2. In a leap year, there are 53 weekly or 27 biweekly pay dates when the first pay date is on either the first or second day of the fiscal year.

For both salaried and wage employees, there can also be **overtime pay**. Generally, overtime pay is governed by federal law, with some exemptions for state and local employees. Management and some professional employees may not be entitled by law to overtime pay but may have guarantees in contracts. Typically, overtime pay refers to work beyond 40 hours a week, regardless of whether the covered employee is salaried or hourly. Overtime pay is at a rate not less than 1.5 times the regular hourly rate.[3] For covered salaried employees, the salary can be converted to an hourly rate by dividing the annual salary by 2,080 hours. When computing expected annual payroll, planned or recurrent overtime should be included.

Pay Scales

With few exceptions, governments and larger organizations do not use individual pay rates. Instead, they use pay scales. A **pay scale** is a table that has a series of **pay grades** and a series of **pay steps**, as shown in Table 27.1. Political appointees (directors) may be nonscale employees. The scale sets an annual rate, which can be converted by pay method (divided by periods for monthly or semimonthly pay, divided by hours for hourly pay, or multiplied [hours times units] for weekly or biweekly pay) to determine the pay rate as transacted. Employees are hired into a pay grade and at a step. Normal conditions typically allow annual progress along steps—usually on or near the employment or promotion anniversary date—until the employee reaches the maximum step. Because many employees progress along steps and because they may begin their employment anytime during the fiscal year, it is necessary to account for the transition between steps for each employee as part of the pay calculation.

TABLE 27.1
Typical Pay Scale (annual pay in $)

	Step				
Grade	1	2	3	4	5
1	27,000	27,675	28,367	29,076	29,803
2	29,160	29,889	30,636	31,402	32,187
3	31,493	32,280	33,087	33,914	34,762
4	34,012	34,863	35,734	36,627	37,543
5	36,733	37,652	38,593	39,558	40,547
6	39,672	40,664	41,680	42,722	43,790
7	42,846	43,917	45,015	46,140	47,294
8	46,273	47,430	48,616	49,831	51,077
9	49,975	51,224	52,505	53,818	55,163
10	53,973	55,322	56,706	58,123	59,576
11	58,291	59,748	61,242	62,773	64,342
12	62,954	64,528	66,141	67,795	69,490
13	67,991	69,690	71,433	73,218	75,049
14	73,430	75,266	77,147	79,076	81,053
15	79,304	81,287	83,319	85,402	87,537
16	85,649	87,790	89,985	92,234	94,540
17	92,500	94,813	97,183	99,613	102,103
18	99,900	102,398	104,958	107,582	110,271
19	107,893	110,590	113,355	116,188	119,093
20	116,524	119,437	122,423	125,484	128,621

To set up the table, we determined a relationship among all of the cells. The table is structured so that each step is 2.5% higher than the step before it. Moreover, the beginning pay step at each grade is 8% higher than the next lower grade. These assumptions allow us to array all the data in the entire table using formulas after entering only the lowest value in the top left cell. In the real world, if such relationships are not so clear in a pay system, you can instead enter the full array itself.

Pay can also be variable during the year or between years for these reasons:

- The entire body of employees may receive a **cost of living adjustment (COLA)**, a pay increase that may be discretionary or an element of a contract. Generally, this increase will be

3. For more information on overtime, see http://www.dol.gov/whd/overtime_pay.htm.

a small increment, maybe 2%–4%. It will apply to the pay scale, thus changing all employees' pay, with the possible exception of a nonscale director. It begins with the effective date, which typically (but not always) coincides with the first day of the fiscal year.

- There may be a provision for bonuses that are not added to the base pay. These could include the following:[4]
  - **Longevity pay**, a small increment once every 5 or 10 years after an employee reaches the maximum step
  - **Efficiency pay**, a distribution of a small portion of budgetary savings resulting from innovative work practices. Efficiency or similar bonuses may be awarded to all employees or all otherwise satisfactory employees in a work group. In general, while this sort of pay may be recognized in the pay plan, it is awarded retrospectively when there is an efficiency savings to recognize.
  - **Performance incentive pay** for exceptionally high performance. This may be allocated to a work unit and distributed in some manner to individuals at the end of a review cycle.

For the remainder of this discussion of pay, it will be assumed that the employee is salaried and is paid monthly.

For employees who have not reached the maximum pay step, pay is determined first by finding the appropriate number of regular hours in each cell in Table 27.2 (the variables labeled in the table are used in subsequent formulas):

TABLE 27.2
Appropriate Hours

Hours Worked	Before Step Increase	After Step Increase
Before COLA	$R_{original}$, H_1	R_{step}, H_3
After COLA	R_{COLA}, H_2	$R_{COLA+step}$, H_4

If there is no COLA or it is granted on the first day of the fiscal year, the table can be collapsed into one row, either the top row if there is no COLA or the bottom row if there is one. If the employee is at the maximum step or is otherwise ineligible for a step increase during the current year, the table can be collapsed to one column. No more than three cells should have entries for any particular employee, even when the step increase and the COLA fall on different dates that are not at the beginning of the fiscal year. For certain employees, the hours each cell must be multiples of are as follows:

Weekly	40
Biweekly	80
Semimonthly	86.67
Monthly	173.33

A similar but separate table is required for overtime hours. Typically, these hours would be substantially less than regular hours. Hourly employees should typically earn overtime only if they are first working full-time (40 hours a week), but when employment cycles seasonally, they may work part-time some weeks and overtime other weeks.

These factors yield the following formulas:

$$\text{Base} = R_{\text{original}} \times H_1 + R_{\text{COLA}} \times H_2 + R_{\text{step}} \times H_3 + R_{\text{COLA+step}} \times H_4$$

4. Other bonus options may also exist.

$$\text{Overtime} = R_{\text{original}} \times 1.5 * H_1 + R_{\text{COLA}} \times 1.5 \times H_2 + R_{\text{step}} \times 1.5 \times H_3 + R_{\text{COLA+step}} \times 1.5 \times H_4$$

$$\text{Inital pay} = \text{Base} + \text{Overtime}$$

In these formulas, R refers to a pay rate and H refers to hours. To the initial pay, we add any of the bonus elements:

$$\text{Pay} = \text{Base} + \text{Overtime} + \text{Longevity} + \text{Efficiency} + \text{Performance}$$

Benefits

Benefits include some that are purchased by the employee and that do not add to the employer's cost. However, other benefits are paid by the employer.

FICA, Medicare, and Unemployment Compensation

There are four benefits that employers must pay because of legal requirements:

- **Federal Insurance Compensation Act (FICA)** is the tax to support Social Security. The employer and the employee must each contribute 6.2% of the value of the employee's salary, up to a capped amount. In 2013, the income maximum for FICA was $113,700.[5,6] On average, this maximum has increased 2.7% over the past 10 years.

- Medicare tax (sometimes labeled **HI**) is 1.45% for the employer, with a like amount for the employee in most cases. Certain high-earning employees may owe more.[7]

- **Unemployment insurance** and **workers' compensation** tax rules differ by state and by type of work, so it is probably best to look for state- and industry-specific information. However, general information is available from the Bureau of Labor Statistics (BLS).[8] No specific amounts are given for these taxes, but their combined proportion is indirectly addressed for private industry,[9] where the total amount "legally required" is 0.55% more than the combination of FICA and HI. Thus, 0.55% of pay is a rough value for these two items.

Health Benefits

Beginning in 2014, the Affordable Health Care Act penalizes many employers who do not provide health insurance. The entirety of this rule is arcane. Although the BLS has data on health

---

5. The source for this information is http://ssa-custhelp.ssa.gov/app/answers/detail/a_id/240/~/social-security-and-medicare-tax-rates%3B-maximum-taxable-earnings.

6. Approximately one fourth of state and local governments (by employee numbers) do not participate in Social Security. They do have some type of pension plan, though it may not be as generous.

7. The source of the additional information is http://www.irs.gov/Businesses/Small-Businesses-&-Self-Employed/Questions-and-Answers-for-the-Additional-Medicare-Tax/.

8. The BLS compensation survey is periodically published at http://www.bls.gov/news.release/ecec.nr0.htm.

9. The referent is private employees because of the information in footnote 5.

insurance, the number is reported as proportional to earnings, which is not realistically related to the cost experienced by employers or employees. BLS shows health insurance as costing roughly 11.7% of earnings. The Kaiser Family Foundation's 2012 survey[10] shows that on average for family coverage, health insurance premiums are $15,745, with worker contributions of $4,316 and the employer's share being $11,429. For individual coverage, the premium is $5,615, and the employer's share is approximately $4,600. The report does not show the distribution between family and individual coverage. Approximately 81% of employees who are offered coverage take it. Based on recent US Census data,[11] approximately 15%–30% of adults are in living arrangements that are associated with being users of individual health insurance rather than family health insurance. Based on all of these facts, we treat the 2012 employer share health premium as follows:

$$(\$11{,}429 \times 77.5\% + \$4{,}600 \times 22.5\%) \times 81\% = \$8{,}013$$

This amount should be inflated by health insurance inflation of 6.5% per year.[12] For very small employers that cannot afford to be underfunded by even a small amount, the take-up rate should be reset to 100%, in which case the premium is roughly $9,892.

Retirement

The other major benefit employers commonly contribute to is a retirement plan. BLS shows that governments contribute about 9% of salary to such plans. This number may be high because some governments are substituting for Social Security. Private employers contribute 3.2%, but this number may be an underestimate with respect to governments. However, it is likely that governments that are not substituting for Social Security are contributing an amount similar to that contributed by private employers. To balance these concerns, the judgment value used here is 5%.

Leave

BLS counts paid leave as a benefit. This benefit leads to an outlay other than salary only if the employee is replaced while on leave or if the employee is paid for some portion of paid leave on separation. Otherwise, the benefit is consumed through payment of salary during leave periods. If the salary is based on full-year pay, no additional cost must be included for this purpose under most circumstances.

Other Benefits

Other benefits may include life insurance, disability insurance, or other unspecified compensation. The cost of these is quite small. The BLS data suggest a ratio of 0.4% of salary for these benefits for public employees.

The formula for benefits is as follows:

$$\text{Health} + r_{\text{FICA}} \times (\text{lesser of}\left[\text{Pay}, \text{FICA Max}\right]) + \text{Pay} \times (r_{\text{HI}} + r_{\text{WC+UI}} + r_{\text{Ret}} + r_{\text{Other}}) =$$

$$\text{Health} + 6.2\% \times (\text{lesser of}\left[\text{Pay}, \text{FICA Max}\right]) + \text{Pay} \times (1.45\% + 0.55\% + 5\% + 0.4\%) =$$

---

10. The source for this information is http://kff.org/health-costs/report/employer-health-benefits-2012-annual-survey/.

11. The source for this information is Table A2 at http://www.census.gov/population/www/socdemo/hh-fam/cps2010.html.

12. Health insurance inflation may slow in future years.

$$\text{Health} + 6.2\% \times (\text{lesser of}\left[\text{Pay}, \text{FICA Max}\right]) + \text{Pay} \times 7.4\%$$

Where the actual benefit rates or fixed premiums are known and differ from the ones shown here, one should use them.

By combining the pay formula and the benefit formula, one gets the total cost of the employer for an employee. These costs can be estimated into the next year using a reasonable inflation rate for the FICA maximum and for health insurance. For numbers that may be based on an estimated inflation rate, one should be aware of the base year of the data. The Kaiser data discussed above have a base year of 2012. The FICA maximum is for 2013.

The most complex part of building an estimating **simulation** is determining how many hours belong in each of the four cells of Table 26.2. In the example, this problem is simplified by assuming that the COLA, if there is one, is granted on the first day of the fiscal year—which is the common practice. Thus, we collapse Table 26.2 into the top row. Still, pay periods must be distributed between the period paid at the lower step and the period paid at the higher step.

One way to do this, which may be labeled the brute force method, is to make a data set that has one record (row) for each future pay date for each employee. Then assign any change in pay (such as COLA or step increase) to the relevant records following the date of the pay change. Include the planned values for each type of cost (such as FICA and health insurance) with each record. The total value of the entire data set can then be summed to estimate the future cost of pay, subject to turnover and vacancy rate issues to be discussed later in this module. For small employers, where the risk of error with respect to assigning dates of pay changes is small, this may be a relatively effective method. For larger employers, this may not be the most efficient method; they may use a more developed simulation with only one record for each position. The point of either the brute force method or the simulation method discussed later is to computerize the full cost of pay and benefits accounting for the many elements that contribute to pay.

Creating a Payroll Simulation

In this section, we discuss simulations with *one record for each employee*. If the analyst prefers a simulation with one record for each instance of a pay issuance, then the goal is to pre-create a mirror pay-and-benefits record for each pay issuance of each employee for the entire fiscal year and to sum the effect, subject to any limitations such as Social Security maximums. While possible, this can become unwieldy for all but the smallest of employers. Mirror pay issuances can be generated through a mirror or test payroll system that does not make actual pay issuances but produces all the necessary duplicate payroll data, or it can be generated using a spreadsheet program or other similar software.

Simulation is a method of forecasting using all the factors and data elements of the existing system. For payroll, where critical elements of information are already in hand, simulation is frequently the optimal method of forecasting. Here we will describe a simulation model to forecast next year's payroll, occasionally referring to the example in the spreadsheet.

The Simulation Model

The simulation model uses data contained in a payroll system, which is a specialized database of employee information. In it, all the employees are listed with critical elements of information, such as position anniversary dates, payroll grades and steps, and start date in the position. As described before, one of the most complicated issues in simulating the payroll is dividing pay periods between those before a step increase and those after a step increase. This division usually occurs with the next pay period after an employee's employment anniversary. If the anniversary is on the first day of a pay period, it will typically occur with that pay period. To simplify this matter, the pay periods for this discussion are monthly. More complex devices for finding precise anniversary dates are used for semimonthly, biweekly, and weekly pay periods; these are discussed at the end of this section. As an example for this discussion, we use a hypothetical payroll system with 15 employees. These employees and their critical data are shown on the left side of Table 27.3.

With this information, the "rules" outlined below, a decision, and a bit of empirical information, we can forecast a payroll for the next fiscal year. For many governmental departments, with the notable exception of organizations that make transfer payments, payroll is the chief source of expenditure. Thus, making an accurate payroll forecast goes a long way toward accurately forecasting the total expenditure needs of the organization. We will return to our example after we review the steps needed to create the forecast.

What Is the Pay Period?

In our example, employees are paid once a month, on the last day of the month. Monthly pay periods are more convenient for simulation than other pay periods, but those can be handled as well. For other pay periods, you will need to determine whether each pay period is before or after the anniversary within the employment anniversary month. This requires some complex Excel (or other software) formulas, which are substantially simplified when using monthly pay periods.

Is There Any Lag With Respect to the Pay?

In our current example, there is no pay period lag. But some governments hold over one or two weeks' pay when employees are hired and pay them at the end of their second or third week. This is a far less common practice with payrolls that use longer pay periods. **Lagging** does not change the fiscal period for which the pay is due; however, it may affect cash management issues.

When Are Increments Recognized?

It is also important to know when step increases, performance increases, and payroll increases due to wage adjustments will be reflected in the pay. In our example, wage adjustments (COLAs) begin with the first payroll in the fiscal year. Step increases begin with the first full pay period after the anniversary date. Our example assumes a longevity payment that is paid as a nonrecurring bonus at the next paycheck after the 10th anniversary date and every 5th anniversary thereafter. A performance bonus is available to be distributed to the work unit accordingly and in some manner at the end of the fiscal period.

Who Gets What Kind of Payments, and Do They Go Into the Employee's Pay Base?

In our example, the agency pays an annual step increase each year for 4 years, until the employee reaches the fifth step. The steps go into the employee's pay base. In real systems, there are usually either more steps, or people are likely able to move from one grade to another, having the effect of producing more steps.

We also posit a 0.5% performance payment that goes to half of the employees. The performance payment occurs every year. It does not go into the employee's pay base, so after 1 year, the employee will see a decline in take-home pay unless the employee receives a step increase. As mentioned earlier, there are also longevity payments at certain anniversaries that do not go into the base.

What Do You Do With Vacant Positions?

Vacant positions (or more broadly, classes of new hires) should be represented at the correct level of pay. Some organizations hire strictly for entry-level positions. Police officers, teachers, and firefighters are common examples of employees hired only at the entry level.

TABLE 27.3
Hypothetical Payroll Data

Information Basis											Estimating Pay					
Emp. No.	Last Name	First Name	Position Title	Pay Grade	Pay Step	Next Step	Length of Service (years)	Anniv. Month	Anniv. Day	Hired	Pay Periods Before Step Increase	Pay Periods After Step Increase	Current Periodic Pay	Future Periodic Pay	Wage & Salary	Longevity Pay at Year 10, 15, 20, 25, 30
45123	Stone	Melinda	Supervisor	19	5	5	10	9	15	9/15/2005	3	9	$9,924.43	$9,924.43	$119,093.16	$5,000
45124	Delrico	Jose	Team Leader	15	3	4	2	7	22	7/22/2013	1	11	$6,943.25	$7,116.83	$85,228.40	$0
45125	Shaft	George	Office Assistant	5	5	5	6	3	1	3/1/2009	9	3	$3,378.88	$3,378.88	$40,546.58	$0
45126	Vacant		Technician	12	3	4	0	6	30		0	12	$5,511.78	$5,649.57	$67,794.84	$0
45127	Mayes	Angie	Technician	12	2	3	1	5	16	5/16/2014	11	1	$5,377.34	$5,511.78	$64,662.54	$0
45128	Othman	Jerry	Technician	12	4	5	3	6	13	6/13/2012	0	12	$5,649.57	$5,790.81	$69,489.72	$0
45129	Bierns	William	Team Leader	15	5	5	16	5	7	5/7/1999	11	1	$7,294.75	$7,294.75	$87,537.02	$0
45130	Thorndike	Melvin	Office Assistant	5	4	5	6	3	24	3/24/2009	9	3	$3,296.47	$3,378.88	$39,804.88	$0
45131	Hagar	Nancy	Technician	12	3	4	4	1	11	1/11/2011	7	5	$5,511.78	$5,649.57	$66,830.28	$0
45132	Kaplan	Linda	Technician	12	4	5	6	1	14	1/14/2009	7	5	$5,649.57	$5,790.81	$68,501.04	$0
45133	Takei	Ralph	Technician	12	4	5	5	3	1	3/1/2010	9	3	$5,649.57	$5,790.81	$68,218.56	$0
45134	Perez	Javier	Team Leader	15	4	5	5	2	5	2/5/2010	8	4	$7,116.83	$7,294.75	$86,113.66	$0
45135	Mead	Florence	Technician	12	1	2	2	6	17	6/17/2013	0	12	$5,246.19	$5,377.34	$64,528.10	$0
45136	Wolff	Dena	Technician	12	3	4	3	3	1	3/1/2012	9	3	$5,511.78	$5,649.57	$66,554.69	$0
45137	Wilson	Michael	Technician	12	2	3	3	3	15	3/15/2012	9	3	$5,377.34	$5,511.78	$64,931.41	$0

Other organizations have much more complex rules about hiring at various steps. For example, many governments hire from within at low steps but hire from outside at higher steps. If a person moves from a lower grade within, that employee may be at step 1. If that employee enters from outside and had been paid a competitive salary, the employee's pay might begin at the top step. In our example, we assume that one new employee will be hired near the beginning of the future year and will be paid at the middle step.

WHEN DOES THE FISCAL YEAR BEGIN?
In our example, it begins on July 1.

Forecasting Salaries

In this section, we discuss how to use the information outlined above to make a payroll forecast. Table 27.3 shows the first stage, which is the estimation of the base salary across two periods: the period labeled "R_{COLA}, H_2" in the bottom left cell of Table 27.2 and the period labeled "$R_{COLA+step}$, H_4" in the bottom right cell of Table 27.2 (before we adjust the pay rate, the same two elements will correspond to the two cells immediately above them). For convenience, we will refer to these as P_1 = Period 1 = before step increase = "R_{COLA}, H_2" and P_2 = Period 2 = after step increase ="$R_{COLA+step}$, H_4."

Pay Periods

Each employee who has not reached the top of the pay scale will be paid at two different pay rates during the course of the fiscal year. The problem to solve is how many pay periods will be at one pay rate and how many will be at another. If you look back at the rule, it states that a person moves from one step to the next at the beginning of the next pay period after the work anniversary, and there is no lag between change in pay status and receipt of paycheck.

In the spreadsheet, we determine this by determining the number of pay periods before the step increase, P_1, and then subtracting that from the total periods to get the periods after the step increase. This would be easy if the pay periods matched the calendar year, since Excel and other software typically count January as 1, February as 2, and so forth. However, our fiscal year begins, as many do, in July. So we have to shift the month number. For June through December, we need to subtract 6 from the month number. For January through May, we need to add 6 to the month number; we include June in the subtraction to make its post-step value zero rather than 12, as the end of the previous cycle. This math expression is shown as follows:

$$= \mathrm{IF}(M > 5, M - 6, M + 6)$$

The letter M represents the month number for the employee's employment anniversary month. In the spreadsheet, it will be the address of a cell that contains that information.[13] This and other formulas and expressions are repeated on multiple rows with changing cell references. This expression uses a conditional statement so that for the month of June or later, the value is reduced by 6 units, and for January through May, the value is increased by 6 units. The result is a new series: June, 0; July, 1; . . . ; December, 7; January, 8; . . . ; May, 11. Thus, this formula provides the number of months we want to pay at the old step when the employee's pay changes at the end of that month.

13. It is this formula that is much more complex for semimonthly, weekly, and biweekly pay cycles.

Combining Pay Periods With Pay Amounts

In the next two columns, we multiply the number of periods before the step increase by the lower step rate and the remainder by the second step rate.[14]

Vacant Positions

We have included one vacant position during the forecast period, so we need to look back at our hiring expectations to estimate values for columns "Step," "Next Step," "Anniversary Month," and "Anniversary Day." We stated that the hypothetical employee would be hired at the third step, and we set the anniversary date as the last day of the fiscal year. If we knew an actual anticipated hiring date, we could instead set the step and anniversary date to reflect the actual anticipated information. The inclusion of these assumptions is shown in the yellow-highlighted cells in row 6 of the Payroll Forecast tab in the exercise spreadsheet.

Longevity Pay

The longevity pay is paid as if it were a bonus on the longevity date. In the Payroll Forecast tab of the spreadsheet, the longevity pay is not included in the base for calculation of the performance pay.

In the spreadsheet, the longevity pay is included in the simulation with a logic formula:

=OR(H3=10,H3=15,H3=20,H3=25,H3=30)*Q$40

Remember that the longevity pay is paid after each 5-year period of service, beginning with the 10-year service anniversary. The expression OR(x, y, . . . , n) returns the value "TRUE" ("1") if any of the statements within the commas are true. The value in cell Q40 is $5,000, the amount of longevity pay. By placing it in a separate cell, we can easily change it if the value changes per future policy.

Performance Pay

Performance pay is a 0.5% increment paid over the whole year, and half the work unit is eligible for it. The performance pay is calculated by totaling the wages and salaries and multiplying by the performance factor, as shown in Table 27.4:

TABLE 27.4
Performance Pay

	$ 1,059,835
Performance Factor	0.50% $ 5,299

Overtime

The remaining calculation to solve is overtime. As with base pay, overtime pay must be distributed before the step increase and after the increase. Overtime is paid at 150% of the hourly rate, so the periodic (monthly) rate is converted to an hourly rate (divided by 173.33) and then multiplied by 1.5. Under some circumstances, additional components of pay are at twice the hourly rate. This is not demonstrated here, but it would involve a similar table. To model this, all the technicians except the vacant position are designated as having 200 hours of overtime early in the fiscal year.

---

14. To facilitate this, we are using the pay scale (Table 1) as a lookup table in the spreadsheet. We have put the data from Table 1 in a named range labeled "MonthlyOld." Another named range, "MonthlyNew," has a pay scale that has been increased by 2.5% above "MonthlyOld." The procedure to determine the full cost of the pay to the new scale, after we finish calculating the cost of the old pay scale, is to simply use the alternate pay scale. The spreadsheet has features (VLOOKUP and INDIRECT) designed to assist with this, which are described in Appendix B of this book.

EXPECTED PAY

Pay is totaled in the Expected Pay column of Table 27.5. It is the sum of Wage & Salary, Longevity Pay, and Overtime. Performance Pay is treated as a separate record (row).

Forecasting Fringe Benefits

Benefits are calculated on the right-hand side of Table 27.5. Below the benefit calculations is a row labeled "Assumptions" in the left column. These assumptions are used in calculating the benefits.

The first use of assumptions in calculating benefits is to limit Expected Pay to the Social Security Wage for calculating FICA. The spreadsheet uses the MIN([Ce111, Ce112]) function. Cell 1 is the entry in the Expected Pay column. The anticipated FICA maximum for the year 2015 is entered in the Assumption row.[15] FICA is calculated as 6.2% of the Social Security Wage column. The Health Insurance column is a fixed amount.[16] Where the analyst knows whether an employee uses individual, family, or no coverage, this information can be used, with the caution that some of these decisions can change over time. The columns for "Medicare," "Other Legally Required," "Retirement," and "Unspecified" use ratios from the Assumptions row.[17]

The total of benefits is shown in the rightmost column of the table. Benefits are also calculated for the performance pay; however, no additional amount is required for health insurance.

Finalizing the Estimate

Below the calculation of benefits in Table 5, we see some final calculations for Total Pay, Filled Rate, Estimated Payroll, Retention, and Budget Estimate. Total Pay is the addition of Expected Pay and Total Benefits for all employees, including the performance factor. The reset of this calculation makes adjustments for turnover (using Filled Rate) and vacancy (using Retention), which are described below. The calculations multiply Total Pay by Filled Rate to get Estimated Payroll and multiply Estimated Payroll by Retention to get the final Budget Estimate.

Other Than Monthly Periods

This calculation is demonstrated with monthly pay periods because the use of spreadsheets to determine the next pay period after the anniversary date is much simpler with monthly periods.

Following is an expression that can be used to find the number of periods before the step increase with semimonthly data. This expression substitutes the labels M_{an} and D_{an} for the cell addresses for the individual employee's employment anniversary month and anniversary day.

$$= \mathrm{IF}(\mathrm{AND}(\mathrm{M}_{an} = 6, \mathrm{D}_{an} < 16), 23, 2 * (\mathit{IF}(\mathrm{M}_{an} > 5, \mathrm{M}_{an} - 6, \mathrm{M}_{an} + 6)) - 1 * (\mathrm{D}_{an} < 16))$$

The expressions for weekly and biweekly periods are developed for the number of periods after the step increase and are considerably more complex. For the periods before the step

15. The amount entered is the 2013 amount, inflated 2 years at 2.7% per year. In practice, one should use the best available data.

16. This amount is the amount demonstrated earlier, inflated to 2015 at 6.5% per year. In practice, one should use the best available data.

17. Actual organizations should use their actual categories with as much detail as possible, and they should possibly individualize some categories where there is differential participation.

TABLE 27.5

Hypothetical Payroll Data With Expected Pay

	Overtime								Benefits								
Employee	Rate Before Step	Rate After Step	Hours Before Step	Hours After Step	Current Step PT	New Step OT	OT	Expected Pay	Social Security Wage	FICA	Health Insurance	Medicare	Other Legally Required	Retirement	Unspecified	Total Benefits
Stone	$85.88	$85.88	0	0	0	0	0	$124,093	$119,923	$7,435	$9,680	$1,799	$683	$6,205	$496	$26,298
Delrico	$60.09	$61.59	0	0	0	0	0	$85,228	$85,228	$5,284	$9,680	$1,236	$469	$4,261	$341	$21,271
Shaft	$29.24	$29.24	0	0	0	0	0	$40,547	$40,547	$2,514	$9,680	$588	$223	$2,027	$162	$15,194
Vacant	$47.70	$48.89	0	0	0	0	0	$67,795	$67,795	$4,203	$9,680	$983	$373	$3,390	$271	$18,900
Mayes	$46.53	$47.70	200	0	$9,307	0	$9,307	$73,969	$73,969	$4,586	$9,680	$1,073	$407	$3,698	$296	$19,740
Othman	$48.89	$50.11	200	0	$9,778	0	$9,778	$79,268	$79,268	$4,915	$9,680	$1,149	$436	$3,963	$317	$20,460
Bierns	$63.13	$63.13	0	0	0	0	0	$87,537	$87,537	$5,427	$9,680	$1,269	$481	$4,377	$350	$21,585
Thorndike	$28.53	$29.24	0	0	0	0	0	$39,805	$39,805	$2,468	$9,680	$577	$219	$1,990	$159	$15,093
Hagar	$47.70	$48.89	200	0	$9,540	0	$9,540	$76,370	$76,370	$4,735	$9,680	$1,107	$420	$3,818	$305	$20,066
Kaplan	$48.89	$50.11	200	0	$9,778	0	$9,778	$78,279	$78,279	$4,853	$9,680	$1,135	$431	$3,914	$313	$20,326
Takei	$48.89	$50.11	200	0	$9,778	0	$9,778	$77,997	$77,997	$4,836	$9,680	$1,131	$429	$3,900	$312	$20,288
Perez	$61.59	$63.13	0	0	0	0	0	$86,114	$86,114	$5,339	$9,680	$1,249	$474	$4,306	$344	$21,391
Mead	$45.40	$46.53	200	0	$9,080	0	$9,080	$73,608	$73,608	$4,564	$9,680	$1,067	$405	$3,680	$294	$19,691
Wolff	$47.70	$48.89	200	0	$9,540	0	$9,540	$76,094	$76,094	$4,718	$9,680	$1,103	$419	$3,805	$304	$20,029
Wilson	$46.53	$47.70	200	0	$9,307	0	$9,307	$74,238	$74,238	$4,603	$9,680	$1,076	$408	$3,712	$297	$19,776
Performance								$5,299	$5,299	$329		$77	$29	$265	$21	$721
Assumptions:	173.3333	1.50							$119,923	6.20%	$9,680	1.45%	0.55%	5%	0.40%	
	Hours	OT Rate					Totals	$1,146,242		$70,808	$145,200	$16,621	$6,304	$57,312	$4,585	$300,830
															Total pay	$ 1,447,072
															Filled rate	0.932
														Estimated payroll		$ 1,348,456
														Retention		0.974
														Estimated payroll after retention		$ 1,313,340

increase, the number of periods after the step increase is subtracted from the total number of periods.

For biweekly, the expression is as follows:

$$=\text{IF (AND } (M_{an} = 6, D_{an} > 24), \text{Periods}_{total}, \text{IF } (M_{an} > 6, \text{TRUNC ((DATE } (\text{Year}_1, M_{an}, D_{an}) - \text{First}_1) / 7), \text{TRUNC ((DATE } (\text{Year}_2, M_{an}, D_{an}) - \text{First}_2) / 7)))$$

For weekly, it is as follows:

$$=\text{IF (AND } (M_{an} = 6, D_{an} > 24), \text{Periods}_{total}, \text{IF } (M_{an} > 6, \text{TRUNC ((DATE } (\text{Year}_1, M_{an}, D_{an}) - \text{First}_1) / 7), \text{TRUNC ((DATE } (\text{Year}_2, M_{an}, D_{an}) - \text{First}_2) / 7)))$$

The variables M_{an} and D_{an} are as with the semimonthly expression. Year_1 is the calendar year of the first day of the fiscal year. Year_2 is the calendar year in which January occurs. First_1 is the first day of the fiscal year. First_2 is the first day of the calendar year in which January occurs. Periods_{total} contains either 26 or 27 for biweekly and either 52 or 53 for weekly. These values are determined from calendar information.

Below, these expressions are shown in spreadsheet form, where the employee data are in row 3 and reference data are in various cells in rows 37, 39, and 42.

For biweekly, the expression is

=IF (AND (I3 = 6, J3 > (30 – 6)), \$N\$42, IF (I3 > 6, TRUNC ((DATE (\$M\$37, I3, J3) – \$Q\$37)/7), TRUNC ((DATE (\$M\$39, I3, J3) – \$Q\$39)/7)))

For weekly, it is

=IF (AND (I3 = 6, J3 > (30 – 6)), \$N\$42, IF (I3 > 6, TRUNC ((DATE (\$M\$37, I3, J3) – \$Q\$37)/7), TRUNC ((DATE (\$M\$39, I3, J3) – \$Q\$39)/7)))

These expressions have been expanded with spaces for readability, but they must be closed up (spaces removed) to be used in a spreadsheet.

Vacancy and Turnover

Now we move to vacancy and filled positions and to turnover and retention. These are two payroll effects resulting when employees resign or accept promotions.

Vacancy is the fraction of pay and benefits not spent because positions are empty for a period of time. Typically, a position that has been authorized but not funded (as when a position is planned to be filled for the first time later in the year) should not be counted as vacant. Position accounting is not as nuanced as financial accounting. When a funded position is vacant, it contributes to vacancy. The amount it contributes is the dollar-weighted amount of time (number of pay periods) it is vacant. The dollar weight is the total cost of the employee (pay and benefits) who resigns or otherwise vacates the position at the time the employee vacates the position divided by the average cost of all employees.

Consider the following: Marilyn O'Riley is earning \$5,540 per month and resigns to accept a position at another organization. Historically, she has received little in the way of overtime, so we will omit that. Her monthly benefits are \$1,560. The total is \$7,100.[18] The average pay and benefits

18. Shown as \$7,101 in Table 3, due to differences in rounding.

for all employees is $7,804. Her dollar-weighted contribution to the **vacancy rate** is .892 for each period her position is empty.[19]

This weighted value is multiplied by the number of periods her position is vacant within the year for which the rate is calculated and added to the total of weights for other vacant positions in that year.[20]

The total of all of the weights for vacant periods for all positions for the year is found and divided by the product of positions times pay periods. This number may be reduced for the number of pay periods for which positions were intentionally left vacant at the beginning of new projects:

$$\text{Vacancy} = \frac{\text{Weights}}{(\text{Positions} \times \text{Periods}) - (\text{New \& intentional} \times \text{Begin periods})}$$

The **filled rate** is the balance after removing vacancy. It is as follows:

$$\text{Filled} = 1 - \text{Vacancy}$$

The filled rate is multiplied by the total pay and benefits to refine the budget value.

We calculate weights rather than just count each position as a unit of 1 for each period it's vacant because we want to avoid substantial financial error. When employees at different pay grades have different vacancy rates, calculating an unweighted vacancy rate can overstate or understate the rate, resulting in reducing the funding by too much or too little.

Table 27.6 shows the calculation of vacancy rate for a 15-person organization. The monthly full cost (pay and benefits) for each employee is shown in the Cost column. These are divided by the average (at the foot of that column) to produce the weights in the next column. For positions that are vacant, the number of periods vacant in the fiscal year are determined and multiplied by the weights to produce the last column, which is totaled and divided by the total number of periods available. The total number of pay periods available (which is equal to the total value of the weights with this method) is the number of employees times pay periods in a year; for 15 employees paid semimonthly it is 15 times 25, or 360. Dividing total weights, 24.533, by 360 results in a vacancy rate of 6.8% and a filled rate of 93.2%.

Turnover and **retention** involve determining how much higher or lower the average pay of replacement employees is compared with the average pay of the positions they replace. For some organizations, such as police or fire departments, all or almost all employees are hired as entry-level trainees. Universities hire most of their faculty as assistant professors. Many organizations have two or three similar positions roughly reflecting new hire, midrank, and senior level. The average rank and pay of employees leaving positions is typically higher than the rank and pay of new hires. In professions where skills are scarce, the situation may be reversed, and long-employed workers may have fewer relevant skills that limit mobility and pay while new hires may have more options and an ability to demand greater pay than the people they replace. In either case, accounting for the payroll impact of replacement workers provides precision in calculating organizational funding. Turnover is the change. Retention is the balance, the amount to budget for.

To determine retention, annualize the costs of incumbents in positions that go vacant. In the Marilyn O'Riley example, annualize the monthly $7,101 to $85,200.[21] Next, annualize the

19. This method can be calculated by day rather than pay period for more precision. When using the pay period approach, if the position is vacant for about half the period, count it as half. If it is just a few days, don't count it. If it is more than three fourths, count it as the whole period.

20. Weights for positions still vacant at the beginning of the next year are counted for that subsequent year.

21. Annualized numbers are rounded in the text. More specific numbers are in the table.

TABLE 27.6
Weighted Vacancy Rate

Weighted Vacancy Rate										
Employee No	Hired	Position	Separated	Beginning of Period	Filled	End of Period	Cost	Weight	Pay Periods Vacant	Periods Weights
45123	9/15/2005	Supervisor					12,533	1.574		
45124	7/22/2013	Team Leader					8,486	1.066		
45125	3/1/2009	Office Assistant					4,645	0.583		
45126	4/1/2008	Technician	3/15/2014	7/1/2013		6/30/2014	7,101	0.892	7	6.241
45127	5/16/2014	Technician					7,809	0.981		
45128	6/13/2012	Technician	11/1/2013	7/1/2013	3/1/2014	6/30/2014	8,157	1.024	8	8.194
45129	5/7/1999	Team Leader					9,094	1.142		
45130	3/24/2004	Office Assistant					4,575	0.574		
45131	1/11/2006	Technician	5/1/2013	7/1/2013	12/12/2013	6/30/2014	8,043	1.010	10	10.099
45132	1/14/2002	Technician					8,224	1.033		
45133	3/1/1999	Technician					8,197	1.029		
45134	2/5/2001	Team Leader					8,967	1.126		
45135	6/17/2003	Technician					7,775	0.976		
45136	3/1/1999	Technician					8,017	1.007		
45137	3/15/1997	Technician					7,841	0.985		
	360	Total Pay Periods					7,964		25	24.533
									Vacant	6.8%
									Filled	93.2%

replacement salary. As an illustration, Marilyn is replaced by Robert Block at the first step of the pay grade. His annualized first-year salary and benefits total $81,200. Later in the year, another employee leaves at an annualized ending cost of $86,700 and is again replaced by an employee costing an annual $81,200. In a third case, the vacating employee's annualized cost is $97,900. Using the actual cost for positions that remained filled throughout the year and the two sets of annualized costs, one for the original incumbents and the other for the new incumbents, we calculate full-year costs for the entire payroll. We do not include vacancy. Retention is calculated as follows:

$$\text{Retention} = \frac{\$\text{Total}_{\text{new}}}{\$\text{Total}_{\text{original}}}$$

Turnover is calculated as follows:

$$\text{Turnover} = 1 - \text{Retention}$$

The retention rate is multiplied by the once-refined budget value (reflecting the filled rate) to get the final budget value.

Table 27.7 shows the calculation of retention and turnover. The last column

TABLE 27.7
Calculation of Retention and Turnover

Employee No.	Grade	Step	Annualized Current Pay & Benefits ($)	Vacancy	Annualized Cost With New Incumbents ($)
45123	19	5	150,391	0	150,391
45124	15	3	106,499	0	106,499
45125	5	5	55,741	0	55,741
45126	12	3	86,695	1	81,200
45127	12	2	93,709	0	93,709
45128	12	4	97,883	1	81,200
45129	15	5	109,122	0	109,122
45130	5	4	54,898	0	54,898
45131	12	3	96,515	1	81,200
45132	12	4	98,685	0	98,685
45133	12	4	98,365	0	98,365
45134	15	4	107,606	0	107,606
45135	12	1	93,299	0	93,299
45136	12	3	96,201	0	96,201
45137	12	2	94,091	0	94,091
			$1,439,701		$1,402,209
				Retention	97.4%
				Turnover	2.6%

substitutes the annualized cost of new employees for the annualized pay of prior employees who have been replaced in the last year. The total of that column is divided by the column containing the annualized cost of the original incumbents, resulting in the rates of 97.4% turnover and 2.6% retention shown in the bottom right corner.

Tracking and Forecasting

Once you have the vacancy and retention rates, make a record of them and track them over the years. With multiple annual rates, average each series. Generally, a useful vacancy or retention rate will be based on 5 years of data. If the rate exhibits a trend, collect many years of data and consider using forecasting methods.

Summary

This module reviews the steps for estimating the cost of pay and benefits using full computation of the payroll, using real or realistic data and accounting for factors that lead to adjustments. The example case demonstrates these steps with a monthly pay system, and expressions are shown for other pay cycles.

Assignments

For these assignments, use the supplied self-checking spreadsheets.

1. You have been assigned to help your IT department develop a budget request for their programmers. You have been given limited data in the spreadsheet *Budget Tools 2e Chapter 26, Payroll self-checking.xls* and will have to build a payroll simulation utilizing Excel. You have the position numbers, names, position titles, grades, and current steps of the employees and the employees' hire dates.

Create a payroll simulation for the budget submission for FY 2015. The fiscal year is July 1 to June 30.

(a) In column G, write a formula that determines the next step for each employee. There are only 5 steps in this pay grade. This number is already found in cell B27, where the top pay step is recorded. In this case, we can use the "minimum" function in Excel. The formula needs to assume the employee will move up one step during the year but not go higher than 5.

Use the formula =MIN((F2+1),$B$27)

Check the logic of the first returned data. Does it make sense? If so, copy that formula down for the rest of the employees. (For each step that follows, you will need to copy the formula down to the rest of the positions once you have ensured that the first row works.)

(b) In columns I and J, separate the values associated with the hire date respectively. Excel will do this automatically if you use the formula =MONTH(K5) in column H and use =DAY(K5) in column J.

(c) In column H, you need to calculate how many years the employee has with the team. This is accomplished in two steps. First, create a reference cell of the year from which you want to measure. In cell H3, enter "2015" because you are building the FY 2015 budget. Then to calculate the length of service in years, use the formula =$H$3-YEAR(K5). This formula will take 2015 (your reference cell) and subtract the hiring year for the employee. The value returned may appear in date format. Use the formatting tools to make it a number rather than a date.

(d) Assume a monthly pay period. Employees are moved to the next step in the first month following their anniversary. Now you will need to build the columns necessary to finish the simulation. Use column L to convert the months from column I into the corresponding fiscal year month (July = 1, August = 2, etc). Column L is labeled "Fiscal Year Month." In cell L3, use the formula using an IF statement: =IF(I5>7,I5–6,I5+6). A hire date of January should now be represented as "7," August as "2," etc. This provides us with the number of months that employees will be paid at their current steps.

(e) In column M, calculate how many months employees will be paid at their new step rates. In cell M3, use the formula =12-L5. The result is that columns L and M represent 12 months of pay.

(f) Next, calculate the base salary dollars associated with each person. In columns A and B, starting in Row 20, a table has been provided that shows the monthly wage for each step in this pay grade. Column N is labeled "Monthly Before-Step Pay." Use the lookup function to return the correct monthly wage for the step that the employee is in currently. The formula in cell N5 will look like this: =VLOOKUP(F5,$A$22:$B$27,2). Have Excel look at the current step, find that value in column A rows 22 through 27, and return the associated wage from column B (the second column). Format this column to show the numbers as dollars.

(g) You need to perform the same procedure for the "after" step. Copy all of the formulas in column N and paste them into column O. Because "$" is part of the formula, the new formulas will keep the right range, and copying one column to the right will move the reference cell to column G rather than column F.

(h) Calculate the total wages for the year by multiplying the value in column L by the wage in column N and adding it to the product of column M and column O. The formula looks like this for cell P3: =L5*N5+M5*05. The column is labeled "Expected Base Pay."

(i) The department provides an $8,000 one-time pay increase when an employee reaches the 10-, 15-, or 20-year anniversary. Column Q is labeled "Longevity Pay." Similar to what you did in step 3, create a reference cell for the $8,000 in cell Q3. Use the OR formula to include or not include the $8,000 based on the years of service in column H. The formula in Q5 will read as follows: OR(H5=10,H5=15,H5=20)*$Q$3. Format the column to show dollars.

(j) Calculate the total wages for the employee by adding columns P and Q in column R, which should be labeled "Total Annual Wages." The formula expression is simply =P5+Q5.

(k) We assume no overtime for this team.

(l) Now we will calculate benefits. The department has a fixed health and dental benefit rate of $9,400. We will use this number in a moment. The department is required to pay portions of retirement, FICA, and Medicare. Unemployment and workers' compensation are handled through the internal risk management department, so they will be omitted from this exercise.

(m) Column S is labeled “Retirement.” Create a reference cell above the label for the rate (make it a percentage). The current retirement rate is 12.5% of the total wages. Calculate this for each employee. The formula in cell S5 will be =R3*$S$3.

(n) Column T is labeled “FICA.” Create a reference cell above the label for the rate of 6.2%. There is also a maximum income associated with FICA. We need a second reference cell. Highlight and color cells R3 and T3 the same color, as you will use R1 as the second reference cell. Put $113,700 in R3. Now calculate the FICA amount using an IF statement like this: =IF(R5>$R$3,$R$3*$T$3,R5*$T$3).

(o) Column U is labeled “Medicare.” Create a reference cell above the label for the rate of 1.45%. Follow the same process in step 13 to calculate the Medicare amounts.

(p) Column V is labeled “Total Benefits.” Create a reference cell above the label for the fixed benefit rate of $9,400. Add the fixed rate to the retirement, FICA, and Medicare to get to the total benefits. The formula will look like this for cell V3: =$V$3+S5+T5+U5.

(q) Column W is labeled “Total.” Add the Total Annual Wages to Total Benefits.

(r) Create a Total row in row 12 and add all the rows of columns R, V, and W.

(s) If the result of step 18 is “#N/A,” you have not done something wrong. You need to fill in an estimation for the vacant position number 57222. Assume the new person will come in as step 1 on January 1, 2015. Now your values should fill in.

What does your budget request need to be to fund these positions for the upcoming fiscal year?

2. The director just came in after a meeting with the central budget and human resources offices. It seems they have just completed a study of similar positions in the marketplace, and all positions in this grade will be getting a 7% increase in addition to any changes in the step. The director needs you to revise your budget request figure immediately. Before you hang your head, you realize you just built a dynamic spreadsheet that will allow you to make this change quickly.

(a) You will begin by repeating the first assignment. You can do this rapidly by using the copy/paste feature. However, only the yellow-highlighted cells are accessible for copying and pasting in the self-checking spreadsheets.

(b) The Step Payroll information is located in shaded cells A22 through B27. In cell B19, enter the current step 1 pay rate of $3,500. In cell B20, enter 1.07 or 7%. Beginning in cell B23, enter the adjusted rate (3500 * 1.07). In cells B24 through B27, increase this by increments of 2.5% using the rate found in cell B21.

(c) As you are getting ready to email your director the new total, you see an email arrive from the benefits director indicating that the fixed rate will increase to $9,650 per employee and that the retirement rate has been increased to 12.75%. Make these changes in the appropriate reference cells.

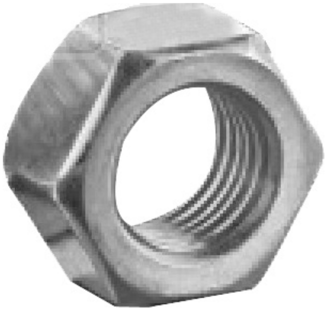
(d) What is your new total?

Additional Readings

Feldt, J. A., & Andersen, D. F. (1982). Attrition versus layoffs: How to estimate the costs of holding employees on payroll when savings are needed. *Public Administration Review, 42*(3), 278–282.

Lyons, N. R. (1977). An automatic data generating system for data base simulation and testing. *ACM SIGSIM Simulation Digest, 8*(4), 8–11.

Scott, D. F., Jr., Moore, L. J., Saint-Denis, A., Archer, E., & Taylor, B. W., III. (1979). Implementation of a cash budget simulator at Air Canada. *Financial Management, 8*(2), 46–52.



MODULE 28
Basic Forecasting Concepts

Learning Objectives:

- *Visualize and graph time series data*
- *Identify and correct outliers*
- *Decompose data*
- *Adjust complex data*
- *Identify three types of time series data that can be forecast*
- *Understand averages of time series data*
- *Make a moving average of time series data*
- *Calculate the following:*
  - *Error*
  - *Squared error*
  - *Mean squared error*
  - *Root mean squared error*
- *Calculate bias*
- *Select a simple forecast*

This module focuses on the skills and preparation you need before you begin to forecast. Some simple preforecasting techniques are shown.

Time Series Data

The techniques shown here and in the next module are appropriate for data that occur regularly in a time sequence. *Regularly* means monthly, quarterly, semiannually, or annually.[1] For the techniques to work well, it is important that data be consistently recorded with time segments. This

1. Weekly, daily, hourly, and so forth data can also be forecast using these techniques, but for issues related to budgeting, there is seldom a need for such forecasts.

means that data belonging with one month should not be recorded as belonging with another month. Such attribution issues may particularly arise with cash management and jurisdictions that use cash budgeting.

In general, these techniques work better with more data. While monthly data is often preferable, monthly data can involve a problem called "seasonality," which will be discussed in Module 30. For the remainder of this module, data will be assumed to be nonseasonal or **deseasonalized**.

Visualizing Data and Data Correction

The first step with your raw data is to graph them. The steps for creating a graph are shown in Appendix C of this book.

FIGURE 28.1
New York City: Government Charges for Services

M8 fx =O8+P8*$M$3

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			
8	1980	1,142		1,142.77	130.46	1,153	(11)	111.14	
9	1981	1,311		1,309.44	130.84	1,273	38	1,452.83	

Figure 28.1 shows a time trend for general government charges for services for New York City. As a forecaster, you should notice that the first three observations are followed by a sharp drop and that the overall series trends upward. Both of these types of information will be used in forecasting. Because this is a long series and the sharp drop comes very near the beginning, the likely corrective action for the sharp drop is to simply start the forecast series after the drop, in year 4 of the series.

FIGURE 28.2
New York City: Government Charges for Services

N8 fx =N7+$N$3*P8

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			
8	1980	1,142		1,142.77	130.46	1,153	(11)	111.14	
9	1981	1,311		1,309.44	130.84	1,273	38	1,452.83	
10	1982	1,415		1,416.13	130.59	1,440	(25)	646.55	

With Figure 28.2, we see that in the year 2005, the value is sharply outside the range of the rest of the series. This observation is not conveniently so long ago that we can just drop all of the series up to this point. But if we include it in the series, our estimate of forecast error, which will eventually help us select which forecast we make, will be overly influenced by this observation. If we have access to the people who manage the data generation process, we may investigate and find out what this unusual observation is about. Perhaps it is a correction and should actually be attributed to several years. Perhaps it reflects a settlement of some sort that would not normally recur. If we can find out why this observation is here, we may be able to reassign the value to other periods when it is no longer so extraordinary, or we may be able to exclude the unusual part altogether because we know, with confidence, that it is a nonrecurring, extraordinary event.

However, we may have to use a data adjustment technique. A well-known technique is known as "Windsorizing" (Armstrong, 1985), in which we first calculate the mean and standard deviation of the series. The Excel expressions for these are as follows:

=AVERAGE()

and

=STDEV()

Excel has several expressions for each of these functions, depending on what might be in the series, but for simple data, these expressions are effective. Within the parentheses is the range of the series, as shown in Figures 28.3 and 28.4.

FIGURE 28.3
Excel Average Formula

FIGURE 28.4
Excel Standard Deviation Formula

To **Windsorize**, we multiply the standard deviation by 3, add it to the mean, and substitute the result for the extraordinary value. This expression is demonstrated in Figures 28.5 and 28.6:

$$W = SD \times 3 + M$$

Figure 28.7 shows the revised observation as a dot superimposed on the original series. The resulting value is still sharply outside the typical value of the series. The goal of Windsorizing is to reduce the extraordinary value to one that is within statistically likely values, not to eliminate

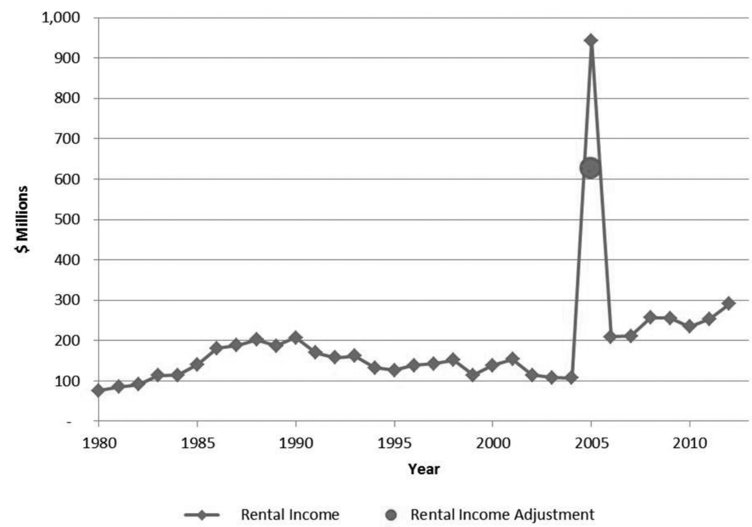
FIGURE 28.5
Windsorizing Formula

187	Average
146.47	Standard Deviation
3	Three
626.14	W=SD*3+M

FIGURE 28.6
Excel Windorizing Formula

F43 | =F41*F39+F37

FIGURE 28.7
New York City: Government Charges for Services (Windsorized)



its unusual effect altogether. We fully eliminate the unusual effect only if our investigation has allowed us to reassign the data or has shown with confidence the value of a nonrecurring event.

The other matters we see in Figures 28.2 and 28.7 include the fact that the series has had a long period of low to no trend and that after the unusual observation, the series adjusted upward but may continue to have a low- to no-trend condition. However, the direction of the low trend, if there is one, has changed. Knowing about this may lead us to be very cautious when it comes time to accept the results of a forecast model.

The goal of graphing data before moving forward is to learn what you can about the series and to observe patterns that may lead to investigating the data or making reasonable corrections. Correcting data simply because it is inconvenient is not reasonable.

Decomposition

TABLE 28.1
New York City: Water and Sewer Revenue, Decomposed

Year	Revenue per Gallon	Gallons per Person	Population	Water & Sewer Revenue ($)
1980	0.0004	78,061.96	7,071,639	233,834,623
1981	0.0005	77,877.30	7,077,279	261,758,831
1982	0.0006	67,419.71	7,088,350	264,582,975
1983	0.0006	70,568.59	7,150,150	319,227,133
1984	0.0007	72,393.82	7,198,277	344,837,391
1985	0.0007	73,930.77	7,232,780	378,209,286
1986	0.0009	66,500.18	7,276,928	445,588,116
1987	0.0009	67,790.31	7,292,432	437,755,375
1988	0.0008	72,425.40	7,289,880	434,646,524
1989	0.0010	74,055.44	7,313,757	545,632,243
1990	0.0011	70,060.46	7,322,564	571,376,190
1991	0.0011	71,146.33	7,304,481	596,377,769
1992	0.0012	74,764.87	7,304,895	644,123,974
1993	0.0014	68,345.23	7,329,079	709,121,849
1994	0.0014	65,980.49	7,570,458	717,845,663
1995	0.0015	64,923.08	7,633,040	738,561,386
1996	0.0015	63,036.86	7,697,182	730,963,206
1997	0.0016	60,942.56	7,773,443	775,318,258
1998	0.0019	55,993.00	7,858,259	822,800,164
1999	0.0017	56,159.55	7,947,660	777,651,962
2000	0.0018	56,316.69	8,018,546	801,255,118
2001	0.0019	56,150.11	8,063,137	842,524,943
2002	0.0020	53,684.84	8,072,000	857,906,603
2003	0.0020	51,374.60	8,068,073	846,351,747
2004	0.0022	49,631.03	8,043,366	884,744,982
2005	0.0022	50,218.21	8,013,368	899,324,043
2006	0.0024	50,563.64	7,993,906	989,545,265
2007	0.0027	48,675.62	8,013,775	1,063,873,430
2008	0.0029	50,530.19	8,068,195	1,202,190,064
2009	0.0032	48,607.87	8,131,574	1,283,505,261

Decomposition is the practice of breaking data up according to meaningful components to improve forecast accuracy. This can be performed with jurisdiction-wide data by separating different revenue sources. For example, one can forecast property tax separately from sales tax. However, there is more to decomposition than separating revenue sources. At a deeper level, data can be divided between different underlying causal processes. In Table 28.1, the New York City Water and Sewer Revenue (right-hand column) is decomposed into three parts: revenue per gallon, gallons per person, and population. Each of these elements can be forecast separately, using appropriate techniques. The total expenditure is recomposed by reversing the decomposition.[2] To decompose the data, the analyst needs appropriate source data. In this example, the source data are total gallons used, which was estimated from other data published on the New York City Open Data website (https://nycopendata.socrata.com/); total revenue from water and sewer, which is available from several New York City sources; and total population, which is constructed from data in the New York City Comprehensive Annual Financial Reports. Gallons per person is total gallons divided by population. Revenue per gallon is water and sewer revenue divided by total gallons. In this example, sewage gallons are not counted because usage information is not available.

Removing Inflation

Similar to decomposition is the removal of effects that are clearly known and understood. Module 7 discusses adjusting for inflation. For some series, such as the water and sewer rate, the

2. A more sophisticated approach would use regression modeling, but this approach is not demonstrated in this text.

underlying cost process may be what you are forecasting, and adjusting for inflation may be misleading. However, for other series (such as personal income tax), it is equally ineffective to forecast a series that contains both inflation and other sources of change. In Table 28.2, income inflation is removed using the Consumer Price Index. Because the analyst is interested in the future revenue, the analyst has **deflated** the older data rather than the current data.[3] To do this, the analyst set the current-year (most recent) observation of the CPI at 100%. The analyst then divided the current year's CPI by each previous year's CPI and placed the result beside each previous year. This results in an inflator adjustment factor. The analyst then multiplied the inflator adjustment factor by the actual personal income tax (PIT) received for each year, resulting in an adjusted PIT. As PIT is partly dependent on the population, the analyst can then decompose the data by dividing by total population.[4] To forecast these data, the analyst would forecast the PIT-per-person data and separately forecast (or acquire a forecast of) population and inflation. The whole forecast is made by reversing the order of calculations or by using the component information in a **regression** model.

TABLE 28.2
New York City: Personal Income Tax (PIT), Adjusted for Inflation

Year	CPI	Inflator Adjustment Factor	Personal Income Tax Millions ($)	Adjusted PIT Millions ($)	Population	PIT per Person
1980	82	273%	879.29	2,400.34	7,071,639	339.43
1981	91	247%	1,018.52	2,520.40	7,077,279	356.13
1982	97	233%	1,159.41	2,702.56	7,088,350	381.27
1983	100	226%	1,331.01	3,005.99	7,150,150	420.41
1984	104	216%	1,546.64	3,348.41	7,198,277	465.17
1985	108	209%	1,739.86	3,637.19	7,232,780	502.88
1986	110	205%	1,815.60	3,726.26	7,276,928	512.07
1987	114	198%	2,163.17	4,283.28	7,292,432	587.36
1988	118	190%	2,088.47	3,971.08	7,289,880	544.74
1989	124	181%	2,445.16	4,435.58	7,313,757	606.47
1990	131	172%	2,537.56	4,367.23	7,322,564	596.41
1991	136	165%	2,798.16	4,621.26	7,304,481	632.66
1992	140	160%	3,233.01	5,183.39	7,304,895	709.58
1993	145	156%	3,474.09	5,408.02	7,329,079	737.88
1994	148	152%	3,555.64	5,396.78	7,570,458	712.87
1995	152	148%	3,601.53	5,315.77	7,633,040	696.42
1996	157	143%	3,919.55	5,619.25	7,697,182	730.04
1997	161	140%	4,377.18	6,134.57	7,773,443	789.17
1998	163	138%	5,152.94	7,111.02	7,858,259	904.91
1999	167	135%	5,527.83	7,463.53	7,947,660	939.09
2000	172	131%	5,611.71	7,330.39	8,018,546	914.18
2001	177	127%	6,164.52	7,829.70	8,063,137	971.05
2002	180	125%	5,005.61	6,258.79	8,072,000	775.37
2003	184	122%	5,029.75	6,148.84	8,068,073	762.12
2004	189	119%	6,068.48	7,226.24	8,043,366	898.41
2005	195	115%	7,200.06	8,292.75	8,013,368	1034.86
2006	202	112%	8,025.81	8,954.95	7,993,906	1120.22
2007	207	108%	8,647.78	9,381.71	8,013,775	1170.70
2008	215	104%	9,927.97	10,372.30	8,068,195	1285.58
2009	215	105%	7,657.18	8,028.44	8,131,574	987.32
2010	218	103%	7,592.66	7,832.32	8,186,443	956.74
2011	225	100%	8,165.97	8,165.97	8,244,910	990.43

Treating Data

Data that occur in a time series are sometimes treated as moving into the future with three types of units.[5]

Level

The first sort of unit is the whole observation (after the sorts of adjustments described in prior sections). We have already seen graphs of such data in Figures 28.1, 28.2, and 28.7. These are the whole data; sometimes, an average of some or all of the whole data in time order, a forecast of the whole data, or

3. This practice is also called **deflating**, but it is called inflating here to agree with a common understanding of what is happening to the older data.

4. If working population data or income-receiving population data are available, that would be better for decomposition; however, forecasts of that data might be less reliable.

5. There can be more, but they are not commonly used in forecasting.

simply a collection of the whole data arranged in a series is called the **level** of the data. *Level* refers to how far the data are from zero. For convenience, this module and the following one will refer to the whole data as x_t, meaning observation x at time period t. If calling attention to some estimate of level, it is labeled L or L_t, which is used when summarizing more than one observation. Use of subscripts implies that there is more than one L or x, and the reference to these subscripts as time periods implies that observations are organized serially, according to a regularly occurring data-generating event, such as monthly tax collection.

Trend

The second sort of unit is the **trend**, which is the change in the data from one period to the next. It is found by taking the later observation, labeled x_{t+1} below, and subtracting the earlier observation, x_t:

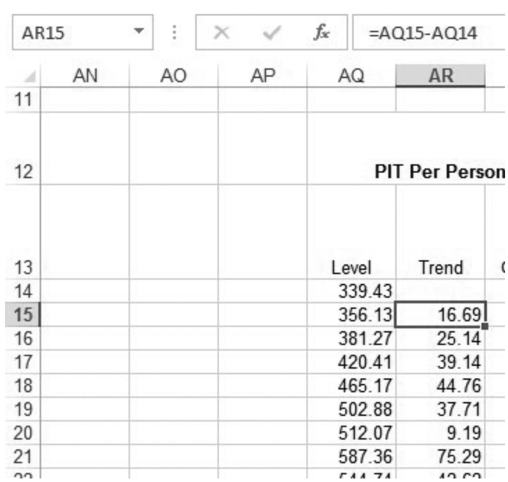
$$s_{t+1} = x_{t+1} - x_t$$

This expression says that s (trend, also called **slope**) at time t +1 is found by taking x (the observation) at time unit +1 and subtracting the observation at the immediately preceding time unit (t). This is one unit of trend.[6] When referring to a summary of more than one unit, trend is labeled S or S_t. A series will have one fewer trend unit than whole units. Figure 28.8 demonstrates the calculation of trend using Excel.

Growth

The third sort of unit is **growth**, which is popularly called percent growth or percent change. Sometimes it is also, incorrectly, called trend. Growth is found by dividing trend by the earlier of the two units in the trend expression:

FIGURE 28.8
Calculation of Trend Using Excel

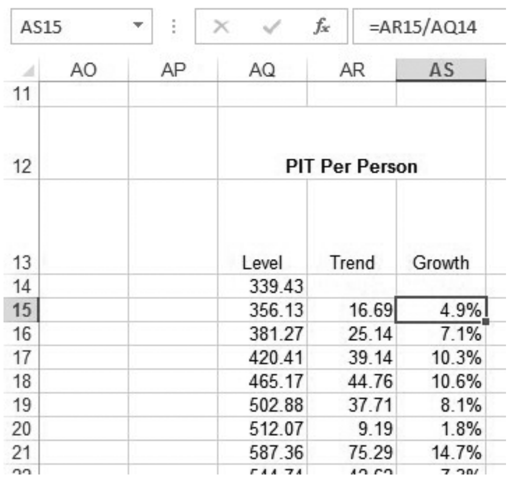


6. In Module 30, we refer to calculating trend units as *differencing*.

$$g_{t+1} = \frac{T_{t+1}}{x_t}$$

What this expression says is that g (growth) at time $t + 1$ is found by taking trend at time $t + 1$ and dividing it by the observation at time t. This is one unit of growth. When referring to a summary of more than one unit of growth, growth is labeled G or G_t. Growth is sometimes expressed as **percent growth**. Manual calculation of percent simply moves the decimal two spaces to the left. However, this requires the opposite decimal movement when doing math with percentages. Spreadsheets simplify this by using a **percent format** such that the actual data are not changed. Figure 28.9 demonstrates calculation of growth using Excel.

FIGURE 28.9
Calculation of Growth Using Excel



AS15 | =AR15/AQ14

	AO	AP	AQ	AR	AS
11					
12			PIT Per Person		
13			Level	Trend	Growth
14			339.43		
15			356.13	16.69	4.9%
16			381.27	25.14	7.1%
17			420.41	39.14	10.3%
18			465.17	44.76	10.6%
19			502.88	37.71	8.1%
20			512.07	9.19	1.8%
21			587.36	75.29	14.7%

Forecasting growth is incidentally popularized because it is the sort of thing that gets noticed in the national arena with respect to the entire economy. This does not make forecasting growth particularly useful in more micro-forecasting environments. This book does not provide recommendations with respect to forecasting growth.

Averages

Sometimes we reserve terms such as *level*, *trend*, and *growth* for estimates made from more than one observation. Here we are using capital letter symbols to refer to summaries or estimates made from more than one unit, while using lowercase symbols to refer to individual units. The following three expressions describe the simplest summary of level, trend, and growth:

$$L = \frac{x_1 + x_2 + \ldots + x_N}{N}$$

$$S = \frac{s_2 + s_3 + \ldots + s_N}{N - 1}$$

$$G = \frac{g_2 + g_3 + \ldots + g_N}{N - 1}$$

All of the elements of these expressions except N are as previously defined. N refers to the number of whole observations. The subscripts are linked to the whole series, so trend and growth begin at the second observation. These expressions indicate that one should add up all the data and divide by the count. Figure 28.3 demonstrated using the average function in Excel.

Averages are summaries of data. They provide interesting information, but they are not forecasts.

Moving Average of Level, Nontrending Data

Inexperienced analysts may be tempted to average a few observations and treat the average as a forecast. Without a better tool, an average seems like a good guess. When the next time to make a forecast comes around, they search for the data again and make a new average. Because the oldest data is always the hardest to find, what they are doing is making an unsystematic moving average. That is, they are making an average of a few observations, dropping the oldest when taking up the newest. As a first step toward forecasting, it is better to do this systematically.

Moving averages are like averages, but they are restricted to shorter periods than the entire length of the series. A systematic **moving average** uses the same number of periods for each calculation. (When we move to trending data, we will see that it is best to use an odd number of observations, unless you are using a moving average of quarterly or monthly data, where 4, 8, 12, or 24 periods reduces the influence of seasonality.) For annual data, one might use 3, 5, or 7 periods if there is a concern that level frequently changes, or one might use a larger number if seeking some protection from occasional unusual observations.

Figure 28.10 demonstrates the sort of data that might be forecast with a medium-length moving average, perhaps of seven periods. As we are using this simple technique and we know that something odd happened before 1990, we exclude those data. Figure 28.11 shows a seven-period

FIGURE 28.10
Forfeitures Data

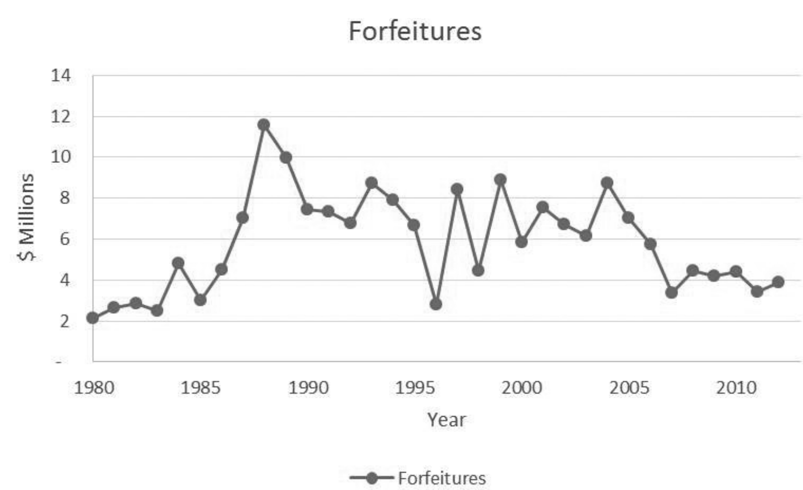


FIGURE 28.11
Forfeitures Data With 7-Period Moving-Average Forecast

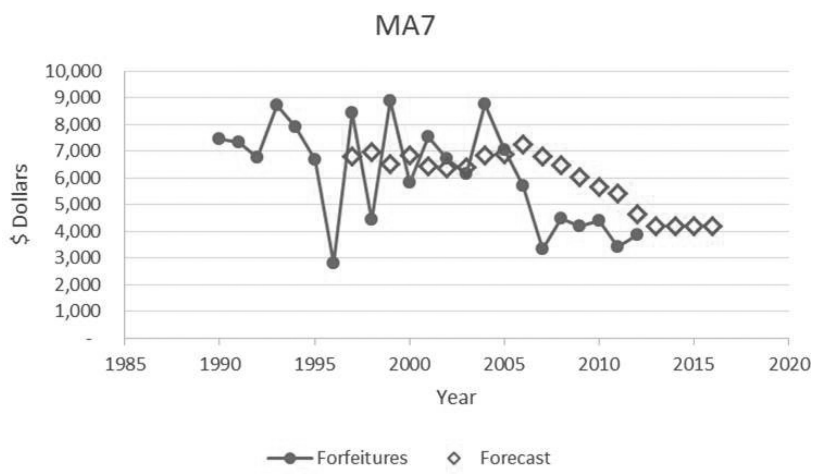


FIGURE 28.12
Forfeitures Data With 7-Period Moving-Average Forecast Using Excel

T19 =AVERAGE(S13:S19)

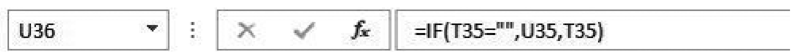
	R	S	T	U	V	W
12						
13	1990	7,468				
14	1991	7,356				
15	1992	6,771				
16	1993	8,738				
17	1994	7,907				
18	1995	6,675				
19	1996	2,816	6,819			
20	1997	8,425	6,955	6,819		
21	1998	4,454	6,541	6,955		
22	1999	8,888	6,843	6,541		
23	2000	5,830	6,428	6,843		
24	2001	7,522	6,373	6,428		
25	2002	6,727	6,380	6,373		
26	2003	6,149	6,856	6,380		
27	2004	8,757	6,904	6,856		
28	2005	7,035	7,273	6,904		
29	2006	5,720	6,820	7,273		
30	2007	3,355	6,466	6,820		
31	2008	4,477	6,031	6,466		
32	2009	4,182	5,668	6,031		
33	2010	4,397	5,418	5,668		
34	2011	3,431	4,657	5,418		
35	2012	3,885	4,207	4,657		
36	2013			4,207		
37	2014			4,207		
38	2015			4,207		
39	2016			4,207		
40						
41						

moving-average forecast. This forecast is relatively effective through 2005 or 2006, and then an event leads to ineffective forecasting. So, perhaps a shorter moving average would be more desirable. However, a shorter moving average will be more volatile (meaning it will move up and down more).

Figure 28.12 demonstrates calculating a seven-period moving average using Excel. Although the moving average is placed in 1996 in the spreadsheet, it is “centered” in 1993; that matters if the analyst uses moving averages for trending data. However, the analyst can know only a rough estimate of 1996 data in time to make a forecast for 1997, as shown in Figure 28.12. The analyst should be careful to update recent historical data to keep the moving average reliable. Figure 28.12 shows the process of dropping down one period to make a forecast.

The forecast remains constant over future years. The expression to achieve this is shown in the formula bar in Figure 28.13.

FIGURE 28.13
Calculating a Moving-Average Forecast Using Excel



Calculating a moving average involves calculating an average of a limited number of periods and offsetting it to the right location. This does not require an especially sophisticated use of spreadsheet functions. However, Excel does have a Data Analysis ToolPak function to produce a moving average. This function is demonstrated in Appendix A to this module.

Moving Average of Level and Trend, Trending Data

Some analysts would like to use moving average with trending data. You can, but doing so requires two moving averages. You need one moving average for the trend and another moving average of the level to find the “takeoff” point for the trend. Figure 28.14 shows trending data that might be forecast with this approach

FIGURE 28.14
General Sales

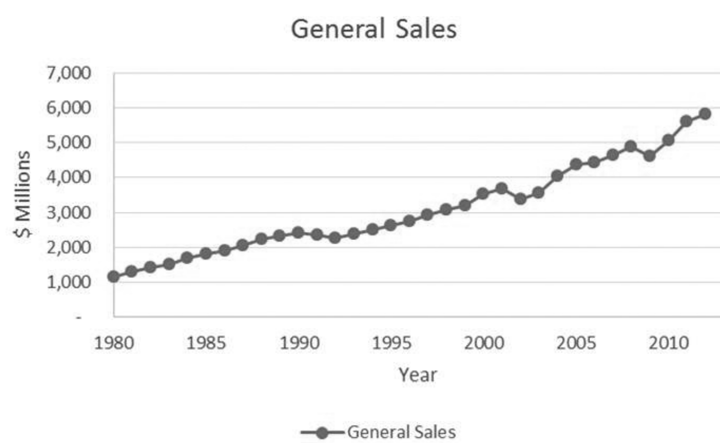


FIGURE 28.15
General Sales Versus Forecast

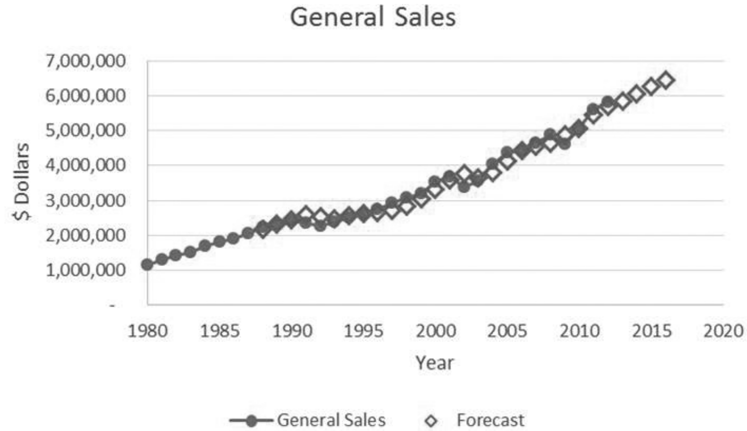


Figure 28.15 shows a forecast using a seven-period moving average for level and trend with these data. The forecast is made using the spreadsheet table shown in Figure 28.16. The seven-period moving averages are offset by one period (set on the row after the seventh period) to help the analyst keep track of the math. The formula bar shown at the top of the figure shows the calculation of the moving average for level. It contains an IF expression to provide a substitute level value when the actual data expire. This is important because forecasts go into periods where there are no data. For level, the substitute value is the previous moving average plus one unit of trend.

FIGURE 28.16
Calculating General Sales Forecast Using Excel

U36 =IF(S35="",U35+V35,AVERAGE(S29:S35))

	R	S	T	U	V	W	X
1						4	
2	Year	General Sales	Trend	Offset Level MA7P	Off Trend MA7P	Forecast	
3	1980	1,142,246					
4	1981	1,311,348	169,103				
5	1982	1,414,855	103,507				
6	1983	1,514,819	99,963				
7	1984	1,686,320	171,501				
8	1985	1,827,791	141,470				
9	1986	1,908,646	80,856				
10	1987	2,044,386	135,740				
11	1988	2,222,942	178,556	1,672,595	128,877	2,188,104	
12	1989	2,329,942	107,000	1,802,823	130,228	2,323,734	
13	1990	2,431,219	101,277	1,933,549	130,727	2,456,456	
14	1991	2,353,759	(77,460)	2,064,464	130,914	2,588,121	
15	1992	2,277,822	(75,937)	2,159,812	95,348	2,541,206	
16	1993	2,402,243	124,421	2,224,102	64,290	2,481,263	
17	1994	2,503,646	101,403	2,294,616	70,514	2,576,671	
18	1995	2,620,932	117,286	2,360,225	65,609	2,622,659	
19	1996	2,742,206	121,274	2,417,081	56,856	2,644,504	
20	1997	2,937,083	194,877	2,475,975	58,895	2,711,555	
21	1998	3,069,378	132,295	2,548,242	72,266	2,837,307	
22	1999	3,204,138	134,760	2,650,473	102,231	3,059,398	
23	2000	3,525,610	321,471	2,782,804	132,331	3,312,128	
24	2001	3,678,734	153,125	2,943,285	160,481	3,585,209	
25	2002	3,373,354	(305,380)	3,111,155	167,870	3,782,634	
26	2003	3,550,673	177,319	3,218,644	107,489	3,648,599	
27	2004	4,042,486	491,813	3,334,139	115,495	3,796,120	
28	2005	4,375,119	332,633	3,492,054	157,915	4,123,712	
29	2006	4,439,362	64,243	3,678,588	186,534	4,424,726	
30	2007	4,644,539	205,177	3,855,048	176,461	4,560,891	
31	2008	4,890,738	246,199	4,014,896	159,847	4,654,284	
32	2009	4,614,926	(275,812)	4,188,039	173,143	4,880,612	
33	2010	5,076,371	461,445	4,365,406	177,367	5,074,876	
34	2011	5,610,243	533,872	4,583,363	217,957	5,455,190	
35	2012	5,826,046	215,803	4,807,328	223,965	5,703,189	
36	2013			5,014,604	207,275	5,843,705	
37	2014			5,221,879	207,275	6,050,980	
38	2015			5,429,154	207,275	6,258,256	
39	2016			5,636,430	207,275	6,465,531	
40							

Figure 28.17 shows the calculation of the seven-period moving average of the trend. As with level, there is an IF statement to provide a substitute value when there is no data for trend. For trend, the substitute value is the previous value of trend.

FIGURE 28.17
Excel Formula for Calculating Seven-Period Moving Average of Trend

Figure 28.18 shows the calculation of the forecast. The forecast is the value of the level moving average plus 4 times the trend moving average. It uses 4 times the trend moving average because the level moving average is centered three periods prior to the current period and one more period is needed to get into the future. In the spreadsheet, the number 4 is in cell W1. It is important to put parameters like this into separate cells, because the number will change as the length of the moving average changes. The correct number is one half the length of the level moving average plus 1. It is a good idea to keep the trend and level moving averages the same length.

FIGURE 28.18
Excel Formula for Calculating Forecast

Error

Although the moving average of level and the moving average of level and trend have been labeled "forecasts," they have very limited power. They are an improvement over using guesses and unsystematic averages, but they can be improved more by calculating errors. The point of calculating errors is to find out how well the forecast performed. In more sophisticated forecasting, it can also provide information about the range of likely point values in the future period. It is extremely rare that a forecast gets the future value exactly right, so all forecasts are wrong by some amount. Knowing how much the error is can help one select a preferred forecast model from a collection of several options.

Error is calculated according to this expression:

$$e_t = x_t - F_t$$

This says that the error, e, is found by subtracting the forecast, F, from the actual, x. All of these have a subscript t, indicating that there is an error for each instance where there is both a forecast and an actual. To evaluate forecasts, we also need the squared error, which is calculated as follows:

$$e_t^2 = e_t \times e_t$$

The squared error is averaged, and this is called the **mean squared error** (**MSE**). To make the number easier to use, one finds the **root mean squared error (RMSE).** The RMSE is very similar to the common standard deviation:

$$MSE = \frac{e_1^2 + e_2^2 + \ldots + e_n^2}{n}$$

$$RMSE = \sqrt{MSE}$$

The first of these expressions says the MSE is the sum of the errors divided by the count. The small n is used because there may not be an error for every observation. The second expression simply says that the RMSE is the square root of the MSE. Calculation of these is demonstrated in Figures 28.19–28.22. The formula bar at the top of Figure 28.19 shows the calculation of error.[7]

Figure 28.20 shows squaring the error in Excel.

FIGURE 28.19

Calculating the Error Using Excel

V20 | fx | =S20-U20

	S	T	U	V	W	X
10	Forfeitures	MA7P	Forecast	Error	Error Squared	
11				(575)	3,020,399	Mean
12					1,738	Root
13	7,468					
14	7,356					
15	6,771					
16	8,738					
17	7,907					
18	6,675					
19	2,816	6,819				
20	8,425	6,955	6,819	1,606	2,580,312	

FIGURE 28.20

Calculating the Squared Error Using Excel

W20 | fx | =V20^2

Figure 28.21, which is from near top of the table, shows the calculation of the average of the squared errors.

7. The data in Figure 28.19, also used in Figure 28.12, have been adjusted by dividing by 1,000 to make this example easier to use. There is no mathematical consequence of this adjustment.

FIGURE 28.21
Calculating the Mean Squared Error Using Excel

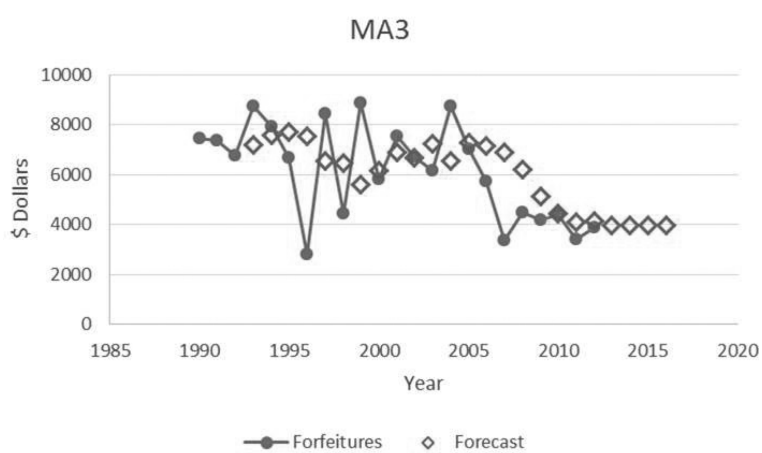
Figure 28.22 shows finding the square root of the MSE and placing this value in the cell below the MSE.

FIGURE 28.22
Calculating the Square Root of the Mean Squared Error Using Excel

The square root is the number that the analyst uses to compare forecasts. A smaller square root implies that over a large number of observations, the **one-period-ahead forecast** was closer to correct. Squared values penalize large errors more than small errors. The reason this is labeled "one-period-ahead forecasts" is that the formulas are written to update the forecast with each new observation, so the error for each observation reflects the forecast made in the previous period.

The result of the error evaluation is RMSE = 1,738. Earlier, while examining the graph in Figure 28.11, we hypothesized that a shorter moving average might have smaller errors but

FIGURE 28.23
Forefeitures Versus Forecast: Three-Period Moving-Average Forecast Based on Error Analysis



might also be more volatile. Two alternate moving averages (MA) were considered: MA = 5 and MA = 3. The results are $RMSE_{MA=5}$ = 1,676 and $RMSE_{MA=3}$ = 1,663. Figure 28.23 shows the three-period moving average that would be selected based on the error analysis.

Bias

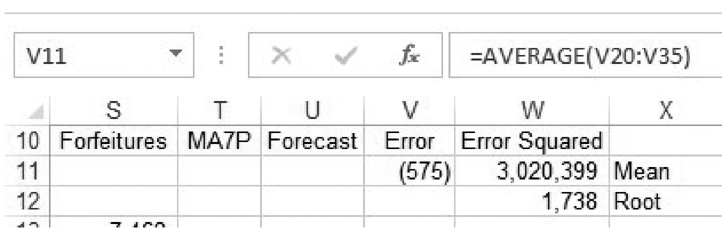
For an unbiased forecast, the measure of central tendency (the forecast) should produce an average error of zero. The **mean error (ME)** is calculated not from the squared error but from the initial calculation of the error, column V in Figure 28.19. The equation for the bias measure is as follows:

$$ME = \frac{e_1 + e_2 + \ldots + e_n}{n}$$

Because all forecasts of future values violate a basic rule of statistics, there is a price to pay, and that price will be found in bias. That is, forecasts are biased. (Some methods appear unbiased because they do not include predicted future values when calculating the mean error, but if these values are systematically included, they will also be biased.) The forecaster's objective is not to eliminate bias, as this cannot be done. Instead, it is to reduce bias without excessively reducing accuracy—that is, by increasing MSE. Sometimes these two objectives—reducing MSE and reducing ME—work against each other. There is no optimal solution. The analyst must become sensitive to the comparative issues and make a reasoned judgment.

Figure 28.24 demonstrates including the bias calculation with the forecast previously demonstrated in Figure 28.19. It shows using Excel's simple averaging formula.

FIGURE 28.24
Calculating the Bias Using Excel


V11 =AVERAGE(V20:V35)

	S	T	U	V	W	X
10	Forfeitures	MA7P	Forecast	Error	Error Squared	
11				(575)	3,020,399	Mean
12					1,738	Root

When this value is included with the three-, five-, and seven-period moving averages, the results (with the RMSE also shown) are as follows:

Period	ME	RMSE
MA3	(266)	1663
MA5	(370)	1676
MA7	(575)	1738

The parenthetical numbers are negative. With these data, both the error measure and the bias estimate support using the three-period model.

Although not demonstrated, for the trending data, the error values are ME = 28 and RMSE = 179. We will return to this information in the next module.

Summary

This module has introduced preliminary forecasting concepts, including examining graphs, correcting outliers, decomposing data, adjusting for known factors that either do not require forecasting or should be separately forecast, determining the type of data to be forecast, calculating the simplest data models using moving averages, and using errors in forecast evaluation. In the next module, we examine another simple forecasting technique that is known to be more effective than the moving average.

Assignments

1. The State of Arizona has asked you to examine tax burdens for residents in preparation for the next forecasting effort. Use the data in Table 28.3.

(a) Prepare a graph showing the total state and local per capita taxes paid per year. Identify any outliers.

(b) Calculate the level, trend, and growth, and their averages.

TABLE 28.3
State of Arizona: State and Local Per Capital Taxes

Year	Total State and Local Per Capita Taxes Paid ($)	Year	Total State and Local Per Capita Taxes Paid ($)
1977	729.10	1994	2017.41
1978	792.06	1995	2105.29
1979	871.42	1996	2133.51
1980	928.18	1997	2147.02
1981	960.51	1998	2230.04
1982	1014.01	1999	2366.24
1983	1058.44	2000	2498.91
1984	1226.99	2001	2553.06
1985	1372.09	2002	2592.34
1986	1470.50	2003	2679.79
1987	1559.85	2004	2787.50
1988	1631.95	2005	3128.02
1989	1718.13	2006	3368.10
1990	1804.99	2007	3750.00
1991	1870.55	2008	3582.34
1992	1878.01	2009	3202.74
1993	1951.56	2010	3006.35

2. In the Arizona tax information in assignment 1, a typographical error in the data has been discovered. The tax per capita in 2006 was actually $6,368.10 rather than $3,368.10.

(a) Prepare a graph showing the total state and local per capita taxes paid per year. Identify any outliers.

(b) Using the Windsorizing technique, adjust the data and prepare a graph showing the original information and the adjusted information.

3. Northland is developing a forecast for its special revenue fund that does not conform to a trend. Use the data in Table 28.4 to calculate a 5-year moving average and then use that result to project revenues for this fund for the next 3 years.

TABLE 28.4
Northland: Special Revenue Fund

1996	$53,421,417	2005	$49,055,130
1997	$64,600,858	2006	$124,202,747
1998	$67,053,747	2007	$86,126,170
1999	$24,316,124	2008	$27,862,519
2000	$66,110,642	2009	$85,704,726
2001	$93,389,137	2010	$81,483,730
2002	$49,686,023	2011	$45,459,605
2003	$20,706,410	2012	$22,672,732
2004	$81,238,032	2013	$58,226,034

4. Prepare a memo to the budget director of Northland providing a brief explanation of the revenue forecasts developed for FY 2014 to FY 2016. Based on the results of assignment 3, provide a recommendation.

5. River County receives revenue through property taxes. The budget director has asked you to build a forecast for the next 3 years. Use the data in Table 28.5.

(a) Prepare a forecast using a 7-, 5-, and 3-year offset moving average of level and trend.

(b) Utilize an analysis of errors to determine which forecast provides the most accurate information.

TABLE 28.5
River County: Property Tax Revenue

Fiscal Year	Property Tax Levy ($000)	Fiscal Year	Property Tax Levy ($000)
1992	34,762.30	2003	45,322.70
1993	32,354.05	2004	44,165.63
1994	30,971.67	2005	50,550.37
1995	31,010.82	2006	62,733.41
1996	32,382.92	2007	67,096.62
1997	33,211.58	2008	70,422.87
1998	36,407.76	2009	74,674.33
1999	38,930.54	2010	74,996.80
2000	43,246.20	2011	68,019.59
2001	48,342.35	2012	62,401.17
2002	45,042.55	2013	54,584.58
		2014	39,842.99

Note: These figures are rounded to the nearest cent.

6. Using a jurisdiction of your choosing, select a revenue stream that has more than one component. (Hint: The budget document usually has a discussion of the jurisdiction's revenue sources.) For one stream of revenue, graph 5 years of history as well as the projected value for that year's budget. Also, decompose the revenue stream into its various parts over the 5 years.

Reference

Armstrong, J. S. (1985). *Long-range forecasting: From crystal ball to computer* (2nd ed.). New York, NY: Wiley.

Additional Readings

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: A handbook for researchers and practitioners.* Boston, MA: Kluwer Academic.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2010). *Forecasting: Methods and applications* (3rd ed.). New York, NY: Wiley.

Williams, D. W. (2008a). Forecasting methods for serial data. In G. Miller & K. Yang (Eds.), *Handbook of research methods in public administration* (2nd ed., pp. 595–665). Boca Raton, FL CRC Press/Taylor & Francis.

Williams, D. W. (2008b). Preparing data for forecasting. In J. Sun & T. D. Lynch (Eds.), *Government budget forecasting: Theory and practice* (pp. 345–376). Boca Raton, FL: CRC Press.

Appendix A: Calculating a Moving Average Using Data Analysis ToolPak

In Figure 28A.1, we have series of annualized data in an Excel spreadsheet.

FIGURE 28A.1

AR16

	AP	AQ	AR
10	Year	Forfeitures	MA3 FC
11			
12			
13	1990	7,468	
14	1991	7,356	
15	1992	6,771	
16	1993	8,738	
17	1994	7,907	
18	1995	6,675	
19	1996	2,816	
20	1997	8,425	
21	1998	4,454	
22	1999	8,888	
23	2000	5,830	
24	2001	7,522	
25	2002	6,727	
26	2003	6,149	
27	2004	8,757	
28	2005	7,035	
29	2006	5,720	
30	2007	3,355	
31	2008	4,477	
32	2009	4,182	
33	2010	4,397	
34	2011	3,431	
35	2012	3,885	
36	2013		
37	2014		
38	2015		
39	2016		
40			
41			

Figure 28A.2 shows the Excel Ribbon selected for the Data tab. The Data Analysis icon is indicated.

FIGURE 28A.2

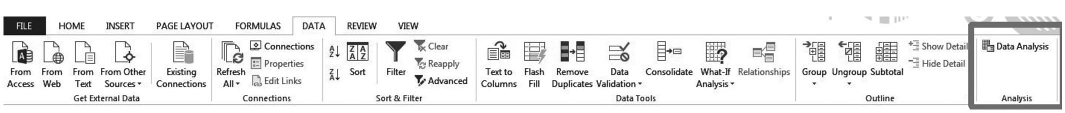


Figure 28A.3 shows the dialog that appears after clicking the Data Analysis icon. Find the Moving Average option and click OK.

FIGURE 28A.3

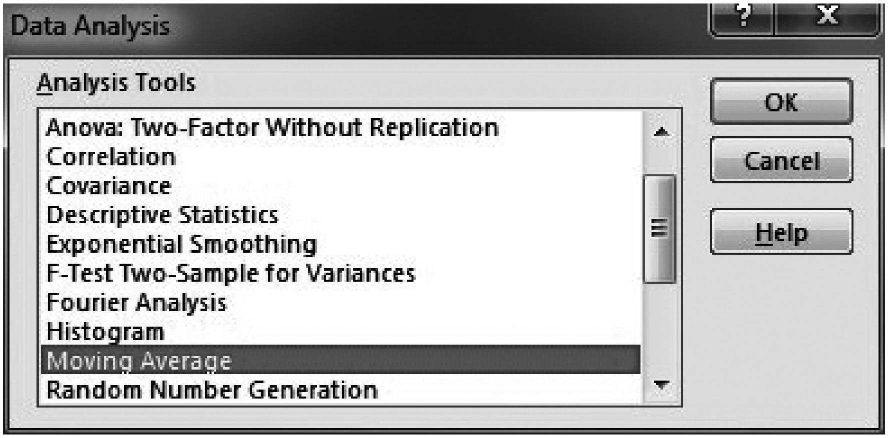


Figure 28A.4 shows the main Moving Average dialog. Click on the indicated icon.

FIGURE 28A.4

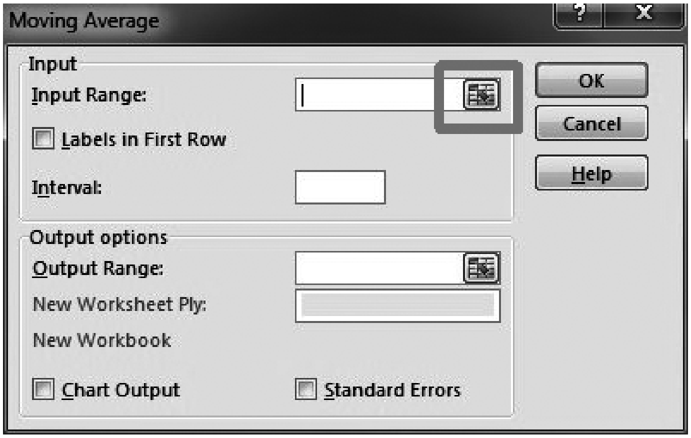
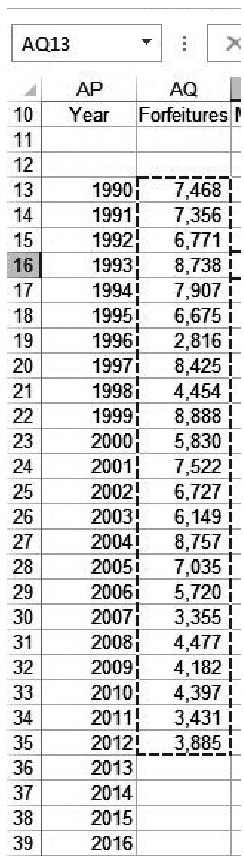


Figure 28A.5 shows the Input Range dialog. While this window is active, you can browse through the spreadsheet and find your input range.

FIGURE 28A.5

Figure 28A.6 shows your input range. The dotted lines indicate that input range AQ13 through AQ35 is highlighted.

FIGURE 28A.6


AQ13

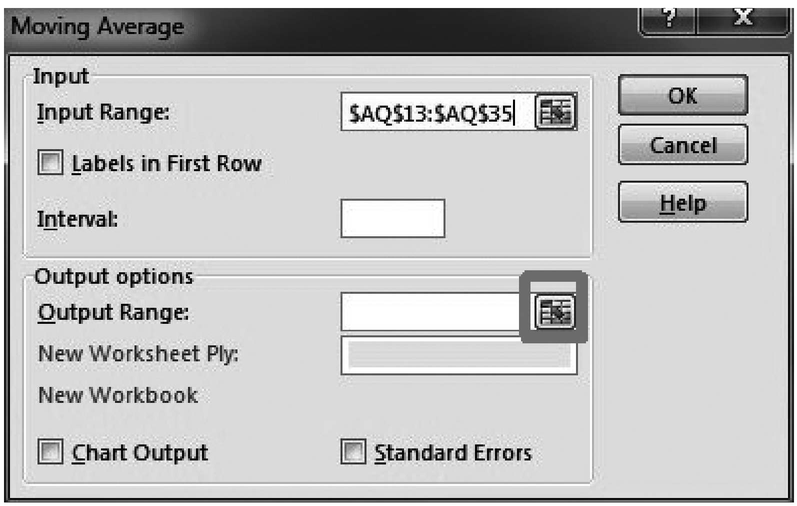
	AP	AQ
10	Year	Forfeitures
11		
12		
13	1990	7,468
14	1991	7,356
15	1992	6,771
16	1993	8,738
17	1994	7,907
18	1995	6,675
19	1996	2,816
20	1997	8,425
21	1998	4,454
22	1999	8,888
23	2000	5,830
24	2001	7,522
25	2002	6,727
26	2003	6,149
27	2004	8,757
28	2005	7,035
29	2006	5,720
30	2007	3,355
31	2008	4,477
32	2009	4,182
33	2010	4,397
34	2011	3,431
35	2012	3,885
36	2013	
37	2014	
38	2015	
39	2016	

Figure 28A.7 shows the highlighted range has populated the Input Range dialog. Click on the indicated icon.

FIGURE 28A.7

Figure 28A.8 shows the return to the main Moving Average dialog. Select the indicated Output Range icon.

FIGURE 28A.8



From the Figure 28A.9 Output Range dialog, search the spreadsheet again.

FIGURE 28A.9

Figure 28A.10 shows the selection of the cell in the column beside the data that is one row below the top of the data. The MA feature automatically offsets for the interval length, but we need to offset one more period to get to the forecast.

FIGURE 28A.10

AR16

	AP	AQ	AR
10	Year	Forfeitures	MA3 FC
11			
12			
13	1990	7,468	
14	1991	7,356	
15	1992	6,771	
16	1993	8,738	
17	1994	7,907	
18	1995	6,675	
19	1996	2,816	
20	1997	8,425	
21	1998	4,454	
22	1999	8,888	
23	2000	5,830	
24	2001	7,522	
25	2002	6,727	
26	2003	6,149	
27	2004	8,757	
28	2005	7,035	
29	2006	5,720	
30	2007	3,355	
31	2008	4,477	
32	2009	4,182	
33	2010	4,397	
34	2011	3,431	
35	2012	3,885	
36	2013		
37	2014		
38	2015		
39	2016		

Figure 28A.11 shows the selected cell populating the Output Range dialog.

FIGURE 28A.11

Figure 28A.12 shows the current status of the main Moving Average dialog.

FIGURE 28A.12

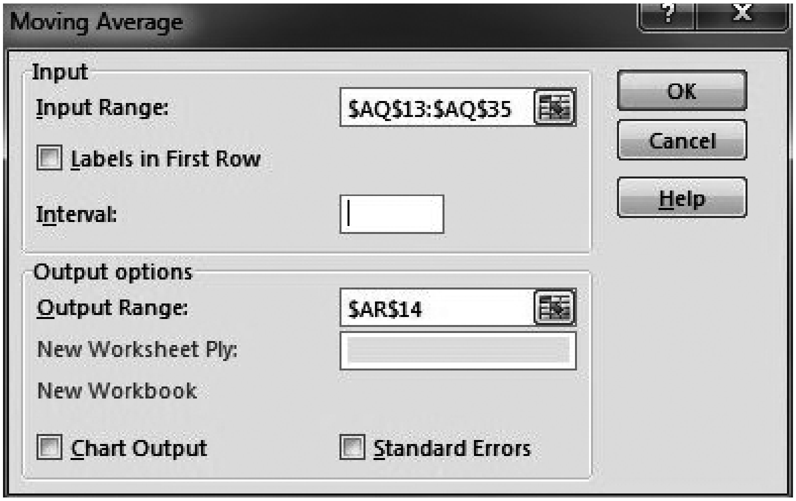


Figure 28A.13 shows entering the interval length of 3. Then click OK.

FIGURE 28A.13

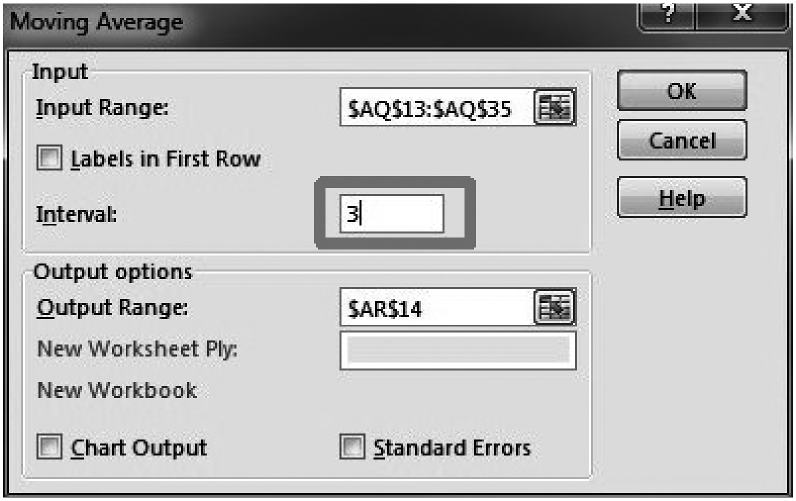


Figure 28A.14 shows the three-period moving average as a forecast. We populate cell AR37 and the following rows with the last forecast value.

FIGURE 28A.14

AR37 | =AR36

	AP	AQ	AR	AS	AT
10	Year	Forfeitures	MA3 FC		
11					
12					
13	1990	7,468			
14	1991	7,356	#N/A		
15	1992	6,771	#N/A		
16	1993	8,738	7,198		
17	1994	7,907	7,622		
18	1995	6,675	7,805		
19	1996	2,816	7,773		
20	1997	8,425	5,799		
21	1998	4,454	5,972		
22	1999	8,888	5,232		
23	2000	5,830	7,256		
24	2001	7,522	6,391		
25	2002	6,727	7,413		
26	2003	6,149	6,693		
27	2004	8,757	6,799		
28	2005	7,035	7,211		
29	2006	5,720	7,314		
30	2007	3,355	7,171		
31	2008	4,477	5,370		
32	2009	4,182	4,517		
33	2010	4,397	4,005		
34	2011	3,431	4,352		
35	2012	3,885	4,004		
36	2013		3,905		
37	2014		3,905		
38	2015		3,905		
39	2016		3,905		

MODULE 29
Forecasting Intermediate Methods

Learning Objectives:

- *Forecast nontrending,* nonseasonal data *using simple exponential smoothing*
- *Forecast trending, nonseasonal data using modified Holt exponential smoothing*
- *Compare exponential smoothing with a moving average*
- *Use judgment in making a forecast*

In this module, we learn how to make forecasts that are typically more accurate and less biased than those based on a moving average. We do this using **exponential smoothing** techniques that are commonly labeled "simple," although they are more sophisticated than the moving average. We also learn how to work with seasonal data. Finally, we briefly examine the use of regression modeling for forecasting.

When forecasting nontrending data, we previously used a moving average, as shown in Figure 28.12 of Module 28. While moving averages are very easy to calculate, they rarely provide the most accurate forecast. In this module, we examine intermediate methods. Intermediate methods can sometimes be considered challenging, but they are typically more accurate than very simple methods. Most can be performed with a spreadsheet, and only a few require knowing more than a little math. For all of these methods, we assume that we are calculating errors and the two error measures—RMSE for accuracy and ME for bias—with whatever forecast we make.

Simple Exponential Smoothing

Simple exponential smoothing (SES) is a method of forecasting nontrending data. It is an improvement over using a moving average. The formula needed to produce SES is as follows:

$$F_t = \alpha A_{t-1} + (1-\alpha)F_{t-1}$$

This formula says that forecast F at time period t is found by taking value α and multiplying it by observation A in the previous period and adding it to the value 1 minus α multiplied by the forecast found in the previous period. The value of α must be between 0 and 1. Its function is to decide how much of the forecast is from the most recent period and how much is from earlier periods. Later, we discuss how to choose (in Excel) its paired mate, $1 - \alpha$. For the first period, there is no forecast. For the second period, the forecast is the value from the first period. Excel knows this formula, and Figures 9.1–9.11 demonstrate using the Data Analysis ToolPak to make an SES forecast.

FIGURE 29.1
Forfeitures Data

L5 =D13/$L$1

	J	K	L	M	N	O	P
1			1000				
2		Year	Forfeitures	ESForecast	Error	Error SQ	
3							Mean
4							RMSE
5		1990	7,467.6				
6		1991	7,356.1				
7		1992	6,771.0				
8		1993	8,738.0				
9		1994	7,907.0				
10		1995	6,674.6				
11		1996	2,816.5				
12		1997	8,425.0				
13		1998	4,454.1				
14		1999	8,887.8				
15		2000	5,829.9				
16		2001	7,522.0				
17		2002	6,727.0				
18		2003	6,149.2				
19		2004	8,756.7				
20		2005	7,035.2				
21		2006	5,719.6				
22		2007	3,355.1				
23		2008	4,476.9				
24		2009	4,182.3				
25		2010	4,397.2				
26		2011	3,431.1				
27		2012	3,885.2				
28		2013					
29		2014					
30		2015					
31		2016					

Figure 29.1 shows the data in the spreadsheet.

FIGURE 29.2
Excel ToolPak Data Analysis Tool

Figure 29.2 shows that we go to the Data Analysis option on the Data Ribbon. This ribbon was previously demonstrated in the Appendix to Module 28.

FIGURE 29.3
Selecting Exponential Smoothing in Excel

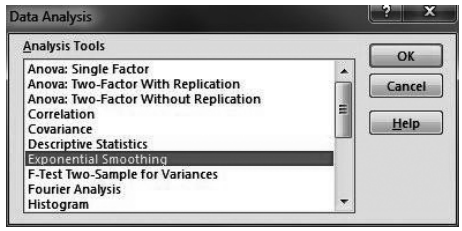


Figure 29.3 shows that with the Data Analysis ToolPak, we select Exponential Smoothing.

FIGURE 29.4
Exponential Smoothing Dialog

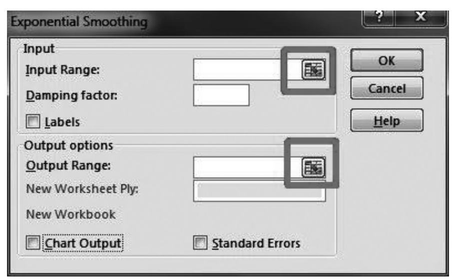


Figure 29.4 shows the Exponential Smoothing dialog. In this dialog, we click on the buttons indicated and search the spreadsheet for the input and preferred output ranges. The results are shown in Figure 29.5.

FIGURE 29.5
Exponential Smoothing Input and Output Ranges

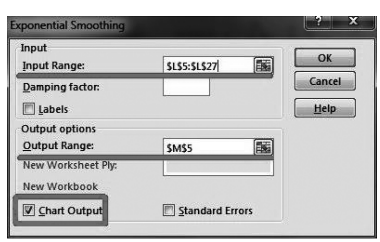


Figure 29.5 shows that we have selected the input range and the output range. Only the first cell of the output range is required. Figure 29.5 shows that it starts on the same row as and in the column beside the input range. Figure 29.5 also shows that the Chart Output option is checked. Do not select the Standard Errors option.

FIGURE 29.6
Exponential Smoothing Output Graph

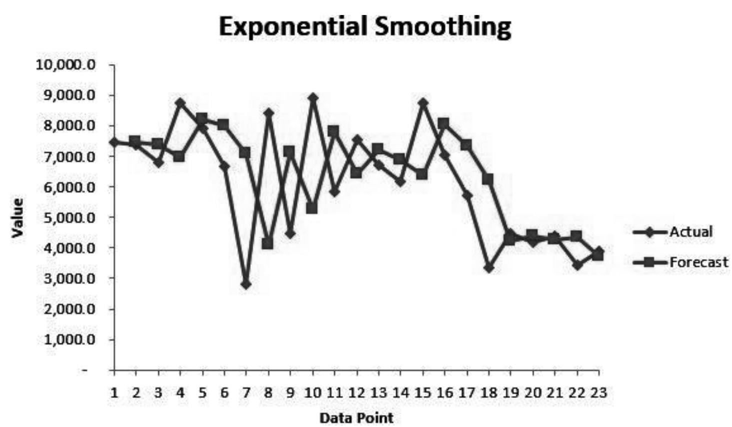


Figure 29.6 shows the output chart, which may be very small, so you may need to expand it to see it. You can use screen grab tools in Excel to move it to a more convenient place on the sheet. The automatically generated chart will not have your original date labels.

FIGURE 29.7
Forfeitures Forecast Using Excel

M5 | #N/A

	J	K	L	M	N	O
2		Year	Forfeitures	ESForecast	Error	Error SQ
3						
4						
5		1990	7,467.6	#N/A		
6		1991	7,356.1	7,467.6		
7		1992	6,771.0	7389.5102		
8		1993	8,738.0	6956.58386		
9		1994	7,907.0	8203.54576		
10		1995	6,674.6	7995.99453		
11		1996	2,816.5	7071.02466		
12		1997	8,425.0	4092.8273		
13		1998	4,454.1	7125.35729		
14		1999	8,887.8	5255.50169		
15		2000	5,829.9	7798.08881		
16		2001	7,522.0	6420.36154		
17		2002	6,727.0	7191.54136		
18		2003	6,149.2	6866.36031		
19		2004	8,756.7	6364.37259		
20		2005	7,035.2	8039.02068		
21		2006	5,719.6	7336.3217		
22		2007	3,355.1	6204.64381		
23		2008	4,476.9	4209.93934		
24		2009	4,182.3	4396.8258		
25		2010	4,397.2	4246.64654		
26		2011	3,431.1	4352.04376		
27		2012	3,885.2	3707.36423		
28		2013				
29		2014				
30		2015				
31		2016				

Figure 29.7 shows the output forecast in column M.

FIGURE 29.8
Forfeitures Forecast Using Excel, With Future Years Added Manually

M28 | =M27

	K	L	M	N	O	P	Q
27	2012	3,885.2	3,707.4	(177.9)	31,634.8		
28	2013		3,707.4				
29	2014		3,707.4				
30	2015		3,707.4				
31	2016		3,707.4				

Figure 29.8 shows that at the bottom of column M, the same value as the last forecast is repeated as the forecast for future periods. This is a manual addition that you supply to complete the table.

FIGURE 29.9
Forfeitures Forecast Using Excel, With Formula Shown

M7 | =0.7*L6+0.3*M6

	K	L	M	N	O	P
2	Year	Forfeitures	ESForecast	Error	Error SQ	
3				(23.3)	3,962,424.4	Mean
4					1,990.6	RMSE
5	1990	7,467.6	#N/A			
6	1991	7,356.1	7,467.6			
7	1992	6,771.0	7,389.5			
8	1993	8,738.0	6,956.6			
9	1994	7,907.0	8,203.5			
10	1995	6,674.6	7,996.0			
11	1996	2,816.5	7,071.0			
12	1997	8,425.0	4,092.8	4,332	18,767,833.0	

Figure 29.9 shows the forecast formula embedded in column M, starting with the second forecast period. Here we see that α and $1 - \alpha$ are included in the formula as fixed numbers. In Figures 29.4 and 29.5, $1 - \alpha$ is labeled "dampening factor."[1] This location is shown again in Figure 29.10. As we made no entry in that location, we use a default value of 0.3. Figure 29.9 also shows error calculations MSE, RMSE, and ME, all of which were defined in Module 28.

Figure 29.9 also shows the forecast results: ME = – 23.3 and RMSE = 1991. The errors before 1997 are excluded to make this analysis comparable to the analysis of the moving average in

1. The term **dampening factor** has a different meaning in academic articles on forecasting.

FIGURE 29.10
Exponential Smoothing Dialog: Damping Factor Option

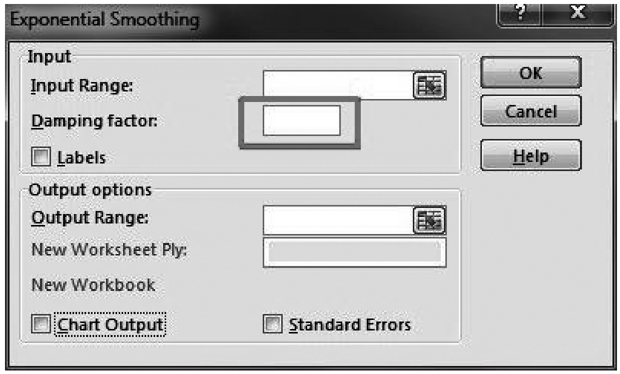
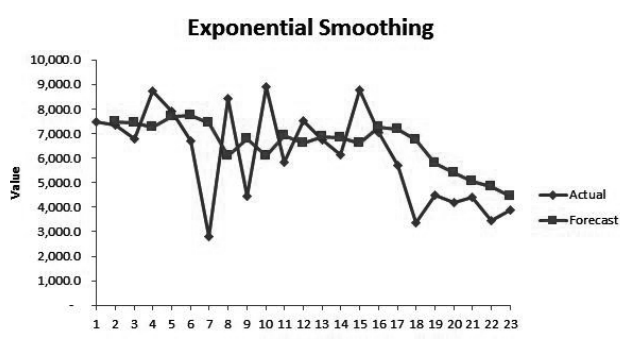


Figure 28.12. Compared with these results (shown in the table below), the forecast is somewhat less accurate, although it is far less biased.

Period	ME	RMSE
MA3	(266)	1663
MA5	(370)	1676
MA7	(575)	1738

However, we used the default dampening factor of 0.3. A little experimentation shows that when the dampening factor is set at 0.7, the values are RMSE = 1683 and ME = 603. This result is within the same general range as the MA forecasts, but it is not better in this instance.

FIGURE 29.11
Exponential Smoothing Output Graph, With Dampening Factor at 0.7



Before selecting this forecast, or any of the MA forecasts, it is beneficial to also look at the graph of the results. Comparing Figure 29.11 (the chart of this forecast) to Figure 29.6, we see that for the recent periods (those leading into our next estimate), the less accurate forecast over the length of the data series will be more accurate now. This paradoxical result comes from the fact that there was a big change in the data in periods 15 through 18 (the mid-2000s) and the more reactive model (which typically has larger errors) found this change, while the less reactive model did not. The forecaster cannot just rely on numbers—judgment here would lead to using the default model or the MA3 model for the present (the reader should also review Figure 28.23 in Module 28). If this forecast is substantially contributory to total revenue, it will bear constant monitoring.

Trending Data

As with moving averages, there is a different version of exponential smoothing for trending data. Trending exponential smoothing is known as **Holt exponential smoothing**. Here we use a slightly simpler version of Holt (T. M. Williams, 1987). We cannot use a built-in Excel function because there is none for trending exponential smoothing. However, exponential smoothing requires relatively little math.

We need four formulas for the forecast and two more to start off at the right value, known as **initialization**.

For the modified Holt exponential smoothing, we use these three formulas:

$$F_t = S_{t-1} + B_{t-1}$$

$$S_t = F_t + \alpha e_t$$

$$B_t = B_{t-1} + \beta e_t$$

$$e_t = A_t - F_t$$

The first formula says the forecast F_t for the current period is equal to the level calculated at the last period S_{t-1} added to the trend calculated at the last period B_{t-1}. The second formula says the level S_t is equal to the forecast F_t plus α times the error e_t. The third formula says the trend B_t is equal to the previous trend plus β times the error e_t.[2] The last formula says the error e_t is equal to the actual A_t minus the forecast F_t. The α value serves the same function as with SES;[3] It is set by the user to decide how much of the level will be from the most recent observation and how much from earlier observations. The value β is set by the user to decide how much of the trend is from the most recent observation and how much from earlier observations.

Below, we demonstrate using these formulas to make a forecast. First, however, we need some formulas to initialize the forecast. *Initialize* means selecting values for S and B for the time period $t = 0$, meaning the time period just before the first actual observation. Holt will perform

---

2. The difference between this version and the unmodified Holt formulation is that the trend formula in Holt is as follows:

$$B_t = B_{t-1} + \alpha\beta e_t$$

3. In fact, if we set β to zero and initialize B_t as zero (thus, effectively deleting all references to B_t from the formulas), this is an alternate expression of SES.

relatively well without initialization, but it performs better when initialized. The formulas are quite simple. First, we want an average of the first few observations in two groups. If the data are monthly or quarterly, we average over the first year and then the second year. If it is annual, we choose a length of time, L, that is not excessive compared with the entire number of observations. In the upcoming example, $L = 5$, so there is an average of the first 5 observations and an average of the second 5 observations. We then use these formulas:

$$B_0 = \frac{\left(\bar{A}_2 - \bar{A}_1\right)}{L}$$

$$S_0 = \bar{A}_1 - B_0 \times \left(\frac{L+1}{2}\right)$$

What the first formula says is that the initial value of trend B_0 is the average of the second group of observations $\bar{A}_2$ (the line above it means average) minus the average of the first set of observations $\bar{A}_1$ divided by the number of observations L in the average. What the second formula says is that the initial value of level S_0 is equal to the first average $\bar{A}_1$ minus one half of the number of observations plus 1.[4]

Figure 29.12 shows us a forecast of the trending data for which we used the two moving averages in Module 28. In subsequent figures, we examine the elements of this forecast.

FIGURE 29.12

Trending Data Forecast Using Excel

T46

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			
8	1980	1,142		1,142.77	130.46	1,153	(11)	111.14	
9	1981	1,311		1,309.44	130.84	1,273	38	1,452.83	
10	1982	1,415		1,416.13	130.59	1,440	(25)	646.55	
11	1983	1,515		1,516.41	130.27	1,547	(32)	1,017.25	
12	1984	1,686	1,414	1,684.34	130.66	1,647	40	1,571.28	
13	1985	1,828		1,827.15	130.79	1,815	13	163.55	
14	1986	1,909		1,911.11	130.30	1,958	(49)	2,430.14	
15	1987	2,044		2,044.24	130.33	2,041	3	8.86	
16	1988	2,223		2,220.52	130.81	2,175	48	2,340.30	
17	1989	2,330	2,067	2,331.01	130.60	2,351	(21)	457.66	
18	1990	2,431		2,432.74	130.29	2,462	(30)	923.63	
19	1991	2,354		2,364.22	128.20	2,563	(209)	43,795.45	
20	1992	2,278		2,288.55	126.06	2,492	(215)	46,054.29	
21	1993	2,402		2,402.86	125.93	2,415	(12)	152.89	
22	1994	2,504		2,504.90	125.68	2,529	(25)	632.36	
23	1995	2,621		2,621.42	125.58	2,631	(10)	93.15	
24	1996	2,742		2,742.45	125.54	2,747	(5)	22.97	
25	1997	2,937		2,933.63	126.23	2,868	69	4,774.96	
26	1998	3,069		3,068.90	126.32	3,060	10	90.69	
27	1999	3,204		3,203.69	126.41	3,195	9	79.46	
28	2000	3,526		3,515.83	128.37	3,330	196	38,222.43	
29	2001	3,679		3,677.01	128.71	3,644	35	1,192.58	
30	2002	3,373		3,394.97	124.39	3,806	(432)	186,939.64	
31	2003	3,551		3,549.11	124.70	3,519	31	980.48	
32	2004	4,042		4,024.05	128.39	3,674	369	135,922.94	
33	2005	4,375		4,363.99	130.61	4,152	223	49,585.85	
34	2006	4,439		4,442.12	130.06	4,495	(55)	3,051.19	
35	2007	4,645		4,640.92	130.79	4,572	72	5,234.84	
36	2008	4,891		4,884.79	131.98	4,772	119	14,168.27	
37	2009	4,615		4,635.02	127.96	5,017	(402)	161,472.53	
38	2010	5,076		5,060.70	131.09	4,763	313	98,216.66	
39	2011	5,610		5,589.32	135.28	5,192	418	175,100.41	
40	2012	5,826		5,820.97	136.29	5,725	101	10,291.99	
41	2013			5,957.26	136.29	5,957			
42	2014			6,093.56	136.29	6,094			
43	2015			6,229.85	136.29	6,230			
44	2016			6,366.14	136.29	6,366			

4. If the number of observations is even, the numerator of the fraction in the parentheses should be adjusted to $L + 1$.

Figure 29.13 shows us the calculation of the average $\bar{A}_1$ in cell L12; $\bar{A}_2$ is shown in cell L17.

FIGURE 29.13
Trending Data Forecast Using Excel: Calculation of $\bar{A}_1$ and $\bar{A}_2$

L12 | =AVERAGE(K8:K12)

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			
8	1980	1,142		1,142.77	130.46	1,153	(11)	111.14	
9	1981	1,311		1,309.44	130.84	1,273	38	1,452.83	
10	1982	1,415		1,416.13	130.59	1,440	(25)	646.55	
11	1983	1,515		1,516.41	130.27	1,547	(32)	1,017.25	
12	1984	1,686	1,414	1,684.34	130.66	1,647	40	1,571.28	
13	1985	1,828		1,827.15	130.79	1,815	13	163.55	
14	1986	1,909		1,911.11	130.30	1,958	(49)	2,430.14	
15	1987	2,044		2,044.24	130.33	2,041	3	8.86	
16	1988	2,223		2,220.52	130.81	2,175	48	2,340.30	
17	1989	2,330	2,067	2,331.01	130.60	2,351	(21)	457.66	

Figure 29.14 shows $\bar{A}_2$ minus $\bar{A}_1$ in cell L3.

FIGURE 29.14
Trending Data Forecast Using Excel: Calculation of $\bar{A}_2$ Minus $\bar{A}_1$

L3 | =L17-L12

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

Figure 29.15 shows a spreadsheet expression in cell L4 to determine the length of *L*. The range must be properly defined for this expression to be effective.

FIGURE 29.15
Trending Data Forecast Using Excel: Calculating *L*

L4 | =COUNT(K8:K12)

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

Figure 29.16 shows calculating B_0 in cell L5 from the information we put in cells L3 divided by L4.The arrows in columns M and N show that this value is also placed in cell N7 as the initial trend value.

FIGURE 29.16

Trending Data Forecast Using Excel: Calculating B_0

L5 | =L3/L4

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

Figure 29.17 shows calculating the quantity $\frac{L+1}{2}$ in cell L6 (note that the letters used in cell addresses are unrelated to the letters used in formulas).

FIGURE 29.17

Trending Data Forecast Using Excel: Calculating the Initial Trend

L6 | =(L4+1)/2

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

Figure 29.18 shows calculating S_0 from $\bar{A}_1$ in cell L12, B_0 in cell L5, and $(L + 1)/2$ in cell L6. This value is also placed in cell M7 as the initial level value.

Figure 29.19 shows the calculation of the formula $S_t = F_t + \alpha e_t$ in cell M8. This formula is copied to all the rows where there are data (see Figure 28.12).

Figure 29.20 shows the calculation of the formula $B_t = B_{t-1} + \beta e_t$ in cell N8, which is copied to all the rows where there are data.

Figure 29.21 shows the adding together of the level S_t and trend B_t from the previous row in cell O8 to get the forecast on the current row. This formula is copied to all the rows where there are data.

The error calculations MSE, RMSE, and ME are as defined previously. This model has been "fit," which means that alternative values of α and β have been tried in order to minimize RMSE. Typically, when fitting modified Holt, β values should be kept small. The author usually does not consider values greater than 0.1 unless graphic evidence shows sharp changes in trend, in which

FIGURE 29.18

Trending Data Forecast Using Excel: Calculating S_0

L7 | =L12-(L5*L6)

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

FIGURE 29.19

Trending Data Forecast Using Excel: Calculating S_t

M7 | =L7

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

FIGURE 29.20

Trending Data Forecast Using Excel: Calculating B_t

N7 | =L5

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			

FIGURE 29.21

Trending Data Forecast Using Excel: Adding Previous Level S_t and Trend B_t

O8 | =M7+N7

	J	K	L	M	N	O	P	Q	R
2	Year	General Sales	Initialize	α	β	Forecast	Error	ErrorSQ	
3			653	0.95	0.01		17	29,915	Mean
4			5					173	RMSE
5			130.56	→	↘				
6			3		↓				
7			1,022.22	1,022.22	130.56	↘			
8	1980	1,142		1,142.77	130.46	1,153	(11)	111.14	

case it may be more effective to discard older data after enough periods have elapsed to know that the trend change has stabilized, unless the analyst thinks another sharp change may be lurking in the future. In the table, the Holt model is compared with the previous seven-period moving average. Both the ME and RMSE are more favorable with the Holt model.

Method	ME	RMSE
7MA	27	179
Holt$_{\alpha = 0.95,\ \beta = 0.01}$	19	173

Summary

Exponential smoothing is a technique applied to time series data to produce forecasts. There is a different version of exponential smoothing for trending data. Trending exponential smoothing is known as Holt exponential smoothing.

Assignments

For SES solutions, use $\alpha = 0.1$. For Holt solutions, use $\alpha = 0.95$ and $\beta = 0.01$.

1. Recall the forecast produced for assignment 3 in Module 28 for the Northland special revenue fund. The original data are provided in Table 29.1.

   a. Using the same initial data and assumptions, refine the forecast using the exponential smoothing method appropriate for nontrending data.

   b. Produce a graph showing the original data, the 5-year moving average forecast from Module 28, and the forecast produced using the exponential method.

   c. Which forecast would you recommend using, and why?

TABLE 29.1
Northland: Special Revenue Fund

1996	53,421,417	2005	49,055,130
1997	64,600,858	2006	124,202,747
1998	67,053,747	2007	86,126,170
1999	24,316,124	2008	27,862,519
2000	66,110,642	2009	85,704,726
2001	93,389,137	2010	81,483,730
2002	49,686,023	2011	45,459,605
2003	20,706,410	2012	22,672,732
2004	81,238,032	2013	58,226,034

2. Recall the forecast produced in assignment 5 in Module 28 for River County's property tax revenue. The original data are provided in Table 29.2.

   a. Using the same initial data and assumptions, refine the forecast using the exponential smoothing method appropriate for trending data.

b. Create a matrix showing the ME and RMSE for the three scenarios of moving-average forecasts from Module 28 and the ME and RMSE produced using the exponential method.

c. Which forecast would you recommend using, and why?

TABLE 29.2
River County: Property Tax Revenue

Fiscal Year	Property Tax Levy ($000)	Fiscal Year	Property Tax Levy ($000)
1992	34,762.30	2004	44,165.63
1993	32,354.05	2005	50,550.37
1994	30,971.67	2006	62,733.41
1995	31,010.82	2007	67,096.62
1996	32,382.92	2008	70,422.87
1997	33,211.58	2009	74,674.33
1998	36,407.76	2010	74,996.80
1999	38,930.54	2011	68,019.59
2000	43,246.20	2012	62,401.17
2001	48,342.35	2013	54,584.58
2002	45,042.55	2014	39,842.99
2003	45,322.70		

Note: These figures are rounded to the nearest cent.

Reference

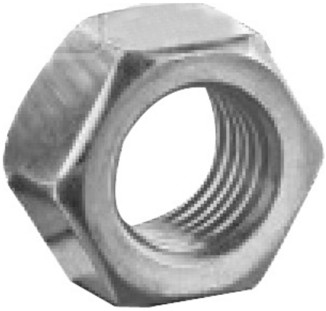
Williams, T. M. (1987). Adaptive Holt-Winters forecasting. *Journal of the Operational Research Society, 38*(6), 553–560.

Additional Readings

Armstrong, J. S. (1985). *Long-range forecasting: From crystal ball to computer* (2nd ed.). New York, NY: Wiley.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2010). *Forecasting: Methods and applications,* New York, NY: Wiley.

Williams, D. W. (2008). Forecasting methods for serial data. In G. Miller & K. Yang (Eds.), *Handbook of research methods in public administration* (2nd ed., pp. 595–665). Boca Raton, FL: CRC Press/Taylor & Francis.



MODULE 30
Forecasting: Advanced Intermediate Methods

Learning Objectives:

- *Identify seasonal data*
  - *Deseasonalize seasonal data*
  - *Reseasonalize forecasts*
- *Make a forecast of nontrending data using a very simple regression model*

In this module, we learn a simple method of working with seasonal data, and we learn how to make a forecast using a simple regression model with a single-period time lag.

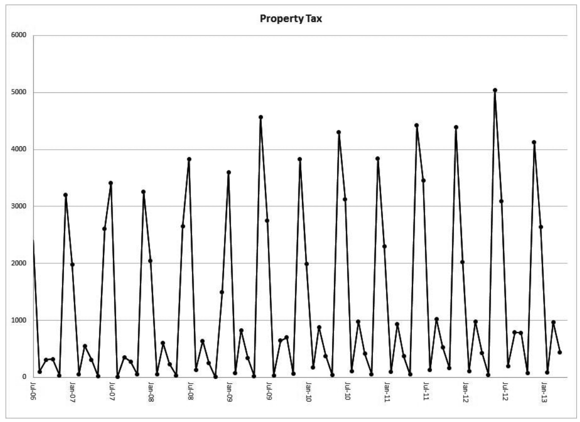
Seasonality

Seasonality arises when time series data exhibit a pattern over a cyclical calendar period, such as a year, quarter, month, or week. Most commonly, the idea of seasonality is associated with years, but practical human behavior can lead to other periodic seasonality as well. For example, businesses may have quarterly profit reports, which could lead to increased activity in the last month of fiscal quarters and matching increased tax collections. In this case, the pattern would be relatively low taxes in January and February, followed by more collection in March, followed by the same pattern in April through June, and so forth.

When data follow any seasonal pattern, forecasting must account for seasonality before methods like moving average, the exponential smoothing methods we have examined, or the regression method we will look at can be used.

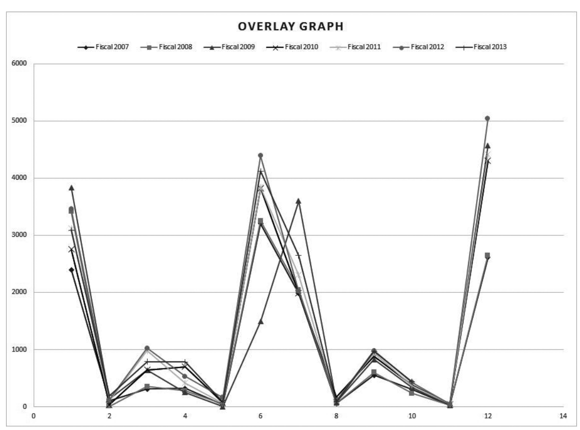
Figure 30.1 shows seasonal data. These data show two seasonal peaks, one in winter and one in summer. Generally, there are two observations in the seasonal high period and four in the seasonal low period, so the data might appear to exhibit semiannual seasonality. However, the summer peaks are higher than the winter peaks, so annual seasonality is likely correct.

FIGURE 30.1
Property Taxes: *Seasonal Data*



Sometimes it is hard to tell seasonality by visual inspection, so it may be useful to make an overlay chart as shown in Figure 30.2.

FIGURE 30.2
Property Taxes: Overlay Graph (FY2007–2013)



To make the overlay chart, the data are grouped by year and month as shown in Figure 30.3.

On the overlay chart, the observations clearly show a distinctive annual pattern, so the data are seasonal. There are advanced statistical techniques to confirm seasonality; however, when the pattern is this strong, there is no need to use such techniques. In addition, the June (period 12) peaks are higher than the December (period 6) peaks, so estimating factors as semiannual would be incorrect.

FIGURE 30.3
Property Taxes: Data by Year and Month

	Overlay						
	Fiscal 2007	**Fiscal 2008**	**Fiscal 2009**	**Fiscal 2010**	**Fiscal 2011**	**Fiscal 2012**	**Fiscal 2013**
July	2400	3416	3835	2754	3122	3455	3091
August	103	10	133	37	110	127	198
September	313	356	642	648	980	1024	790
October	323	278	248	702	413	532	784
November	34	50	6	67	57	162	78
December	3207	3256	1492	3826	3842	4391	4124
January	1985	2044	3602	1993	2304	2021	2647
February	56	57	81	174	99	106	86
March	554	604	826	884	938	980	972
April	308	236	339	376	372	425	440
May	26	33	24	48	50	45	
June	2609	2651	4572	4303	4420	5043	

When there is no pattern or the analyst is unsure, it is more reasonable to judge that the data are not seasonal. Miller and Williams (2003, 2004) have demonstrated that seasonal methods are likely to find seasonality when there is none or are likely to overestimate seasonality. So the analyst should be cautious. Excel does not include a seasonal estimation package, and many methods require sophisticated software. Although there is an exponential smoothing method that incorporates seasonality, it may be subject to severe overestimation over time. The simplest safe method for estimating seasonality in a spreadsheet is the classical decomposition method (Makridakis, Wheelwright, & Hyndman, 1998).

The steps to implementing classical decomposition are demonstrated in Figures 30.4–30.20.

Figure 30.4 shows the top part of the entire computation. There are 82 observations of these data. The order of magnitude is not reported.

FIGURE 30.4
Property Tax Data: Decomposition

T83

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	
24	20	Feb-08	57	1117.5	1122.6	0.051			0.083	690.9	
25	21	Mar-08	604	1127.8	1139.7	0.530			0.620	974.6	
26	22	Apr-08	236	1151.6	1150.3	0.205			0.272	866.8	
27	23	May-08	33	1149.1	1147.3	0.029			0.038	857.5	
28	24	Jun-08	2651	1145.4	1071.9	2.473			2.962	895.1	
29	25	Jul-08	3835	998.4	1063.3	3.607			2.563	1496.5	
30	26	Aug-08	133	1128.3	1129.3	0.118			0.085	1572.2	
31	27	Sep-08	642	1130.3	1139.5	0.563			0.557	1153.6	
32	28	Oct-08	248	1148.8	1153.0	0.215			0.369	671.7	
33	29	Nov-08	6	1157.3	1157.0	0.005			0.060	99.7	
34	30	Dec-08	1492	1156.6	1236.6	1.207			2.550	585.2	
35	31	Jan-09	3602	1316.7	1271.6	2.833			1.842	1955.0	
36	32	Feb-09	81	1226.6	1222.6	0.066			0.083	981.8	
37	33	Mar-09	826	1218.6	1218.8	0.678			0.620	1332.8	
38	34	Apr-09	339	1219.1	1238.0	0.274			0.272	1245.1	
39	35	May-09	24	1256.9	1259.5	0.019			0.038	623.6	
40	36	Jun-09	4572	1262.0	1359.3	3.364			2.962	1543.6	
41	37	Jul-09	2754	1456.5	1389.5	1.982			2.563	1074.7	
42	38	Aug-09	37	1322.4	1326.3	0.028			0.085	437.4	
43	39	Sep-09	648	1330.2	1332.6	0.486			0.557	1164.3	

Figure 30.5 and subsequent figures show calculations in the formula bars at the top of the figures and in the cell where the calculation is made in the more complete figure. In Figure 30.5 we calculate a 12-period moving average and place it beside the seventh observation of the series counting from the beginning. Because of the critical location and the relative ease of calculating a moving average, it is better to do this with =Average([Range]), than to use the Data Analysis ToolPak function.[1]

FIGURE 30.5
Property Tax Data: 12-Period Moving Average (12MA)

D11 | =AVERAGE(C5:C16)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Copy this formula until you reach a row where the range of the moving average (the cell within the Excel parentheses that follows the colon) is the last cell that contains data. Do not go beyond that location. This is shown in Figure 30.6, where the entries in column D end at row 81. Although the formula will compute as if correct for subsequent rows, the moving average will have fewer than 12 observations.

Figure 30.7 shows a 2-period moving average following the 12-period moving average. This is needed with even-numbered observations to center the moving average beside an original observation rather than halfway between two observations. The first 2-period moving average is set beside the first 12-period moving average. Referring back to Figure 30.6, we see that the column stops one row before the 12-period moving average column stops.

---

1. If you are calculating seasonality with quarterly data (four observations), you would use a four-period moving average and locate the first MA at period 3. For all even-numbered periods, this step is one period past half. For odd-numbered periods (daily, within weeks, and monthly within quarter), place the MA at the middle observation and skip the next step. This method cannot be used when the number of periods varies (weeks within months).

FIGURE 30.6

Property Tax Data: 12-Period Moving Average (12MA), Last Date for Which It Can Be Calculated

	A	B	C	D	E	F	G	H	I	J
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized
76	72	Jun-12	5043	1496.0	1484.9	3.396			2.962	1702.7
77	73	Jul-12	3091	1473.8	1499.8	2.061			2.563	1206.2
78	74	Aug-12	198	1525.9	1525.1	0.130			0.085	2340.5
79	75	Sep-12	790	1524.3	1523.9	0.518			0.557	1419.5
80	76	Oct-12	784	1523.6	1524.2	0.514			0.369	2123.4
81	77	Nov-12	78	1524.8					0.060	1296.7
82	78	Dec-12	4124						2.550	1617.6
83	79	Jan-13	2647						1.842	1436.7
84	80	Feb-13	86						0.083	1042.4
85	81	Mar-13	972						0.620	1568.3
86	82	Apr-13	440						0.272	1616.0
87										

FIGURE 30.7

Property Tax Data: 2-Period Moving Average (2MA)

E11 =AVERAGE(D11:D12)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

FIGURE 30.8

Property Tax Data: Seasonal Factors

F11 =C11/E11

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Figure 30.8 shows how to compute raw seasonal ratios by dividing the original data by the correctly centered two-period double moving average.[2]

FIGURE 30.9

Property Tax Data: Average Seasonal Factors

G11 =AVERAGE(F11,F23,F35,F47,F59,F71)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Figure 30.9 shows averaging the seasonal ratios for each seasonal period (month). This set requires hunting appropriate cells in the spreadsheet and is subject to the errors of omitting a relevant cell or selecting an inappropriate cell. Also, if copy and paste is used to fill the column for the 12 seasonal periods (for annual data), there is a danger that the formula will refer to stray data below the calculation table in lower cells. The analysts should proofread this step carefully.

Figure 30.10 shows calculating the total of the seasonal factors. For annual data, the desired sum is 12. If, as is likely, the total is not 12, the seasonal factors should be normalized as shown in the next steps.

Figure 30.11 shows a fixed value of 12 (for 12 seasonal periods) in a convenient location in the spreadsheet.

Figure 30.12 shows how to divide the desired total of the seasonal factors (12) by the calculated total. This produces an adjustment factor to be used to correct the seasonal factors.

Figure 30.13 shows the formula =G11*H$4 (in the formula bar), which is used to adjust the seasonal factor. This formula is copied to the 12 seasonal periods.

Figure 30.14 shows that we have totaled the 12 seasonal factors and have confirmed that they total to 12.

---

2. It is called *double moving average* because it is a moving average of a moving average value. For odd-numbered periods, the double moving average should not be calculated—the original moving average should be used for this step.

FIGURE 30.10

Property Tax Data: Total Average Seasonal Factors

G3 =SUM(G11:G22)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

FIGURE 30.11

Property Tax Data: Inputting the Number 12 to Use in Calculations

H2 12

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

FIGURE 30.12

Property Tax Data: Adjustment Factor for Correcting Seasonal Factors

H4 | =H2/G3

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

FIGURE 30.13

Property Tax Data: Normalized Seasonal Factors

H11 | =G11*$H$4

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

FIGURE 30.14

Property Tax Data: Total Average Normalized Seasonal Factors

H3 | =SUM(H11:H22)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

FIGURE 30.15

Property Tax Data: Inputting Dampening Factor of 0.99

I4 | 0.99

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Earlier, we discussed the possibility that seasonal factors overestimate seasonality. There are sophisticated methods for correcting for this, but they require extensive math. A simple (if less perfect) method is to dampen the seasonal factors (Armstrong, 2001). For the current data, dampening is not required; the pattern is clear, and seasonal factors are unlikely to be substantially overestimated. However, to demonstrate the calculation of dampening, a dampening factor is included in this example. Figure 30.15 shows the dampening factor of 0.99. Values of less than 1 reduce the distance from 1 of the adjustment of the seasonal factor. The adjustment of 0.01 may seem small, but it may be excessive in this instance. Dampening without the aid of statistical methods requires substantial care. Examine the overlay graph. If the lines lie pretty much on top of each other, dampen little or perhaps not at all (i.e., set the factor at 1). If you are convinced there is a pattern but that you must screen out some "noise" to see it, then you may want to dampen somewhat. If you feel a need to dampen by as much as 0.8 (for substantial noise) or even more, you may not actually have a seasonal pattern, or the pattern may change too frequently to be modeled with classical decomposition.

FIGURE 30.16

Property Tax Data: Damped Seasonal Factors

I11 | =H11*$I$4+(1-$I$4)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Figure 30.16 shows the implementation of the dampening factor with the formula =H11*$I$4+(1-$I$4). This is the expression:

$$I_t' = I_t \times d + 1 \times (1 - d)$$

This means that the damped seasonal factor I' for period t is equal to the seasonal factor I for period t times the dampening factor d plus 1 times the quantity 1 minus the dampening factor. In the spreadsheet calculation, we omit the superfluous multiplication by 1.

FIGURE 30.17

Property Tax Data: Total Damped Seasonal Factors

I3 | =SUM(I11:I22)

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Figure 30.17 shows that we total the damped seasonal factors to verify that they still total to the number of seasonal periods (12).

FIGURE 30.18

Property Tax Data: Application of Seasonal Factors to All Observations

I10 | =I22

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonal s	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Figure 30.18 shows that to put seasonal factors beside all 82 observations of our original data, we repeat the factors we calculated from the averaged, normalized, and damped calculations (periods 7 through 18 in Figure 30.18).

FIGURE 30.19
Property Tax Data: Deseasonalized Data

J11 =C11/I11

	A	B	C	D	E	F	G	H	I	J	K
1	Period number (from start)	Month	Property Tax	12MA	2MA	Seasonal Ratios	Average Seasonals	Normalized Seasonals	Damped Seasonals	De-seasonal-ized	
2								12			Factor
3							11.971	12.000	12.000		Sum
4								1.0024203	0.99		Factor
5	1	Jul-06	2400						2.563	936.5	
6	2	Aug-06	103						0.085	1217.5	
7	3	Sep-06	313						0.557	562.4	
8	4	Oct-06	323						0.369	874.8	
9	5	Nov-06	34						0.060	565.2	
10	6	Dec-06	3207						2.550	1257.9	
11	7	Jan-07	1985	993.2	1035.5	1.917	1.847	1.851	1.842	1077.4	
12	8	Feb-07	56	1077.8	1074.0	0.052	0.073	0.073	0.083	678.8	
13	9	Mar-07	554	1070.1	1071.9	0.517	0.614	0.616	0.620	893.9	
14	10	Apr-07	308	1073.7	1071.8	0.287	0.264	0.265	0.272	1131.2	
15	11	May-07	26	1069.9	1070.6	0.024	0.029	0.029	0.038	675.6	
16	12	Jun-07	2609	1071.3	1073.3	2.431	2.974	2.982	2.962	880.9	
17	13	Jul-07	3416	1075.3	1077.8	3.169	2.572	2.578	2.563	1333.0	
18	14	Aug-07	10	1080.3	1080.3	0.009	0.075	0.075	0.085	118.2	
19	15	Sep-07	356	1080.3	1082.4	0.329	0.551	0.552	0.557	639.7	
20	16	Oct-07	278	1084.5	1081.5	0.257	0.362	0.363	0.369	752.9	
21	17	Nov-07	50	1078.5	1078.8	0.046	0.051	0.051	0.060	831.2	
22	18	Dec-07	3256	1079.1	1080.8	3.012	2.559	2.565	2.550	1277.1	
23	19	Jan-08	2044	1082.6	1100.0	1.858			1.842	1109.4	

Figure 30.19 shows that we compute the deseasonalized data (the data we will forecast) by dividing the raw data in column C by the damped seasonal factors in column I. The formula for deseasonalization is as follows:

$$A'_t = \frac{A_t}{I'_t}$$

This formula can be read to say the deseasonalized actual A' for period t is equal to the actual A for period t divided by the damped seasonal factor I' for period t.

Figure 30.20 shows the deseasonalized data as darker lines overlaid on the original data, now shown with dotted lines.

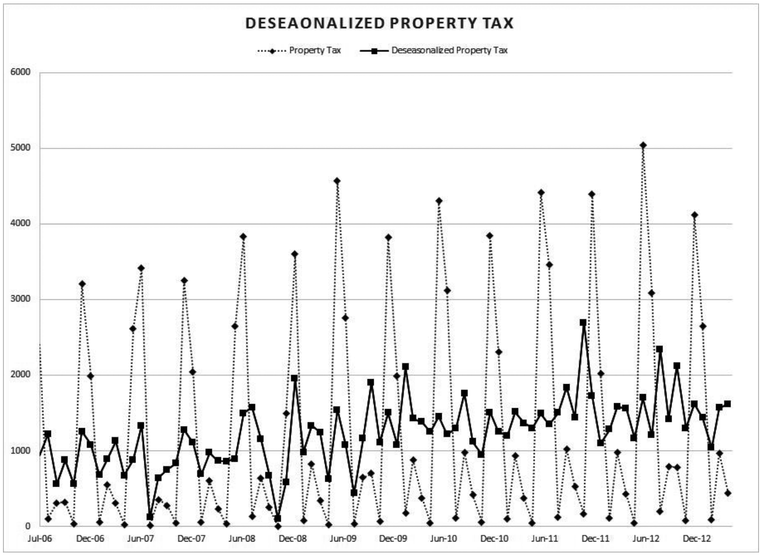
Seasonal *factors* computed with this method should not be forecast; they should be repeated for each seasonal period.

To **reseasonalize** the forecast (to make it follow the same pattern as the original data), use the formula

$$F'_t = F_t \times I'_t$$

This formula can be read to say the forecast F' for period t is equal to the forecast F for period t times the damped seasonal factor I' for period t.

FIGURE 30.20
Property Tax Data: Seasonal and Deseasonalized Data Overlay Graph



Simple Cross-Sectional Regression With Lagged Input *X* Variables

Regression is a statistical procedure that relates input variables, also called **predictor variables**, to an output (**predicted**) variable. The general form of a regression model is as follows:

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon$$

This says that the predicted **output variable** Y is equal to the sum of an intercept α plus a series of coefficients $\beta_{1,2\,...\,n}$ multiplied by their respective predictor variables $x_{1,2\,...\,n}$ plus a random element ε. When we make a regression model, we do not model the random element (because it is random), but we do try to verify that it is relatively random.

This model is linear, which means that on a graph, the data should group up in a line (they do not all have to fall directly on the line, and when there is more than one variable, it may not be possible to see the line on an unaltered graph). Because of this linearity requirement, we transform much of the data before we use it. Anything we transform before forecasting, we have to reverse-transform after the regression to have a usable forecast.

Before we can make a regression model, we need the predicted variable Y and the predictor variables $x_{1,2\,...\,n}$.[3] We do not need a variable for the intercept. The predicted variable is the value

3. It is assumed that the reader of this book has access to other material for basic statistical concepts. Regression-based forecasting should not be attempted by those without statistical training.

we want to forecast. The predictor variables are variables that we believe may help predict (explain) the predicted variable. In the following example and the example in Appendix D of this book, the explanatory variables are very simple. In actual practice, more complex models may be used; however, the analyst should keep in mind that most data series are relatively short, so the number of predictor variables should be kept to a minimum.[4]

The predictor variables need to have values we can know at the time we are making the forecast. There are generally two ways to do this. First, we can use **lagging** (the method we use here), which consists of using a prior-period predictor variable for a later-period predicted variable. Lagging is often effective but is usually weaker than using current-period cross-sectional modeling. Still, it is judged here to be preferable to the alternate method, which is to forecast the predictor variables into the period of the predicted variable and avoid or reduce lagging. To do this, one must have more forecasts, one for each predictor variable. The model fit may be better, but there will be a hidden cost, which is that the uncertainty associated with using estimated predictor variables, rather than measured predictor variables, is usually unknown. Consequently, the alternate approach can lead to completely unexpected catastrophic failure.

In the following, we examine forecasting using Excel. In Appendix D, we review a more challenging case. In Appendix E, we examine types of forecasting not demonstrated in this book.

Regression and Forecasting in Excel Using Data Analysis ToolPak

FIGURE 30.21

Water and Sewer Data

Year	Log Rev Per Gal/Lag 1	Thousand Gallons Per Cap/Lag 1	Log Revenue
1981	-3.373	78.06	8.418
1982	-3.323	77.88	8.423
1983	-3.257	67.42	8.504
1984	-3.199	70.57	8.538
1985	-3.179	72.39	8.578
1986	-3.150	73.93	8.649
1987	-3.036	66.50	8.641
1988	-3.053	67.79	8.638
1989	-3.084	72.43	8.737
1990	-2.997	74.06	8.757
1991	-2.953	70.06	8.776
1992	-2.940	71.15	8.809
1993	-2.928	74.76	8.851
1994	-2.849	68.35	8.856
1995	-2.843	65.98	8.868
1996	-2.827	64.92	8.864
1997	-2.822	63.04	8.889
1998	-2.786	60.94	8.915
1999	-2.728	55.99	8.891
2000	-2.759	56.16	8.904
2001	-2.751	56.32	8.926
2002	-2.730	56.15	8.933
2003	-2.703	53.68	8.928
2004	-2.690	51.37	8.947
2005	-2.654	49.63	8.954
2006	-2.651	50.22	8.995
2007	-2.611	50.56	9.027
2008	-2.564	48.68	9.080
2009	-2.530	50.53	9.108

4. Excessive numbers of predictor variables relative to total observations can undermine the validity of the regression model.

Figure 30.21 shows the data that are to be used. These are derived from the water and sewer decomposition problem previously found in Module 28: Basic Forecasting Concepts. The revenue per gallon and revenue data have been converted with the log function, which simply requires the use of =Log([Cell Address]). Only positive numbers can be logged, and numbers less than 1 will have negative logs. The function statement allows for a second **argument**, but when it is omitted, the base is 10.

These data are logged to remove an expected upward curve in the data. Regression is designed for linear analysis, and it fails when the mass of the data curves as it travels. There are many devices for addressing this, but logging the data is a well-known simple method.

The labels at the top of the input data in Figure 30.21, Log Rev Per Gal and Thousand Gal Per Cap, each end in "/Lag 1." *Lagging* means using data from a prior row, in this case from the first prior row. Sophisticated software has functionality that handles this. In Excel, it is done manually—that is, on the row with the date and output (*Y* variable) from 1981; the input data are from 1980. This means that we have no input data for the output year 1980. We lag the data because that is the point—we want to make a forecast for the next year after the input data run out.

FIGURE 30.22
Excel LOG Function

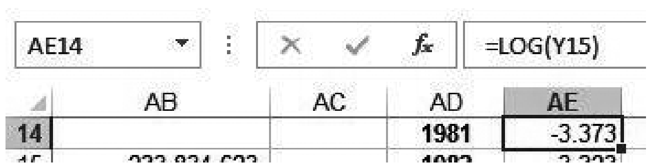


Figure 30.22 shows using =LOG([cell address]) to log data.

FIGURE 30.23
Excel Data Analysis ToolPak Dialog: Selecting Regression

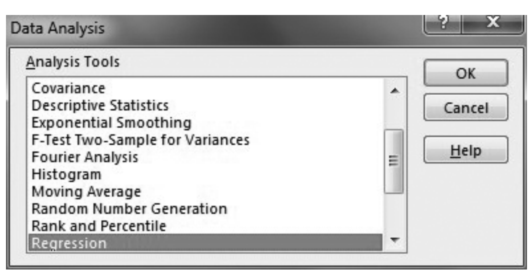


Figure 30.23 shows selecting the Regression dialog from the Data Analysis ToolPak menu.

Figure 30.24 shows the Regression dialog.

Figure 30.25 shows that we have selected the data in column AG (Log Revenue) as our *Y* (predicted) variable and the data in columns AE and AF (Log Rev Per Gal/Lag 1 and Thousand

FIGURE 30.24
Excel Data Analysis ToolPak: Regression Dialog

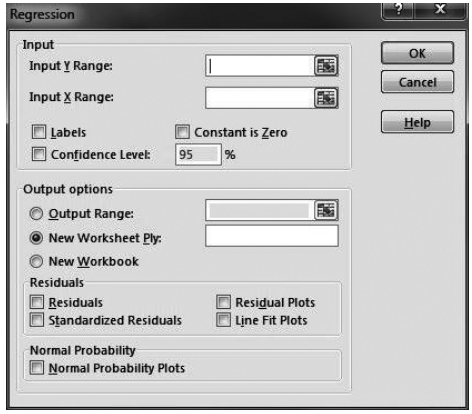
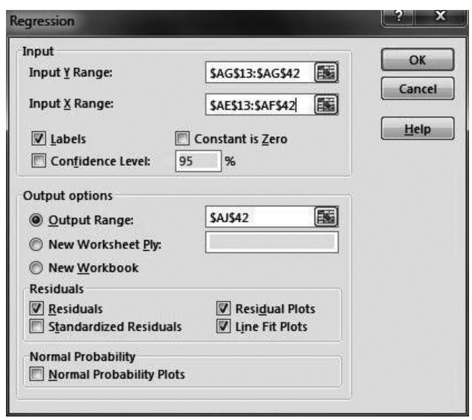


FIGURE 30.25
Excel Data Analysis ToolPak: Regression Dialog With Input and Output Ranges and Options Selected



Gallons Per Cap/Lag 1) as our input data. Under Input, we checked Labels, which means that the top row of the data contains labels for each column. Under Output, we checked the boxes for Residuals, Residual Plots, and Line Fit Plots, which all affect the output that will be produced.

We specified an output range on the same page we began. If we click the second radio button under Output (New Worksheet Ply:) and do not give a name, a generically named new sheet with the output will appear immediately to the left of the tab of the current sheet in the workbook.

Figure 30.26 shows the output when first generated. The charts may appear anywhere on the active part of the worksheet page. It is up to the user to format the sheet to make it usable.

FIGURE 30.26

Water and Sewer Data: Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9912
R Square	0.982477
Adjusted F	0.981129
Standard E	0.025787
Observatio	29

ANOVA

	df	SS	MS	F	ignificance F
Regressior	2	0.969344	0.484672	728.8694	1.47E-23
Residual	26	0.017289	0.000665		
Total	28	0.986633			

	Coefficients	tandard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	11.32724	0.077456	146.2408	1.93E-39	11.16803	11.48646	11.16803	11.48646
Log Rev P	0.986702	0.046442	21.24587	5.9E-18	0.891239	1.082165	0.891239	1.082165
Thousand	0.00531	0.001136	4.676132	7.9E-05	0.002976	0.007645	0.002976	0.007645

RESIDUAL OUTPUT

Observatior	ted Log Re	Residuals
1	8.413584	0.004318
2	8.461617	-0.03906
3	8.471804	0.032296
4	8.545698	-0.00808
5	8.574642	0.00309
6	8.611337	0.037597
7	8.684915	-0.04368
8	8.675021	-0.03688
9	8.66839	0.06851
10	8.763557	-0.00664

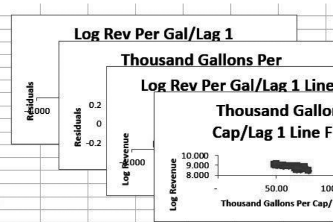


FIGURE 30.27

Water and Sewer Data: Regression Output, Formatted

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9912
R Square	0.9825
Adjusted R Square	0.9811
Standard Error	0.0258
Observations	29

ANOVA

	df	SS	MS	F	Significance F
Regression	2	0.9693	0.4847	728.9	0.0000
Residual	26	0.0173	0.0007		
Total	28	0.9866			

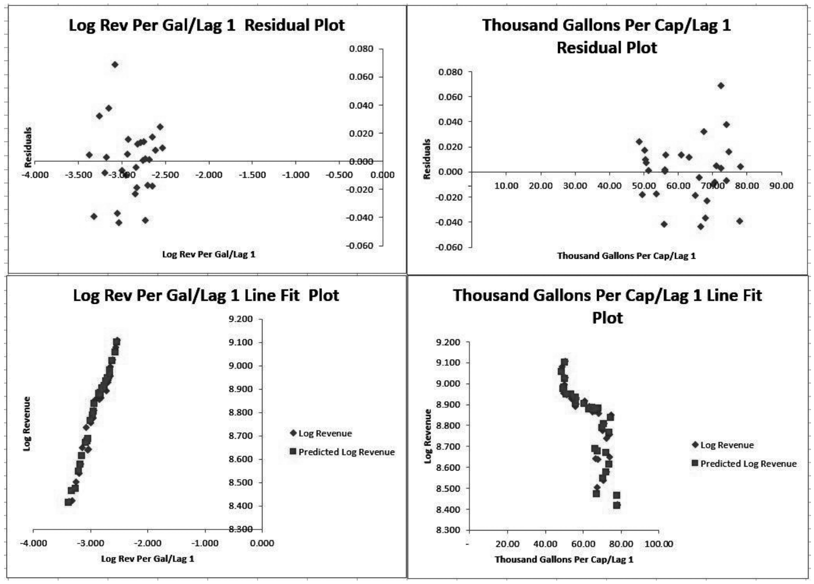
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	11.3272	0.077	146.24	0.0000	11.168	11.486	11.168	11.486
Log Rev Per Gal/Lag 1	0.9867	0.046	21.25	0.0000	0.891	1.082	0.891	1.082
Thousand Gallons Per Cap/Lag 1	0.0053	0.001	4.68	0.0001	0.003	0.008	0.003	0.008

Figure 30.27 shows the statistical output after formatting. The user is interested in the adjusted *R* square, the significance of *F*, and the coefficients and the *p*-values of the *t* statistics. *R* square and adjusted *R* square cannot exceed 1 and generally should not appear to be approaching 1 unless there are very good reasons.

We have attempted to remove spurious[5] reasons for an extremely high *R* square by converting dollar values to real dollars and logging the values. However, if values are this high, then the

5. Spurious correlation arises when something not included in the model is the reason for the appearance of correlation. For time series regression, spurious correlation often results from the simple tendency of things moving in time to increase in size. When time series regression models have input (*X*) and output (*Y*) variables that increase in time, they often attain very high *R* square without having a particularly good explanation of why. In a related matter, the analyst should never use a time index as an independent (*X*) variable, as it is a sure way to import spurious correlation.

FIGURE 30.28
Water and Sewer Data: Regression Plots



analyst should be suspicious. After reviewing the remainder of the output from this model, we will look at a different model.

Figure 30.28 shows us the plots we asked for. The analyst must reformat some to see them this way, and they may still be obscure. For the two on the top, we are looking for the dots to be randomly distributed above and below the center line. A visual inspection shows no reason to doubt this, although the distant dot shown in both plots suggests there might be an outlier. An **outlier** is a value that is so far from the rest of the data that it may cause the coefficients to be incorrectly estimated.

In the box labeled 1 in Figure 30.29, we have the predicted values for the original observations and the residuals. *Residual* is the regression word for what we have been calling error, the difference between the predicted value and the original value.

In box 2 of Figure 30.29, I have calculated the Durbin Watson statistic (DW). When the regression is of time-ordered data, DW should always be calculated. In this text, we do not extensively discuss DW, but briefly, the point of DW is to find out whether the residuals are correlated with each other. If they are, we must correct the regression. An example is given in Appendix D. DW has a maximum value of 4. When the number of observations is reasonably large and the number of **independent variables** is reasonably small, values above 2 are acceptable. If you get values below 2, you should learn more about DW or use the correction shown in Appendix D. If you use a large number of independent variables (more than 5), you should also learn more about DW.

In Box 3 of Figure 30.29, we reverse the logging to get predicted values in the same form as the original data.

FIGURE 30.29

Water and Sewer Data: Regression Predictions, Residuals, Durbin Watson Statistic, and Antilogged Predictions

RESIDUAL OUTPUT		1	Durbin Watson	2		3
			2.587			
Observation	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		Antilog Predicted
1	8.413583615	0.004317728		0.000019		259,169,335
2	8.461617292	-0.039055397	0.00188	0.001525		289,479,152
3	8.471803603	0.032296194	0.00509	0.001043		296,349,094
4	8.545698155	-0.008083804	0.00163	0.000065		351,316,183
5	8.574642447	0.003089741	0.00012	0.000010		375,528,106
6	8.61133709	0.037596509	0.00119	0.001413		408,636,438
7	8.68491548	-0.043683992	0.00661	0.001908		484,078,149
8	8.67502101	-0.036884799	0.00005	0.001360		473,174,149
9	8.668389605	0.068510421	0.01111	0.004694		466,003,958
10	8.763557413	-0.006635274	0.00565	0.000044		580,172,865
11	8.785346462	-0.009825016	0.00001	0.000097		610,023,354
12	8.803933653	0.005035811	0.00022	0.000025		636,698,245
13	8.834869818	0.015851049	0.00012	0.000251		683,706,672
14	8.879029898	-0.022998817	0.00151	0.000529		756,884,999
15	8.872915551	-0.004528954	0.00034	0.000021		746,303,625
16	8.882886795	-0.018991278	0.00021	0.000361		763,636,705
17	8.877487287	0.011992724	0.00096	0.000144		754,201,318
18	8.901864079	0.01343029	0.00000	0.000180		797,744,978
19	8.932698672	-0.0419134	0.00306	0.001757		856,443,410
20	8.903279598	0.000491219	0.00180	0.000000		800,349,353
21	8.911924494	0.013658273	0.00017	0.000187		816,440,414
22	8.931454824	0.001985186	0.00014	0.000004		853,994,010
23	8.944884967	-0.017334072	0.00037	0.000300		880,815,538
24	8.94586363	0.000954478	0.00033	0.000001		882,802,654
25	8.971725671	-0.017809467	0.00035	0.000317		936,969,968
26	8.978408654	0.017027011	0.00121	0.000290		951,499,697
27	9.019314849	0.007575113	0.00009	0.000057		1,045,477,882
28	9.055568072	0.024405062	0.00028	0.000596		1,136,496,421
29	9.098870192	0.009527461	0.00022	0.000091		1,255,654,599

In Figures 30.30–30.34, we see how to calculate DW, and in Figure 30.35 we see how to reverse the log.

FIGURE 30.30

Durbin Watson Statistic: Difference of Residuals, Squared

AM68 | =(AL68-AL67)^2

	AK	AL	AM	AN	AO	AP
64			Durbin Watson			
65			2.587			
66	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		**Antilog Predicted**
67	8.414	0.004		0.000019		259,169,335
68	8.462	-0.039	0.00188	0.001525		289,479,152
69	8.472	0.032	0.00509	0.001043		296,349,094

Figure 30.30 shows part of how to calculate DW. We build a column of values that are the residual minus the residual on the row above squared:

$$d_t = \left(e_t - e_{t-1}\right)^2$$

FIGURE 30.31
Durbin Watson Statistic: Residual Squared

AN68 | =AL68^2

	AK	AL	AM	AN	AO	AP
64			Durbin Watson			
65			2.587			
66	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		**Antilog Predicted**
67	8.414	0.004		0.000019		259,169,335
68	8.462	-0.039	0.00188	0.001525		289,479,152
69	8.472	0.032	0.00509	0.001043		296,349,094

Figure 30.31 shows the next part of DW. We build a column of values that are the residual squared:

$$w_t = e_t^2$$

FIGURE 30.32
Durbin Watson Statistic:
Total of Difference of Residuals, Squared (*D*)

AM66 | =SUM(AM67:AM95)

	AK	AL	AM	AN	AO	AP
64			Durbin Watson			
65			2.587			
66	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		**Antilog Predicted**
67	8.414	0.004		0.000019		259,169,335
68	8.462	-0.039	0.00188	0.001525		289,479,152
69	8.472	0.032	0.00509	0.001043		296,349,094

Figure 30.32 shows that we total the first column as follows:

$$D = \sum d_t$$

FIGURE 30.33
Durbin Watson Statistic: Total of Residual Squared (*W*)

AN66 | =SUM(AN67:AN95)

	AK	AL	AM	AN	AO	AP
64			Durbin Watson			
65			2.587			
66	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		**Antilog Predicted**
67	8.414	0.004		0.000019		259,169,335
68	8.462	-0.039	0.00188	0.001525		289,479,152
69	8.472	0.032	0.00509	0.001043		296,349,094

Figure 30.33 shows that we total the second column as follows:

$$W = \sum w_t$$

FIGURE 30.34
Durbin Watson Statistic: Divide *D* by *W*

AM65 | f_x =AM66/AN66

	AK	AL	AM	AN	AO	AP
64			Durbin Watson			
65			2.587			
66	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		**Antilog Predicted**
67	8.414	0.004		0.000019		259,169,335
68	8.462	-0.039	0.00188	0.001525		289,479,152
69	8.472	0.032	0.00509	0.001043		296,349,094

Figure 30.34 shows that we divide the total of the first column by the total of the second column:

$$DW = \frac{D}{W}$$

The result of this calculation is the Durbin Watson statistic.

FIGURE 30.35
Antilog of Predicted Revenue

AP67 | f_x =10^AK67

	AK	AL	AM	AN	AO	AP
64			Durbin Watson			
65			2.587			
66	*Predicted Log Revenue*	*Residuals*	0.04473	0.01729		**Antilog Predicted**
67	8.414	0.004		0.000019		259,169,335
68	8.462	-0.039	0.00188	0.001525		289,479,152
69	8.472	0.032	0.00509	0.001043		296,349,094

Figure 30.35 shows the calculation of an antilog. The antilog is the base number (we used the default base of 10) raised to the power of the log. In Excel, “raised to the power” is expressed with the caret symbol (^) found on the 6 key at the top of the keyboard.

Figure 30.36 shows the inputs to the regression equation (located in the bottom left corner of Figure 30.27). The equation can be read as follows:

$$\text{Log Rev} = 11.3272 + 0.9867 \times (\text{Log Rev Per Gal/Lag1}) + 0.0053 \times (\text{Thousand Gallons Per Cap/Lag 1})$$

FIGURE 30.36
Water and Sewer Data: Input to Regression Equation

	Coefficients
Intercept	11.3272
Log Rev Per Gal/Lag 1	0.9867
Thousand Gallons Per Cap/Lag 1	0.0053

Because we noticed a potential problem with the model when reviewing Figure 30.27 (excessively high regression results indicating possible spurious correlation), we will not use this model to make predicted values. Instead, we will examine a differenced model.

FIGURE 30.37
Water and Sewer Data: Differenced Data

Year	Differenced Rev Per 1000 Gal/Lag 1	Differenced Thousand Gallons Per Cap/Lag 1Corrected	Differenced Thousand Gallons Per Cap/Lag 1	Differenced Revenue
1982	0.051	-0.185	-0.185	2,824,144
1983	0.079	-9.328		54,644,158
1984	0.079	3.149	3.149	25,610,258
1985	0.029	1.825	1.825	33,371,895
1986	0.046	1.537	1.537	67,378,830
1987	0.213	-7.431	-7.431	(7,832,741)
1988	-0.035	1.290	1.290	(3,108,851)
1989	-0.062	4.635	4.635	110,985,719
1990	0.184	1.630	1.630	25,743,947
1991	0.106	-3.995	-3.995	25,001,579
1992	0.034	1.086	1.086	47,746,205
1993	0.032	3.619	3.619	64,997,875
1994	0.236	-6.420	-6.420	8,723,814
1995	0.021	-2.365	-2.365	20,715,723
1996	0.053	-1.057	-1.057	(7,598,180)
1997	0.016	-1.886	-1.886	44,355,052
1998	0.130	-2.094	-2.094	47,481,906
1999	0.233	-4.950	-4.950	(45,148,202)
2000	-0.128	0.167	0.167	23,603,156
2001	0.032	0.157	0.157	41,269,825
2002	0.087	-0.167	-0.167	15,381,660
2003	0.119	-2.465	-2.465	(11,554,856)
2004	0.062	-2.310	-2.310	38,393,235
2005	0.174	-1.744	-1.744	14,579,061
2006	0.019	0.587	0.587	90,221,222
2007	0.213	0.345	0.345	74,328,165
2008	0.279	-1.888	-1.888	138,316,634
2009	0.221	1.855	1.855	81,315,197

Figure 30.37 shows the data used for estimating the differenced model. **Differencing** is simply subtracting the earlier value of the serial variable from the next later value, $X_t - X_{t-1}$. In Module 28, we calculated trend units by differencing. The second column from the left is missing an observation, for 1983, which is the outlier we observed when inspecting the earlier graphs. We did not use that column in making the forecast. Excel will cancel a Data Analysis

ToolPak regression that has an empty cell. We used that column to Windsorize the data (as explained in Module 28) and produce an adjusted column immediately to the left. If we had the resources, we would likely want to find out what led to this seemingly bizarre value, and we might want to have the regression start with data in the subsequent year.

FIGURE 30.38

Water and Sewer Data: Regression Using Differenced Data

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.403912							
R Square	0.163145							
Adjusted R Square	0.096197							
Standard Error	37828179							
Observations	28							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	6974210828088330	3487105414044170	2.43688035	0.10792608			
Residual	25	35774278163315400	1430971126532620					
Total	27	42748488991403700						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	35755354	9781961	3.66	0.001	15609028	55901679	15609028	55901679
Differenced Rev Per 1000 Gal/Lag 1	64618347	81311912	0.79	0.434	-102846670	232083364	-102846670	232083364
Differenced Thousand Gallons Per Cap/Lag 1Corrected	5400148	2452971	2.20	0.037	348160	10452137	348160	10452137

The *t* statistic beside Differenced Rev Per 1000 Gal/Lag 1 confirms that the outlier is problematic. Outliers can be much more problematic when working with differenced data. The next step is to make a model excluding the data from 1982 and 1983.

FIGURE 30.39

Water and Sewer Data: Regression Using Differenced Data, Without Outlier

SUMMARY OUTPUT								
Regression Statistics							DW	
Multiple R	0.558329						2.264	
R Square	0.311731							
Adjusted R Square	0.251882							
Standard Error	35145134.68							
Observations	26							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	1.29E+16	6.43E+15	5.2086	0.01362130			
Residual	23	2.84E+16	1.24E+15					
Total	25	4.13E+16						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	32414502	9539327	3.40	0.002	12680900	52148103	12680900	52148103
Differenced Rev Per 1000 Gal/Lag 1	113400509	79483046	1.43	0.167	-51022699	277823717	-51022699	277823717
Differenced Thousand Gallons Per Cap/Lag 1	8887837	2761438	3.22	0.004	3175368	14600307	3175368	14600307

Figure 30.39 shows the regression after omitting the first two observations so that the outlier is not included. The input data can be seen on the left side of Figure 30.42. Here we see that the *t* statistic on the Differenced Rev Per 1000 Gal/Lag 1 is still statistically insignificant, although substantially improved. Having poor relative *t* statistics is a known effect of using differenced models, and if there are other good reasons for retaining the variable (in this case, the expectation that revenue will be related to the amount of service delivered), then one may do so. While this model is sharply less extraordinary than the original model, the analyst can be confident that spurious time indexing correlation is not included. The DW statistic (it is another user

adjustment to put it on this part of the page) is above 2, so concerns over serial correlation of errors are excluded.

FIGURE 30.40

Water and Sewer Data: Plots of Regression Using Differenced Data

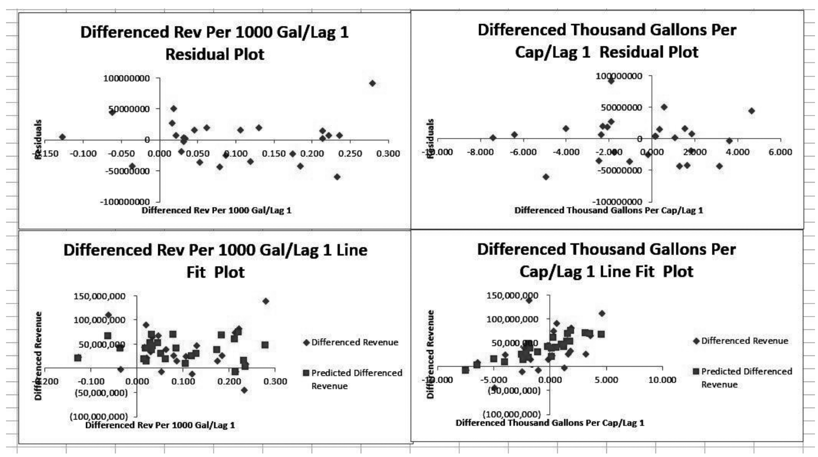


Figure 30.40 shows the plotted residuals and fit. The residual plots show no particular pattern. The model fit plots are satisfactory.

FIGURE 30.41

Water and Sewer Data: Regression Model Using Differenced Data, Without Outlier

	Coefficients
Intercept	32414502
Differenced Rev Per 1000 Gal/Lag 1	113400509
Differenced Thousand Gallons Per Cap/Lag 1	8887837

Figure 30.41 blows up a part of Figure 30.39 and shows the regression model. It can be read as follows:

$$\text{Differenced Revenue} = 32{,}414{,}502 + 113{,}400{,}509 \times (\text{Differenced Rev Per 1000 Gal/Lag 1}) + 8{,}887{,}837 \times (\text{Differenced Thousand Gallons Per Cap/Lag 1})$$

Figure 30.42 shows the implementation of the regression model in Excel to compute predicted values in column BA. The Excel formula is shown in the formula bar. The use of anchors

FIGURE 30.42

Water and Sewer Data: Predicted Values

BA40 =$AZ$43+$AZ$44*AP42+$AZ$45*AQ42

	AW	AX	AY	AZ	BA	BB
13	Year	Rev Per 100	Differenced Thousand Gallons Per Cap/Lag 1	Differenced Revenue	Predicted (Change to Next year)	
14	1984	0.079	3.149	25,610,258	69,362,260.6	
15	1985	0.029	1.825	33,371,895	51,933,573.2	
16	1986	0.046	1.537	67,378,830	51,241,334.9	
17	1987	0.213	-7.431	(7,832,741)	(9,416,659.8)	
18	1988	-0.035	1.290	(3,108,851)	39,879,226.2	
19	1989	-0.062	4.635	110,985,719	66,549,161.1	
20	1990	0.184	1.630	25,743,947	67,786,335.2	
21	1991	0.106	-3.995	25,001,579	8,967,243.6	
22	1992	0.034	1.086	47,746,205	45,901,353.4	
23	1993	0.032	3.619	64,997,875	68,183,916.3	
24	1994	0.236	-6.420	8,723,814	2,152,401.2	
25	1995	0.021	-2.365	20,715,723	13,829,165.6	
26	1996	0.053	-1.057	(7,598,180)	29,053,291.2	
27	1997	0.016	-1.886	44,355,052	17,480,858.3	
28	1998	0.130	-2.094	47,481,906	28,555,496.5	
29	1999	0.233	-4.950	(45,148,202)	14,886,307.5	
30	2000	-0.128	0.167	23,603,156	19,416,831.4	
31	2001	0.032	0.157	41,269,825	37,445,340.1	
32	2002	0.087	-0.167	15,381,660	40,751,774.9	
33	2003	0.119	-2.465	(11,554,856)	23,977,134.2	
34	2004	0.062	-2.310	38,393,235	18,929,957.4	
35	2005	0.174	-1.744	14,579,061	36,695,060.8	
36	2006	0.019	0.587	90,221,222	39,732,808.1	
37	2007	0.213	0.345	74,328,165	59,678,090.4	
38	2008	0.279	-1.888	138,316,634	47,295,659.8	
39	2009	0.221	1.855	81,315,197	74,010,205.7	
40	2010	0.298	-1.922		49,173,849.5	
41						
42		Regression Model				
43			Intercept	32414502		
44			Dif Rev per Tho Gal	113400509		
45			Dif Tho Gal/Cap Lag1	8887837		

($) in the formula makes it possible to write it once and copy it to the entire column. The entry in row 40 shows the use of this formula to predict the value for 2010. The analyst should recognize that because this is a differenced model, the value is the expected change (increase or decrease) from 2009. Stochastically, it is the change from the midpoint of the local series, so it would be good to have a three-period average of the whole data series for 2007 to 2009 and add three increments of this number to that value to get the estimate of the expected 2010 value; however, the ordinary use is to treat this as the expected change from the actual 2009 value. The data in columns AX and AY are manually lagged by 1 year, which means that the data shown in the 2010 row are from 2009. This is the forecasting purpose, making the projection before the year begins. In real-world experience, when using annual data, the current year will not be available, or near the end of the year, there may be preliminary estimates of the current-year data. This means that the lag may need to be more than one period. In Appendix D, there is an example with a 2-year lag.

Summary

This module reviews two critical skills, deseasonalization and its reverse, and the use of simple cross-sectional regression for forecasting. These advanced intermediate techniques round out the forecasting skills the analyst will use. Where these techniques and those of Modules 28 and 29 do not meet the analyst's needs, it is likely best to obtain the services of professional forecasters.

Assignments

1. Businesses in the city of Techville are required to remit their sales tax collections every month to the city tax department. To ensure that business owners are in compliance, the city assesses a penalty on businesses that are out of compliance with this regulation at the beginning of each quarter. The exercise spreadsheet includes the revenue data for the collection of sales tax receipts from businesses.

   a. Prepare a graph of the revenue collection from July 2006 to June 2013. Does there appear to be seasonality in the data?
   b. Prepare an overlay graph of the revenue collection to validate your assumption.
   c. Go through the appropriate steps to deseasonalize the data in preparation for the next budget cycle forecasting. Upon reaching the step of having deseasonalized data for July 2006 to June 2013, prepare a chart similar to Figure 30.20 showing the original data with the deseasonalized data.

2. Valley County is developing a forecast for its sales tax revenue. The data they believe will increase the accuracy of the forecast include the number of single-family home permits, the average monthly rent for an apartment in a multifamily facility, and the employment levels in the county.

   a. Using the data provided and the regression functionality in Excel, develop a model (formula) for predicting sales tax revenues.
   b. Calculate the Durbin Watson statistic. Are the residuals correlated with each other, and why?
   c. Reverse the log of the predicted value (sales tax).
   d. Provide the predictive formula derived from these calculations.
   e. Calculate the predicted change in sales tax revenue next year using differences.

References

Armstrong, J. S. (2001). Extrapolation of time-series and cross-sectional data. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 217–244). Boston, MA: Kluwer Academic.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). New York, NY: John Wiley & Sons.

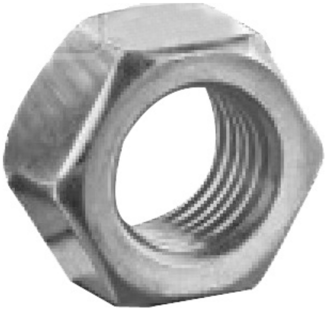
Miller, D. M., & Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting, 19*(4), 669–684.

Miller, D. M., & Williams, D. W. (2004). Damping seasonal factors: Shrinkage estimators for the X-12-ARIMA program. *International Journal of Forecasting, 20*(4), 529–549.

Additional Readings

Makridakis, S., & Wheelwright, S. C. (Eds.). (1983). *The handbook of forecasting: A manager's guide* (2nd. ed.). New York, NY: Wiley.

Williams, D. W. (2008). Seasonality. In J. Rabin & E. M. Berman (Eds.), *Encyclopedia of public administration and public policy* (pp. 1746–1756). New York, NY: Taylor & Francis.



APPENDIX D
Regression

FIGURE D.1

Income Data Logged in Base 10

AC37

	X	Y	Z	AA
2		Lag2	Lag2	
3	**Year**	**Log Real PCI**	**Log Population**	**Log Real Personal Income Tax $M**
4	**1982**	4.476938	6.84952	3.4317747
5	**1983**	4.481991	6.849866	3.4779882
6	**1984**	4.492074	6.850545	3.5248382
7	**1985**	4.508138	6.854315	3.5607662
8	**1986**	4.532095	6.857229	3.5712732
9	**1987**	4.544552	6.859305	3.6317766
10	**1988**	4.563479	6.861948	3.5989088
11	**1989**	4.574937	6.862872	3.64695
12	**1990**	4.59547	6.86272	3.6402061
13	**1991**	4.605035	6.864141	3.6647602
14	**1992**	4.610721	6.864663	3.7146136
15	**1993**	4.592361	6.863589	3.7330381
16	**1994**	4.597581	6.863614	3.7321345
17	**1995**	4.59168	6.865049	3.7255664
18	**1996**	4.592954	6.879122	3.7496786
19	**1997**	4.602145	6.882698	3.7877845
20	**1998**	4.611241	6.886332	3.8519319
21	**1999**	4.623885	6.890613	3.8729442
22	**2000**	4.637028	6.895326	3.8651269
23	**2001**	4.64394	6.900239	3.8937452
24	**2002**	4.655396	6.904096	3.7964902
25	**2003**	4.653781	6.906504	3.7887932
26	**2004**	4.646625	6.906981	3.8589124
27	**2005**	4.646723	6.90677	3.9186986
28	**2006**	4.663095	6.905438	3.9520633
29	**2007**	4.675289	6.903815	3.972282
30	**2008**	4.696588	6.902759	4.0158749
31	**2009**	4.715278	6.903837	3.9046312
32	**2010**	4.712812	6.906776	3.8938905
33	**2011**	4.690242	6.910175	3.9120079
34	**2012**	4.704746	6.913095	
35	**2013**	4.708642	6.916186	

Figure D.1 shows income data that were previously adjusted for inflation in Module 28: Basic Forecasting Concepts. Now all three columns of data have been logged (base 10) to remove an expected upward turning curve in the data.

FIGURE D.2

Regression of Logged Income Data

SUMMARY OUTPUT

DW
0.895

Regression Statistics	
Multiple R	0.945804178
R Square	0.894545543
Adjusted R Square	0.886734102
Standard Error	0.052225702
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.624697619	0.312348809	114.5173493	6.47704E-14
Residual	27	0.073643146	0.002727524		
Total	29	0.698340765			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-22.13516495	5.76058762	-3.842518578	0.000670314	-33.95491442	-10.31541548	-33.95491442	-10.31541548
Log Real PCI	1.331813209	0.348227407	3.824550229	0.000702692	0.617309589	2.04631683	0.617309589	2.04631683
Log Population	2.871039471	1.042470356	2.754073011	0.010402661	0.732066983	5.010011959	0.732066983	5.010011959

Figure D.2 shows preliminary output from the regression. The Durban Watson statistic of 0.895 indicates significant serial correlation in the error. This means that the forecast model is defective and cannot be used.

FIGURE D.3

Regression of Logged Income Data: Residual Plots

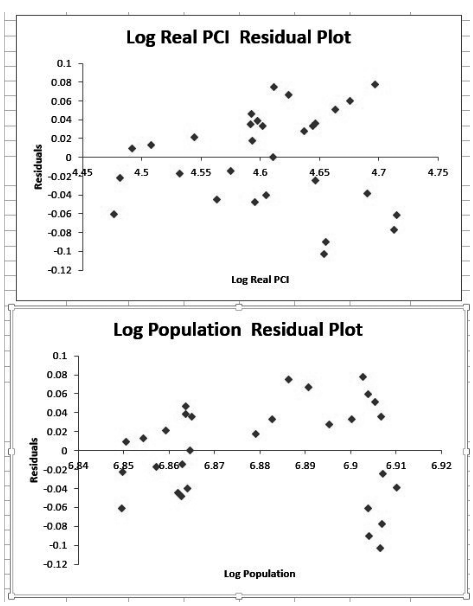


Figure D.3 shows the residual plots. The log population residual plot shows that across this axis, the error exhibits an inverse U shape. As the population rises, the error first tends to be positive and then tends to be negative. Because the population is correlated with time (even when logged), this implies the same shape over time. We cannot fix this by logging the data—we have already logged the data. We could try differencing, which might result in an improvement. However, the Durbin Watson statistic suggests an alternate solution, shown in Figure D.4.

FIGURE D.4
Regression of Logged Income Data: Calculation of *R*

AH66 | =AE66*AE65

	AE	AF	AG	AH
62		Durbin Watson		ρ
63		0.895		0.517215754
64	*Residuals*	0.06592	0.07364	0.03809
65	-0.060747544		0.003690	
66	-0.02225873	0.00148	0.000495	0.0014
67	0.009213755	0.00099	0.000085	-0.0002

Figure D.4 shows a step in the calculation of *R*, which is an estimator of ρ. What we do is multiply each residual (error) by the error in the row before.

$$e_t \times e_{t-1}$$

There is no entry in the first row; the formula is copied down the column to compute the multiplication for the remaining output residuals. Afterward, we total these as shown in Figure D.5.

FIGURE D.5
Regression of Logged Income Data: Total *R*/Total *W*

AH63 | =AH64/AG64

	AE	AF	AG	AH
62		Durbin Watson		ρ
63		0.895		0.517215754
64	*Residuals*	0.06592	0.07364	0.03809
65	-0.060747544		0.003690	
66	-0.02225873	0.00148	0.000495	0.0014
67	0.009213755	0.00099	0.000085	-0.0002

Figure D.5 shows that after we have totaled $e_t \times e_{t-1}$, we divide the total by the squared errors (the *W* in the calculation of the Durban Watson statistic).

FIGURE D.6

Logged Income Data: Adjusted

AP4 =Y5-Y4*$AH$63

	AP	AQ	AR
3	Log Real PCI Adjusted	Log Population Adjusted	Log Real Personal Income Tax $M Adjusted
4	2.1664	3.3072	1.7030
5	2.1739	3.3077	1.7260
6	2.1848	3.3111	1.7377

Figure D.6 shows that we use the *R* (the quantity calculated in the previous steps) to adjust the data (Canavos, 1984, pp. 446-453). The formulas are as follows:

$$y'_t = y_t - r \times y_{t-1}$$

$$x'_t = x_t - r \times x_{t-1}$$

When we use these formulas, we remove some of the serial correlation in the error, and we lose one observation at the beginning of the series.

FIGURE D.7

Regression of Adjusted Logged Income Data

SUMMARY OUTPUT

DW
1.577

Regression Statistics	
Multiple R	0.821441853
R Square	0.674766717
Adjusted R Square	0.649748772
Standard Error	0.042193697
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.096034476	0.048017238	26.97130886	4.55549E-07
Residual	26	0.046288009	0.001780308		
Total	28	0.142322485			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-12.25136027	3.839651647	-3.190747858	0.003686289	-20.14387726	-4.358843275	-20.14387726	-4.358843275
Log Real PCI Adjusted	0.718930009	0.516250518	1.392599104	0.175536915	-0.342238129	1.780098147	-0.342238129	1.780098147
Log Population Adjusted	3.753643146	1.43529929	2.61523375	0.014646268	0.803343201	6.703943091	0.803343201	6.703943091

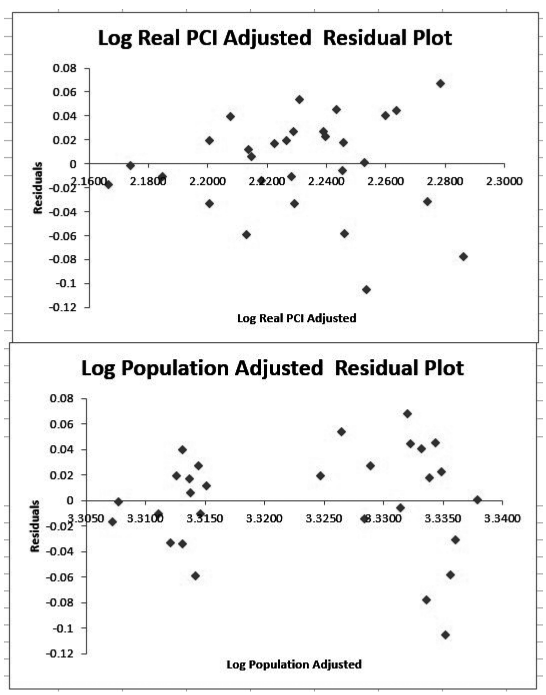
Figure D.7 shows the output of the regression with the modified data. The Durbin Watson statistic of 1.577 still suggests possible serial correlation, but it is sharply improved over 0.895. The analyst may want to consider other ways to further remove serial correlation. Differencing (of the original data, not the modified data) is a possible solution. For the purposes of this discussion, we treat this analysis as improved, if not solved, in order to show the completion of this procedure.

The *R* square and the *t* statistics are reasonably good, although as in the chapter example, there is a coefficient with a relatively poor *p*-value. It is retained because it makes sense that this variable belongs in the model. In real-world application, the analyst may spend considerable effort examining why this statistic has a poor result.

What can be done with Excel is review the residual plots in Figure D.8. The residual plots show two things. First there is a suspect negative-valued outlier between the values 2.24 and 2.26. Second, the residuals are in a funnel shape. The funnel shape is called ***heteroskedasticity***, which means that the model could be improved through the identification of missing variables or through technical management of the model[1] with more sophisticated software. Excel does not provide technical management. The funnel shape is due to the fact that the errors continue to lie more above the middle than below, offset by the weight of the distant outlier. The best way the analyst could improve this model at this point is to identify the outlier and determine whether there is some explanatory variable that could be included in the model that would sharply reduce the effect of this and possibly other unusual observations. In fact, the outlier is associated with fiscal year 2002, and the explanation may be the fiscal distress that followed the September 11, 2001, terrorist attack. Because the city also experienced other fiscal distress following the economic crash at the end of that decade, it may be that the model would benefit from an indicator variable (valued 0 when not in use and 1 when in use) for economic distress.

FIGURE D.8

Regression of Adjusted Logged Income Data: Residual Plots



1. Such as use of "robust standard errors."

Figure D.9 shows the addition of the distress variable. It is generally valued zero, but it is valued 1 for years known to reflect severe fiscal distress after a crisis. The distress variable is not used to model ordinary cyclical activity. The first modeled event is a singular crisis, and the second one has occurred, in recent historical times, in an 80-year cycle. Ordinary cycles are not modeled but possibly should be modeled as an improvement.

FIGURE D.9

Adjusted Logged Income Data With Distress Variable

Year	Log Real PCI Adjusted	Log Population Adjusted	Distress	Log Real PIT $M Adjusted
1983	2.1664	3.3072	0	1.7030
1984	2.1739	3.3077	0	1.7260
1985	2.1848	3.3111	0	1.7377
1986	2.2004	3.3121	0	1.7296
1987	2.2005	3.3126	0	1.7847
1988	2.2130	3.3142	0	1.7205
1989	2.2146	3.3138	0	1.7855
1990	2.2292	3.3131	0	1.7539
1991	2.2282	3.3146	0	1.7820
1992	2.2289	3.3144	0	1.8191
1993	2.2076	3.3131	0	1.8118
1994	2.2223	3.3137	0	1.8013
1995	2.2137	3.3151	0	1.7952
1996	2.2181	3.3284	0	1.8228
1997	2.2266	3.3247	0	1.8484
1998	2.2309	3.3265	0	1.8928
1999	2.2389	3.3289	0	1.8807
2000	2.2455	3.3314	0	1.8620
2001	2.2456	3.3339	0	1.8946
2002	2.2535	3.3352	1	1.7826
2003	2.2459	3.3356	1	1.8252
2004	2.2396	3.3348	0	1.8993
2005	2.2434	3.3344	0	1.9228
2006	2.2597	3.3331	0	1.9253
2007	2.2635	3.3322	0	1.9282
2008	2.2785	3.3320	1	1.9613
2009	2.2861	3.3336	1	1.8276
2010	2.2740	3.3360	1	1.8744
2011	2.2527	3.3379	0	1.8980

Figure D.10 shows the new model statistics. The R square values and t test p-values are in the reasonable range. The Durbin Watson statistic is also reasonable.

Figure D.11 shows plots of the residuals. In general, these residuals are reasonably well distributed. There is a suggestion of one possible outlier, which may be one of the five observations included among the distress years. Further investigation may improve this model more. However, the model is not defective.

Figure D.12 shows using the regression output to produce predicted values. As this model is lagged 2 years, the last two observations are forecasts from the perspective of the data. The forecaster may be in fiscal year 2012 while making this forecast, so the budgetary forecast may be for 2013. The Excel formula for the forecast is shown in the formula bar and involves multiplying the coefficients by the variables and adding across (including adding the intercept).

Figure D.13 shows the retransformation of the data to remove the adjustment made in Figure D.6, using the same value R, which we estimated in Figures D.4 and D.5.

FIGURE D.10

Regression of Adjusted Logged Income Data With Distress Variable

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8857
R Square	0.7844
Adjusted R Square	0.7586
Standard Error	0.0350
Observations	29

DW
2.208

ANOVA

	df	SS	MS	F	Significance F
Regression	3	0.11164	0.03721	30	0.0000
Residual	25	0.03068	0.00123		
Total	28	0.14232			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-13.151	3.198	-4.112	0.000	-19.737	-6.565	-19.737	-6.565
Log Real PCI Adjusted	1.348	0.464	2.908	0.008	0.393	2.303	0.393	2.303
Log Population Adjusted	3.606	1.192	3.024	0.006	1.150	6.062	1.150	6.062
Distress	-0.076	0.021	-3.566	0.001	-0.120	-0.032	-0.120	-0.032

FIGURE D.11

Regression of Adjusted Logged Income Data With Distress Variable: Residual Plots

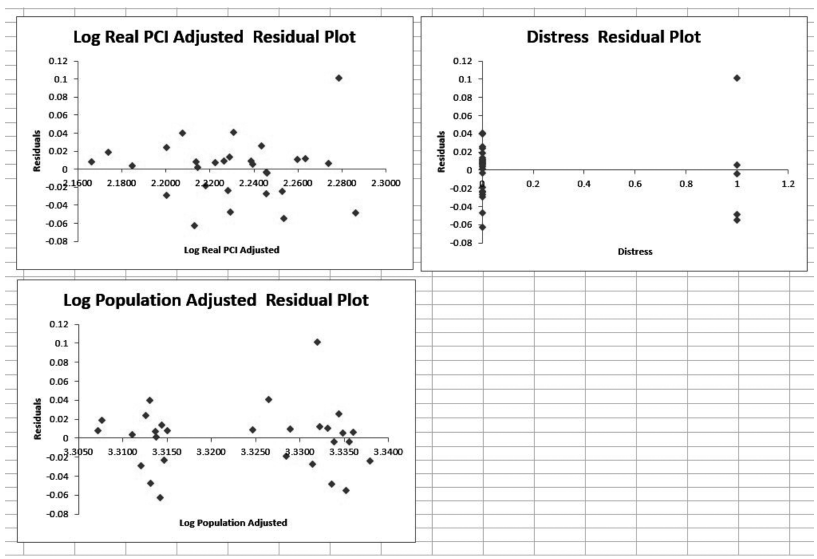


Figure D.14 shows taking the antilog of the retransformed data. Taking the antilog puts the data back in their original form. When this is done for fiscal years 2012 and 2013, the analyst has the desired forecasts.

FIGURE D.12

Predicted Results Based on Regression of Adjusted Logged Income Data With Distress Variable

AT34 =\$AR\$36+\$AR\$37*AP34+\$AR\$38*AQ34+\$AR\$39*AR34

	AO	AP	AQ	AR	AS	AT	AU	AV
3	Year	Log Real PCI Adjusted	Log Population Adjusted	Distress	Log Real PIT $M Adjusted	Predicted	B0'>B0	Antilog
4	1983	2.1664	3.3072	0	1.7030	1.695326189	3.511560707	3,247.6
5	1984	2.1739	3.3077	0	1.7260	1.707196583	3.536148074	3,436.8
6	1985	2.1848	3.3111	0	1.7377	1.734149305	3.591975749	3,908.2
7	1986	2.2004	3.3121	0	1.7296	1.758717891	3.64286512	4,394.1
8	1987	2.2005	3.3126	0	1.7847	1.760862231	3.647306731	4,439.2
9	1988	2.2130	3.3142	0	1.7205	1.783346216	3.693878229	4,941.7
10	1989	2.2146	3.3138	0	1.7855	1.78400043	3.695233314	4,957.2
11	1990	2.2292	3.3131	0	1.7539	1.801417839	3.731310317	5,386.5
12	1991	2.2282	3.3146	0	1.7820	1.805400952	3.739560613	5,489.9
13	1992	2.2289	3.3144	0	1.8191	1.805631984	3.740039154	5,495.9
14	1993	2.2076	3.3131	0	1.8118	1.772072242	3.670526236	4,683.0
15	1994	2.2223	3.3137	0	1.8013	1.794000614	3.71594688	5,199.3
16	1995	2.2137	3.3151	0	1.7952	1.787537478	3.702559668	5,041.5
17	1996	2.2181	3.3284	0	1.8228	1.841438191	3.814205217	6,519.4
18	1997	2.2266	3.3247	0	1.8484	1.839585593	3.810367897	6,462.0
19	1998	2.2309	3.3265	0	1.8928	1.851875159	3.835823503	6,852.1
20	1999	2.2389	3.3289	0	1.8807	1.871238875	3.87593193	7,515.1
21	2000	2.2455	3.3314	0	1.8620	1.88915008	3.913031742	8,185.2
22	2001	2.2456	3.3339	0	1.8946	1.898230304	3.931839778	8,547.5
23	2002	2.2535	3.3352	1	1.7826	1.837705223	3.806473051	6,404.3
24	2003	2.2459	3.3356	1	1.8252	1.829032764	3.788509626	6,144.8
25	2004	2.2396	3.3348	0	1.8993	1.893633542	3.922318421	8,362.2
26	2005	2.2434	3.3344	0	1.9228	1.897101622	3.929501918	8,501.6
27	2006	2.2597	3.3331	0	1.9253	1.914694116	3.965941578	9,245.7
28	2007	2.2635	3.3322	0	1.9282	1.916349834	3.9693711	9,319.0
29	2008	2.2785	3.3320	1	1.9613	1.859885525	3.852415524	7,118.9
30	2009	2.2861	3.3336	1	1.8276	1.876087204	3.885974364	7,690.9
31	2010	2.2740	3.3360	1	1.8744	1.868321101	3.869888289	7,411.2
32	2011	2.2527	3.3379	0	1.8980	1.922279604	3.981653542	9,586.4
33	2012	2.2789	3.3390	0		1.961760984	4.063432059	11,572.6
34	2013	2.2753	3.3406	0		1.962597617	4.065164992	11,618.9
35		Regression Model						
36			Intercept	-13.151				
37			Log Real PCI Adjusted	1.348				
38			Log Population Adjusted	3.606				
39			Distress	-0.076				

FIGURE D.13

Retransformed Income Data

AU34 =AT34/(1-\$AH\$63)

	AO	AP	AQ	AR	AS	AT	AU	AV
3	Year	Log Real PCI Adjusted	Log Population Adjusted	Distress	Log Real PIT $M Adjusted	Predicted	B0'>B0	Antilog
4	1983	2.1664	3.3072	0	1.7030	1.695326189	3.511560707	3,247.6
5	1984	2.1739	3.3077	0	1.7260	1.707196583	3.536148074	3,436.8
31	2010	2.2740	3.3360	1	1.8744	[illegible]	[illegible]	7,411.2
32	2011	2.2527	3.3379	0	1.8980	1.922279604	3.981653542	9,586.4
33	2012	2.2789	3.3390	0		1.961760984	4.063432059	11,572.6
34	2013	2.2753	3.3406	0		1.962597617	4.065164992	11,618.9

FIGURE D.14
Retransformed Income Data With Antilog

AV34 | =10^AU34

	AO	AP	AQ	AR	AS	AT	AU	AV
3	Year	Log Real PCI Adjusted	Log Population Adjusted	Distress	Log Real PIT $M Adjusted	Predicted	B0'>B0	Antilog
4	1983	2.1664	3.3072	0	1.7030	1.695326189	3.511560707	3,247.6
5	1984	2.1739	3.3077	0	1.7260	1.707196583	3.536148074	3,436.8
32	2011	2.2527	3.3379	0	1.6980	1.922270604	3.981663542	9,586.4
33	2012	2.2789	3.3390	0		1.961760984	4.063432059	11,572.6
34	2013	2.2753	3.3406	0		1.962597617	4.065164992	11,618.9

This forecast is made in real (inflation-adjusted) dollars, so it must be multiplied by the expected inflation for future years to get the full forecast.

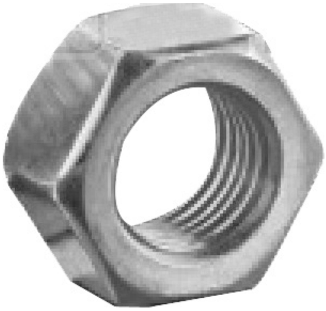
Additional Readings

Armstrong, J. S. (1985). *Long-range forecasting: From crystal ball to computer* (2nd ed.). New York, NY: Wiley.

Armstrong, J. S. (2001). Extrapolation of time-series and cross-sectional data. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 217–244). Boston, MA: Kluwer Academic.

Canavos, G. C. (1984). *Applied probability and statistical methods*. Boston, MA: Little, Brown.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). New York, NY: John Wiley & Sons.



APPENDIX E
Forecasting Techniques Not Demonstrated

In this appendix, we list three types of forecasting techniques not demonstrated in the text: those not to be used, those that can be learned through the literature, and those that require instruction. In addition, there are many variants of methods that have been tried and fallen by the wayside over the years. Methods not mentioned at all are not so much demonstrated as being ineffective as simply not common.

Techniques Not Demonstrated Because They Should Not Be Used

Holt-Winters Exponential Smoothing—This is a method for including seasonality within an exponential smoothing model. The author's anecdotal experience using this method suggests that when using the more common multiplicative version of this method, it can easily become misfit and provide very misleading forecasts for periods beyond the next few. As the purpose is to predict at least one full seasonal cycle, the technique requires more management than it is worth. A more reasonable method is to deseasonalize first, then to forecast.

Time Index Regression—The only purpose of this popular method is to illicitly import spurious correlation.

Seasonal Indicator Variables—Sometimes called "dummy" variables, seasonal indicators can absorb seasonal impact in approximately the correct size, but they will lead to overestimation of the strength of the model (*R* square, adjusted *R* square). Also, using these variables requires more sophistication than demonstrated here to dampen them, as discussed in the section on seasonality. Finally, including both seasonal predictor and predicted variables can lead to spurious correlation similar to time-indexed spurious correlation. The data likely should be deseasonalized.

Techniques to Pursue Through the Literature

Damped Trend—This method is considered an improvement on Holt exponential smoothing by some scholars (McKenzie & Gardner, 2010).

Additive Seasonality—In this book we examined seasonality with ratios, which is commonly known as multiplicative seasonality. When the magnitude of seasonal effect does not depend on the size of the original data series and simultaneously there is a trend, multiplicative seasonality will overstate or understate seasonality over some range of the series. An alternative is additive seasonality, which does not exhibit this difficulty but does exhibit the inverse (failing to account for the size of seasonal effect that does depend on the size of the data series). Most full textbooks on forecasting will address both types of seasonality.

Techniques Requiring Instruction

ARMA/ARIMA

Any technique involving autoregressive (integrated) moving average (AR(I)MA) techniques or any variant, such as vector ARIMA or any other vector method, requires extensive instruction before use.

Advanced Seasonal Methods

Advanced methods that require instruction include the following:

- Census methods known as X-11, X-12, or X-13
- Methods that involve Bayes methods, vectors, or averaging
- Methods involving unit root
- Methods involving SEATS or ARIMA

Adaptive Techniques

Methods involving adaptive techniques (a common term for these is Kalman filters, but there are also other adaptive techniques) requires instruction.

Advanced Regression Methods and Other Advanced Models

Any method that requires the use of any regression model or any advanced technique not demonstrated in this book requires instruction. These include but are not limited to simultaneous equations, neural networks, path analysis, regression-style simulation models, and macroeconomic models.

Reference

McKenzie, E., & Gardner, E. S., Jr. (2010). Damped trend exponential smoothing: A modelling viewpoint. *International Journal of Forecasting, 26*(4), 661–665.