Example 3-2—Regression Analysis of Revenue From Sales of Steam

Background Information

B.15 US SteamCo is a public utility in the northeastern United States. It produces steam and pumps it under high pressure to apartment buildings to provide heat in winter and power for air conditioners in summer. US SteamCo's regulator, the local public service commission, determines pricing, which varies by season and demand, and includes various fees, some fixed, and adjustments for variations in US SteamCo's fuel costs. The schedule is different for November to May (heating season) and for June to October (cooling season). Pricing schedules are revised at least annually.

Set the Objectives for and Plan the Procedure

Financial Statement Item or Account and Related Assertions

B.16 This substantive analytical procedure was intended to be a source of audit evidence regarding revenue earned by the company by providing steam for heating and cooling. The assertions addressed by this procedure were as follows:

- *Occurrence*. All revenue transactions that have been recorded have occurred and pertain to the company.
- *Completeness*. All revenue transactions and events that should have been recorded have been recorded.
- *Accuracy*. Amounts and other data related to revenue transactions and events have been recorded properly.
- *Cutoff*. Transactions and events related to revenue have been recorded in the correct accounting period.

(*Note: "Events" refers to, for example, adjustments to revenues by means of journal entries.*)

Assessed Risk of Material Misstatement

B.17 As required by paragraph .26 of AU-C section 240, the auditor performed procedures appropriate in the circumstances encountered in this audit to respond to the presumption that risks of fraud exist in revenue recognition.

B.18 This analytical procedure was designed to respond to a moderate level of risk of material misstatement of revenue. The level of assurance to be provided by this procedure took into account various matters, such as other audit procedures performed to provide assurance regarding revenues. These procedures included the following:

- Tests of the effective operation of controls over revenues and revenue adjustments (all relevant assertions)
- Confirmation of accounts receivable from customers (occurrence, accuracy, cutoff)
- Various procedures to audit cash receipts regarding amounts billed (occurrence)

Exhibit B-3

US SteamCo: Revenue and Production 2011–2014

		Revenue $	*Production Mlb*	*Cool DD*	*Heat DD*
Base (Training) Data	Jan-11	23,304,096	659,640	0	1,021
	Feb-11	24,552,000	686,144	0	794
	Mar-11	14,804,728	416,464	0	714
	Apr-11	11,265,320	324,456	1	354
	May-11	5,945,904	151,104	72	114
	Jun-11	10,379,712	479,264	239	4
	Jul-11	13,617,528	708,472	486	0
	Aug-11	14,092,512	687,728	346	0
	Sep-11	12,085,472	567,424	195	19
	Oct-11	11,957,032	392,328	24	223
	Nov-11	10,153,800	254,504	0	383
	Dec-11	15,093,880	358,888	0	662
	Jan-12	22,474,848	566,528	0	846
	Feb-12	18,027,296	473,016	0	680
	Mar-12	14,575,856	410,248	0	438
	Apr-12	9,064,488	264,424	18	294
	May-12	6,582,400	217,264	102	69
	Jun-12	10,418,400	499,288	254	11
	Jul-12	14,122,824	732,520	484	0
	Aug-12	14,779,280	725,056	430	0
	Sep-12	12,003,432	608,376	184	9
	Oct-12	4,809,952	369,656	31	173
	Nov-12	7,439,448	192,032	0	591
	Dec-12	18,035,344	420,920	0	681
	Jan-13	20,113,488	531,336	0	899
	Feb-13	23,725,040	587,704	0	851
	Mar-13	20,387,368	505,272	0	755
	Apr-13	17,811,224	427,312	0	366

(*continued*)

US SteamCo: Revenue and Production 2011–2014—*continued*

		Revenue $	*Production Mlb*	*Cool DD*	*Heat DD*
	May-13	6,082,328	155,032	87	128
	Jun-13	12,064,192	394,400	278	4
	Jul-13	16,033,016	749,152	505	0
	Aug-13	14,311,800	642,688	340	0
	Sep-13	12,291,192	594,184	129	31
	Oct-13	9,845,128	372,080	49	172
	Nov-13	8,933,832	292,648	0	577
	Dec-13	17,613,912	562,320	0	826
Projection Data	Jan-14	19,228,840	606,400	0	1,123
	Feb-14	26,792,280	714,128	0	938
	Mar-14	19,935,840	805,600	0	866
	Apr-14	13,468,000	375,856	0	412
	May-14	7,344,128	279,296	49	88
	Jun-14	11,196,216	517,600	230	0
	Jul-14	13,929,472	749,472	380	0
	Aug-14	12,352,176	663,432	322	0
	Sep-14	12,628,944	701,656	178	14
	Oct-14	9,361,000	411,728	24	166
	Nov-14	10,164,048	293,536	0	579
	Dec-14	18,377,456	567,048	0	752

Nature of the Auditor's Expectation

B.19 The auditor's expectation was an estimate of revenue from sales of steam for each month (the dependent or test variable). The independent variable (predictor) was the quantity of steam produced in 2011–2014. Steam production is measured by mass in units of thousands of pounds (Mlb). This was internal data from a source outside the company's financial reporting system. In addition, cooling and heating degree days (DD) data (external data) was used in the preliminary analysis. Degree days measure how many days outside air temperatures were higher or lower than a specified base temperature over a period. Exhibit B-3 shows details regarding these variables.

Desired Precision of the Auditor's Expectation and What Will Be Considered a Significant Difference

B.20 The desired precision for this substantive analytical procedure was performance materiality.

Type of Regression Analysis to Be Used

B.21 A time-series regression analysis was used to develop monthly expectations of revenue from steam sales. The regression model used is described in paragraphs B.29–B.34. The auditor used a regression tool available in widely used software to perform the regression.

B.22 The regression was intended to provide a level of assurance by establishing, in this case, whether steam production was an accurate predictor of revenue.

B.23 In evaluating the level of assurance that a regression analysis may provide, one consideration is how the standard error obtained from the regression compares to the auditor's performance materiality. In this example, if the standard error level obtained is less than performance materiality, this provides further confidence regarding use of the regression. On the other hand, if the standard error is a high percentage of performance materiality, the auditor would consider limiting how much assurance the auditor intends to derive from the regression. Regression analyses that factor in performance materiality, and the auditor's desired level of assurance, would ordinarily adjust achieved assurance for the standard error.

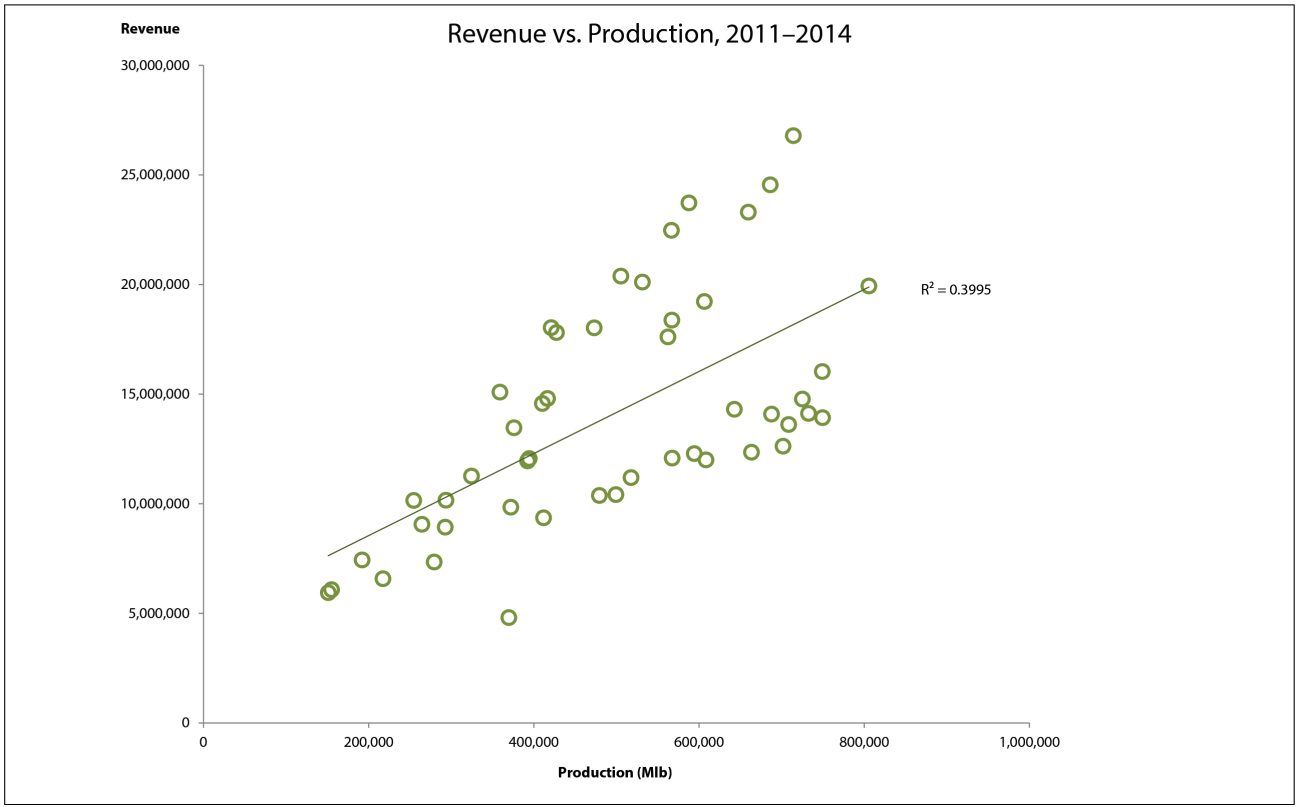
B.24 Residuals in a regression analysis represent the variability in the test variable (in this example, revenue) that is not explained by the regression model. The statistical results of the regression analysis are combined with the results of other procedures using the professional judgment of the auditor. Other procedures might include further statistical analysis of the residuals.

Graphics and Tables

B.25 The auditor decided to use the graphics shown in exhibits B-4–B-12 to show key aspects of developing and applying the regression model.

Exhibit B-4

Scatter Plot of Revenue vs. Production, 2011–2014, With Regression Line



Apply the Regression Analysis

Exploratory Analysis—Understanding the Data

B.26 To start exploring the data, the auditor regressed revenue against production for the entire dataset, 2011–2014, as depicted in exhibit B-4. In this example, the correlation between the variables, measured by R^2 was rather weak relative to the auditor's expectations.[2] The line appears to be averaging between two separate data series. The auditor found that this is precisely what was happening because the relationship between revenue and production differed in heating and cooling months.

[2] R^2 is a number between 0 and 1 and measures the degree to which changes in the dependent variable can be estimated by changes in the independent variables. A more precise regression is one that has a relatively high R^2 (close to 1). Determining an acceptable R^2 is a matter of professional judgment. Most regression analysis involving financial data have R^2 values above .5 and many have values in the .8 to .9 range.

Exhibit B-5

Time Series Comparing Steam Production With Cooling and Heating Degree Days

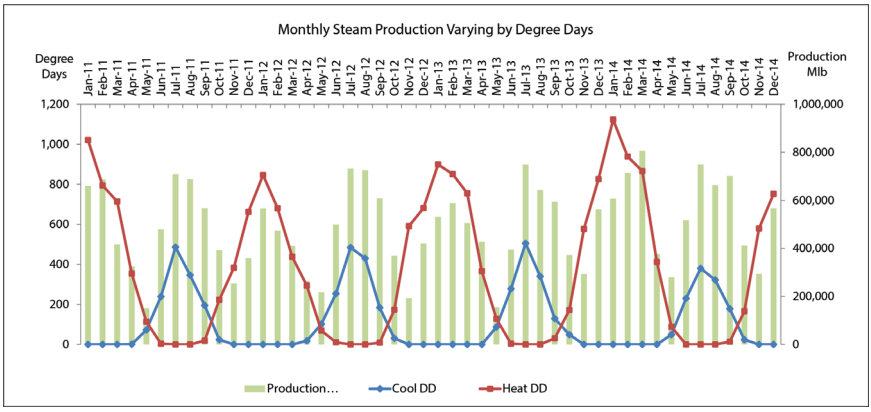
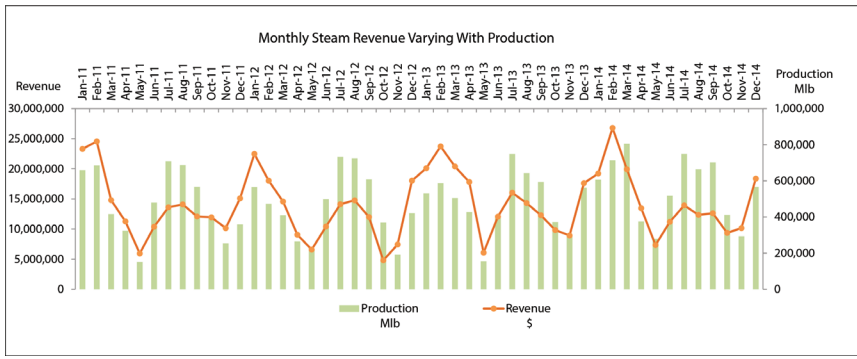


Exhibit B-6

Time Series Comparing Steam Revenue With Production

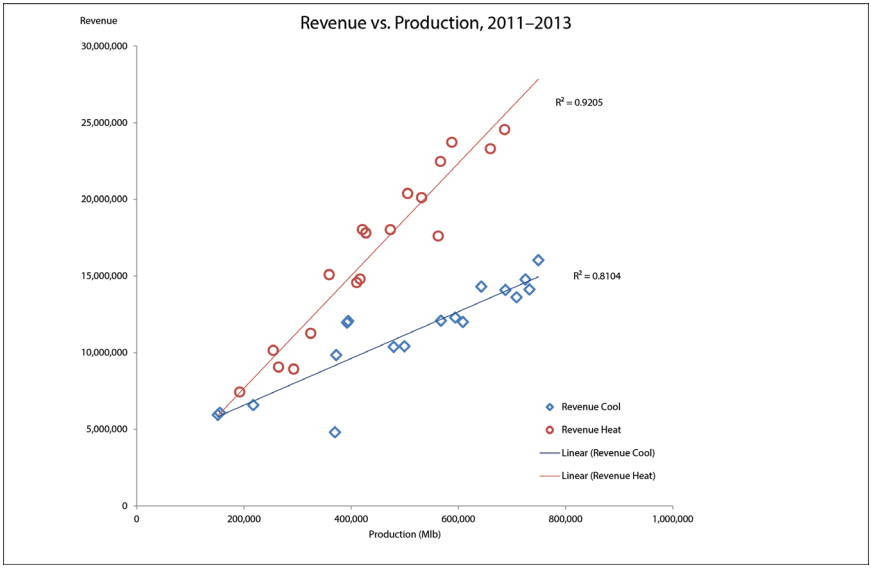


B.27 To better understand this seasonal behavior, the auditor created two time series charts to compare production with degree days and to compare revenue with production as depicted in exhibits B-5 and B-6, respectively. These exploratory charts accorded visually with the auditor's expectations based on an understanding of the entity's business. Production ramps up in winter and summer and winds down in spring and fall. Also, revenues in winter are lower than in summer even though production is roughly comparable.

B.28 With a clarified understanding of the business and data, the auditor began to build a predictive model for projecting revenue.

Exhibit B-7

Separate Regressions for Cooling and Heating, Base Data (2011–2013) Only

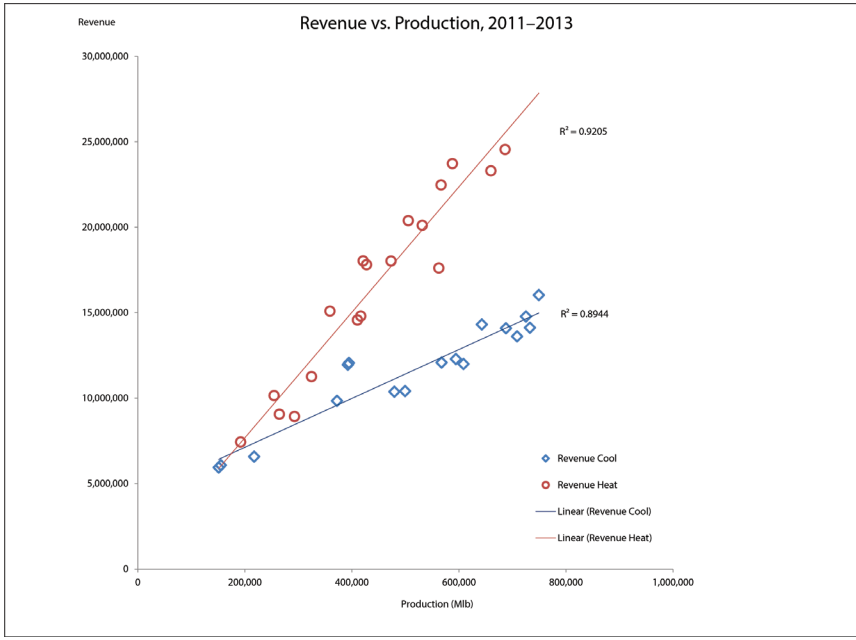


Model-Building

B.29 The auditor dealt with differences in the cooling and heating seasons by treating the observations as two separate data series. To use the analysis as audit evidence (and to avoid circular reasoning), the auditor built the model only from the data from the 2011–2013 base years (that is, not the data for the year currently under audit). Exhibit B-7 shows that the separate regressions provide a much better fit to their respective series than the single regression in exhibit B-4. However, a quick review of the chart reveals an aberration—one cooling month data point is located significantly below the others. Further investigation reveals that this data point is for October 2012, the month when Superstorm Sandy hit the east coast. U.S. steam production was seriously disrupted in that month, and results are highly atypical.

Exhibit B-8

Regressions After Eliminating the Superstorm Sandy Effect (October 2012)



Improving the Model

B.30 The auditor eliminated the anomalous verifiable October 2012 data point to improve the model. The result is shown in exhibit B-8, where it is apparent that the fit has improved for the cooling series, compared with exhibit B-7.

B.31 Additional refinements to the two series might be to add variables that reflect officially scheduled prices. For example, marginal pricing could be used to create a synthetic predicting variable consisting of production times marginal price. Also, other predicting variables could be introduced to account for pricing components that are semi-fixed and do not depend on consumption. A refinement that might be desirable in this case would be to create a dummy variable to distinguish between heating months and cooling months. Such a variable would have the value 1 in heating months and 0 in cooling months (or vice versa). The effect can be visualized as creating a third axis in exhibit B-8. The heating points would be pushed a distance of 1 into the 3-D diagram, and a flat plane would be fitted that runs through the center of both the heating points and the cooling points. In that way, one regression model with two predicting variables would be created, rather than two functions with one predicting variable each.

B.32 Multivariate models in which the test variable is regressed simultaneously against several predicting variables can be an effective way to account for the many factors that affect the test variable. However, this makes the model much more complex. The goal is to include as many variables as are needed, but

no more than that, to create a predictive model that credibly explains the behavior of the test variable. A powerful model that is also sparse (that is, having a low number of variables) is the ideal. Deciding which variables and how many is more art than science. Some regression packages (in software) can select such a set, at least according to quantitative criteria.

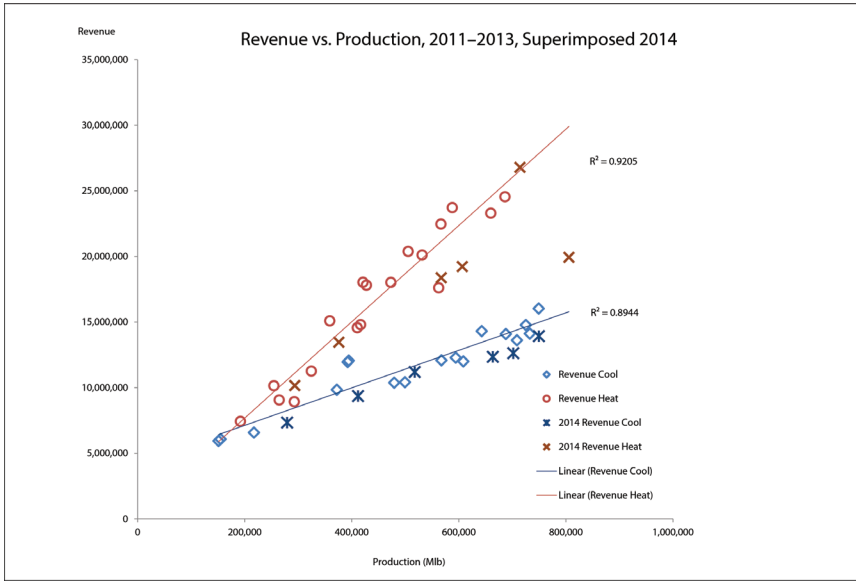
B.33 For simplicity, this example assumes that the auditor deems the model depicted in exhibit B-8 to be adequate for a substantive analytical procedure. The regression equations based on the 2011–2013 data and depicted as lines in the exhibit are as follows:

- Cooling: Projected Revenue = 14.302 × Mlb Produced + 4,266,060
- Heating: Projected Revenue = 36.699 × Mlb Produced + 345,199

B.34 The auditor used a regression package for these and other calculations. However, the slopes and intercepts of the two regression functions can be computed using functions in widely available software. We can see that the slope of the heating season regression function is greater than that of the cooling season function: 36.699 versus 14.302. This reflects the fact that an additional Mlb in heating season is more expensive than an additional Mlb in cooling season. On the other hand, the intercept term of the cooling season function exceeds that of the heating season function: 4,266,060 versus 345,199. This suggests that there is a larger fixed component to revenues in cooling versus heating months. These coefficients and constants were directionally as expected based on, for example, the results of procedures performed by the auditors to obtain an understanding of the entity's business.

Exhibit B-9

Regression Model From 2011–2013 Base Data With 2014 Data Superimposed



Confirmatory Analysis of Audit Revenues

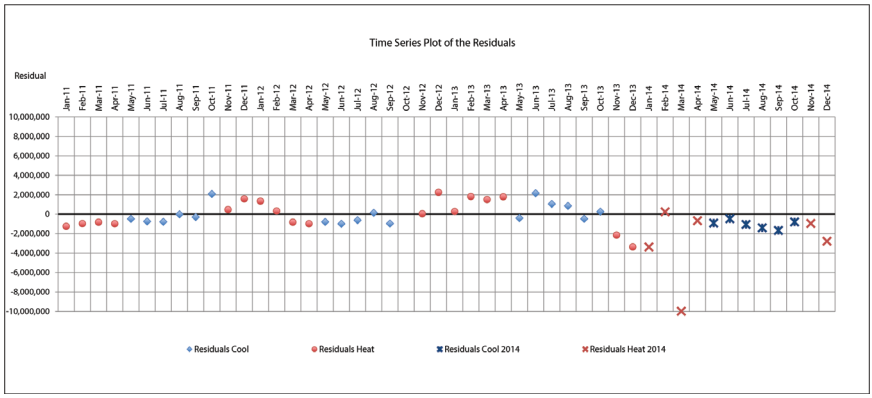
B.35 Having created the model from the base data, the auditor plugged in the 2014 data as shown in exhibit B-9. For each observation, the difference between recorded and projected revenue is the residual unexplained behavior calculated as follows:

Residual = Recorded Revenue — Projected Revenue.

B.36 Each residual can be represented graphically as the distance (positive or negative) between the data point and the regression line measured on the vertical line rising perpendicularly from the value of the predicting variable on the horizontal axis and running parallel to the vertical axis.

Exhibit B-10

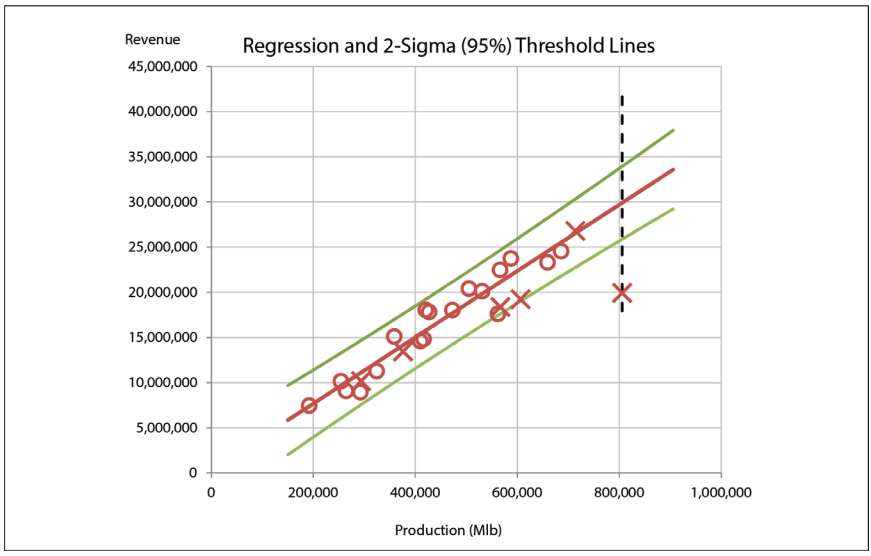
Time Series Plot of Residuals



B.37 Residuals represent the variability in the test variable that is not explained by the regression model. Therefore, it is useful to review a timeline plot of the residuals to see if any patterns emerge that might indicate specific missing model elements and also to detect such things as large residuals at quarter or year-end that reverse in the following period. The residuals for the base data used to develop the regression model always sum to zero (because the regression function passes through the mean). However, the residuals for the projection data usually do not. The monthly residuals for US SteamCo are shown in exhibit B-10.

Exhibit B-11

Testing for Significant Fluctuations From the Regression Line



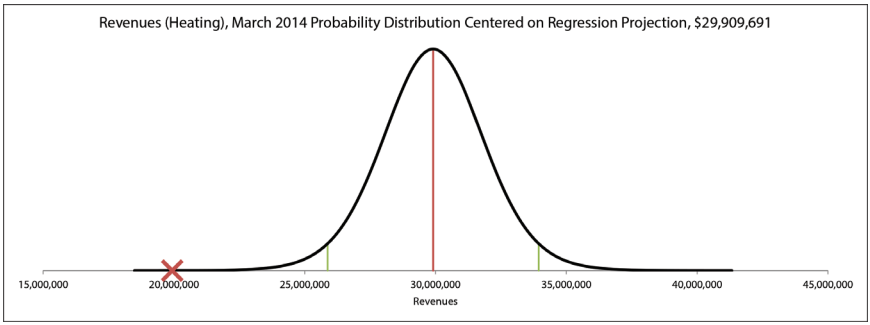
Analysis of Residuals

B.38 The auditor performed an analysis of residuals. Clearly, heating revenue for March 2014 was significantly less than projected by the predictive model. The auditor also determined whether there were other significant residuals by using the statistical characteristics of the model developed from the base data, in particular, the historical degree of fluctuation from the model. It is reasonable to consider a fluctuation as essentially random for any given monthly production statistic. Also, its probability follows an approximate bell curve centered on the regression projection. Therefore, two-sigma (two-standard deviation) thresholds can be established around regression estimates, within which actual revenue is expected to fall approximately 95 percent of the time. Exhibit B-11 shows the regression function for heat revenues, together with upper and lower two-sigma thresholds. Note that the threshold lines in exhibit B-11 bend away from the regression line (two-sigma becomes larger) as the predicting variable moves away from the center of the base observations. In this case, average production in the base period is 441,000 Mlb. At that midpoint, the two-sigma band is at its narrowest as measured on a vertical line perpendicular to the horizontal and parallel to the vertical axis. For values further from the center, projections become more variable and the two-sigma band becomes wider.

B.39 In round numbers, about 800,000 Mlb of steam was produced in March 2014 and the regression model projected revenues of just less than $30 million. Recorded sales were about $20 million, approximately $10 million less than projected. Recorded revenues would be expected to fall somewhere on the dotted vertical line illustrated in exhibit B-11, and 95 percent of the time they are expected to fall within the two-sigma thresholds.

Exhibit B-12

Probability Distribution of Potential Revenues for March 2014



B.40 Exhibit B-12 shows the probability distribution of sales at the 800,000 Mlb point. The distribution is centered on the regression projection, which is slightly less than $30,000,000. Two-sigma at this point on the regression line is about $4 million. Accordingly, the lower and upper thresholds are at about $26 million and $34 million, respectively. Probability is represented by the relative area under the curve. Thus, by definition of two-sigma, the area between the thresholds is 95 percent. The probability that random effects will cause revenues to fall below the lower threshold is $2^1/_2$ percent. At that level, the fluctuation is statistically significant enough to warrant further investigation. The probability that actual revenue would randomly be as low as $20 million (X marks the spot) is vanishingly small, and it is almost certain that something specific caused the fluctuation.

B.41 The probability distribution in exhibit B-11 can be imagined as sitting vertically on exhibit B-12 with its peak facing skywards, its horizontal axis on the dotted line and centered on the regression line. The distribution is based on a *t*-distribution with degrees of freedom equal to the number of base observations for heating months after eliminating October 2012 (Sandy) minus two, that is, 16 degrees of freedom.

B.42 The auditor also looked for patterns in the data that might suggest the potential for misstatement, for example, runs of data points above or below the regression line. From exhibit B-12 it appeared that 2014 revenue was consistently less than projected. The auditor investigated this matter and the individually significant fluctuations identified. Procedures performed included, for example, making inquiries of management and obtaining corroborating audit evidence through examination of documentation relating to production and invoicing.

Evaluate and Respond to the Results of the Procedure

B.43 The auditor's expectation was an estimate of revenue from sales of steam for each month (the dependent or test variable). The recorded revenue from steam sales was significantly different from the auditor's expectation. The

results of this regression analysis caused the auditor to reassess aspects of the effective operation of some of the entity's controls relevant to revenue and redesign the nature and extent of tests of details to respond to the changes in the assessed risk of material misstatement.
