1978

Semilog Transformation

Makridakis, Spyros

John Wiley & Sons Inc

6/5/1 Semilog Transformation

In business and economic series it is often true that a constant rate of growth prevails. This can happen with the sales of a company, GNP, consumption patterns, etc. For example, if the growth in GNP is 5% a year, it implies a compounded yearly rate of growth of 5%, a pattern that is exponential. Table 6-4 shows the revenues of an antipollution company (Lanard) which follows a typical exponential pattern of growth. (These data are graphed in Figure 6-1.) Regression can be used to estimate a forecasting equation for this nonlinear pattern and to find the exact rate of growth.

As Figure 6-1 shows, the pattern of actual revenues is far from linear. However, an exponential pattern can be described by the following form:

$$\text{revenues} = e^{a+bX}, \qquad (6\text{-}6)$$

where X is time ($1946 = 1, 1947 = 2, \ldots, 1975 = 30$).

Equation (6-6) is equivalent to (6-7) except that the $\log_e$'s, natural or base e logarithms, have been taken off both sides:

$$\log_e(\text{revenues}) = a + bX[\log_e(e)], \qquad (6\text{-}7)$$

$$\text{or} \quad Y = a + bX, \qquad (6\text{-}8)$$

$$\log_e(\text{revenues}) = Y, \text{ and } \log_e(e) = 1.$$

Thus, applying a semilogarithm transformation to (6-6) gives the linear form required for a regression equation as described by (6-8).

Transforming the original sales figures to the corresponding natural logarithm ($\log_e$), a simple regression between the transformed revenue figures (column 4 of Table 6-4) and time, X, can be used to estimate the values of a and b. This is a typical simple regression model, and the resulting parameter values can be found to be

$$Y = 4.54 + .083X.$$

The computed F-test is 964 and $R^2 = .972$. Both t-tests are significant as shown in Table 6-5. To estimate the sales for 1975, this regression equation can be used as follows:

$$Y = 4.54 + .083(30) = 7.03.$$

However, since $\quad Y = \log_e(\text{revenues})$,

$$\text{revenues} = \text{antilog}_e(Y),$$

TABLE 6-4 REVENUES OF LANARD COMPANY

(1) Year	(2) Time Period X	(3) Sales (in $1000s)	(4) Natural Logarithm of Sales Y
1946	1	115.182	4.74652
1947	2	67.6176	4.21387
1948	3	104.482	4.64901
1949	4	126.062	4.83678
1950	5	154.174	5.03808
1951	6	174.861	5.16399
1952	7	193.988	5.26779
1953	8	186.968	5.23094
1954	9	223.893	5.41117
1955	10	251.291	5.52661
1956	11	261.78	5.5675
1957	12	232.868	5.45047
1958	13	266.132	5.58399
1959	14	308.049	5.73026
1960	15	283.709	5.64795
1961	16	324.676	5.78283
1962	17	422.233	6.04556
1963	18	387.273	5.95913
1964	19	448.078	6.10497
1965	20	517.24	6.24851
1966	21	558.97	6.3261
1967	22	588.052	6.37682
1968	23	581.686	6.36593
1969	24	685.469	6.5301
1970	25	744.772	6.61308
1971	26	792.753	6.67551
1972	27	826.263	6.71691
1973	28	900.894	6.80339
1974	29	1026.42	6.93383
1975	30	1093.86	6.99746

or $$\text{revenues} = \text{antilog}_e(7.03) = 1130.$$

Similarly, the revenues for 1976 can be estimated as

$$Y = 4.54 + .083(33) = 7.279,$$

and $$\text{revenues} = \text{antilog}_e(7.279) = 1480.$$

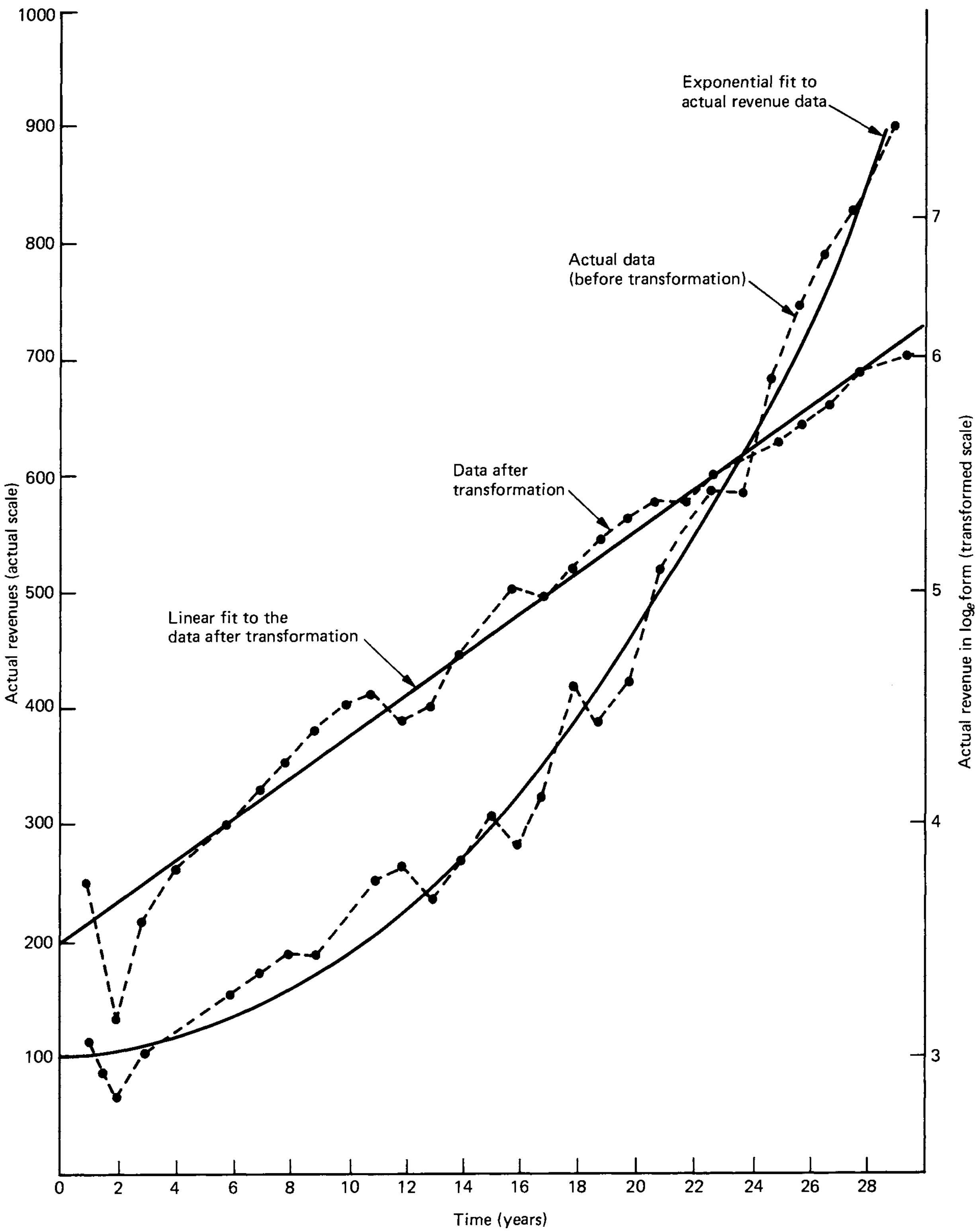

FIGURE 6-1 REVENUES OF LANARD COMPANY—LINEAR AND SEMILOG SCALES

TABLE 6-5 DEPENDENT VARIABLE, $\log_e$ (REVENUES), OF LANARD—SEMILOG FIT

Variable	Regression Coefficient	Standard Error	*t*-Test
Constant (*a*)	4.53804	4.72203E-02	96.1035
Time (*b*)	8.25889E-02	2.65986E-03	31.0501
$R^2 = 0.972$		$R = 0.986$	*F*-Test = 964.1

$\log_e$ (sales) Actual	Predicted	Residuals	Percentage Error
4.74652	4.62063	.125889	2.65324E−02
4.21387	4.70322	−.489348	−.116128
4.64901	4.78581	−.136793	−2.94242E−02
4.83678	4.86839	−3.16192E−02	−6.53725E−03
5.03808	4.95098	8.70982E−02	.017288
5.16399	5.03357	.130421	2.52558E−02
5.26779	5.11616	.151633	2.87849E−02
5.23094	5.19875	3.21882E−02	6.15342E−03
5.41117	5.28134	.129829	2.39929E−02
5.52661	5.36393	.162682	2.94361E−02
5.5675	5.44652	.120987	.021731
5.45047	5.52911	−7.86356E−02	−1.44273E−02
5.58399	5.6117	−2.77035E−02	−4.96124E−03
5.73026	5.69428	3.59745E−02	6.27799E−03
5.64795	5.77687	−.128924	−2.28267E−02
5.78283	5.85946	−7.66358E−02	−1.32523E−02
6.04556	5.94205	.103505	1.71208E−02
5.95913	6.02464	−6.55107E−02	−1.09933E−02
6.10497	6.10723	−2.26188E−03	−3.70498E−04
6.24851	6.18982	5.86874E−02	9.39224E−03
6.3261	6.27241	5.36885E−02	8.48684E−03
6.37682	6.355	2.18198E−02	3.42174E−03
6.36593	6.43758	−7.16534E−02	−1.12558E−02
6.5301	6.52017	9.92990E−03	1.52063E−03
6.61308	6.60276	1.03154E−02	1.55985E−03
6.67551	6.68535	−9.84049E−03	−1.47412E−03
6.71691	6.76794	−5.10283E−02	−7.59608E−03
6.80339	6.85053	−.047142	−6.92920E−03
6.93383	6.93312	7.10964E−04	1.02536E−04
6.99746	7.01571	−1.82433E−02	−2.60713E−03

TABLE 6-6 SIMPLE LINEAR REGRESSION OF REVENUES AND TIME FOR LANARD

Variable	Coefficients	Standard Error	t-Test
Constant	−60.3383	32.5956	−1.85112
Actual sales	31.5265	1.83607	17.1707

$R^2 = 0.913$ $R = 0.956$ F-Test = 294.8

Actual	Predicted	Residuals	Percentage Error
115.182	−28.8117	143.994	1.25014
67.6176	2.7148	64.9028	.959851
104.482	34.2413	70.2404	.672275
126.062	65.7679	60.2944	.47829
154.174	97.2944	56.8797	.368932
174.861	128.821	46.0405	.263297
193.988	160.347	33.6401	.173414
186.968	191.874	−4.90576	−2.62385E−02
223.893	223.401	.492798	2.20104E−03
251.291	254.927	−3.63629	−1.44705E−02
261.78	286.454	−24.6734	−9.42523E−02
232.868	317.98	−85.1122	−.365496
266.132	349.507	−83.3746	−.313283
308.049	381.033	−72.9845	−.236925
283.709	412.56	−128.851	−.454166
324.676	444.086	−119.411	−.367785
422.233	475.613	−53.3802	−.126424
387.273	507.139	−119.867	−.309514
448.078	538.666	−90.5878	−.20217
517.24	570.192	−52.9526	−.102375
558.97	601.719	−42.749	−7.64782E−02
588.052	633.245	−45.1932	−7.68524E−02
581.686	664.772	−83.0857	−.142836
685.469	696.298	−10.8291	−1.57981E−02
744.772	727.825	16.9474	2.27551E−02
792.753	759.352	33.4012	4.21332E−02
826.263	790.878	35.3848	4.28251E−02
900.894	822.405	78.4894	8.71239E−02
1026.42	853.931	172.485	.168046
1093.86	885.458	208.398	.190517

The parameter b in $Y = a + bX$ is the slope of the line and indicates the number of units increase in Y that occurs with a one-unit increase in X. On the other hand, b in $Y = e^{a+bX}$ approximates the percentage growth in Y caused by a one-unit increase in X. Thus .083 indicates that the revenues of Lanard have been growing about 8.3% a year on the average (the actual growth is $8.3 + (8.3)^2/2 + (8.3)^3/6 = .0865$).

A natural question arising from the above examples is how to choose the most appropriate transformation to apply to the original data from the very large number of nonlinear functions available. One approach can be illustrated by assuming that a regression between the actual revenues of Lanard and time is run without any transformation. The result is shown in Table 6-6. What is interesting to note is the pattern in the residuals and the percentage error. About one-fourth of the residuals have a positive sign in the beginning, followed by about half with a negative sign, and finally the last quarter again have a positive sign. This shows a definite nonrandom pattern, which implies that the actual data are above the regression line in the beginning, then below, and finally above again. This pattern is a clear indication that the linear form does not fit the data well. Another indication is the percentage error, which is initially large, becomes smaller, changes sign, then rises, declines, and finally rises again. (Still another indication of the inappropriateness of a linear model to describe the data is given by the Durbin-Watson test, which will be discussed in the Chapter 6 Appendix, Section 3.)

From Figure 6-2, it can be seen that if the actual data is an exponential

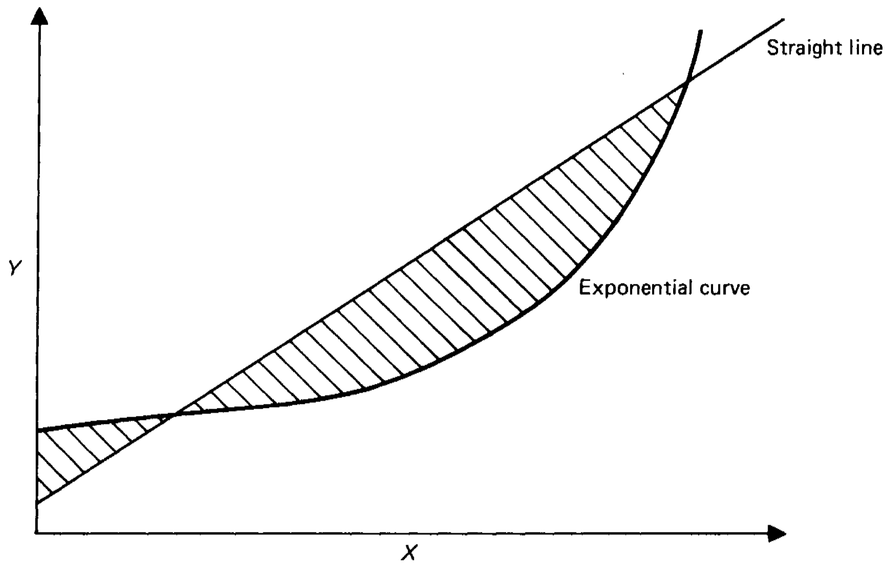

FIGURE 6-2 LINEAR APPROXIMATION OF EXPONENTIAL PATTERN

curve, describing it by using a straight line will give the type of error pattern and percentage errors observed in Table 6-6. In order to identify the best transformation, several possible functional forms can be plotted, previous knowledge can be applied, or the residuals can be examined. Some combination of these procedures is usually best for determining what pattern remains in the residuals once a linear form has been fitted to the data.

6/5/2 Polynomial Transformations

To further illustrate the use of nonlinear functions, the production run cost figures of the Carlisle Corporation (Table 6-7) can be examined to find the functional relationship between the total cost and the number of units produced. Economic theory would suggest that cost functions are either linear

TABLE 6-7 PRODUCTION COSTS FOR CARLISLE

Number of Units Produced (in 1000s) X_1	Total Cost of Production (in \$1000s) Y	Number of Units Produced (in 1000s) X_1	Total Cost of Production (in \$1000s) Y
7.3865	2094.15	7.89939	2295.63
11.1283	3015.36	11.3904	3201.87
9.16186	2407.32	10.5926	2789.65
6.21721	2045.15	11.9572	3488.61
13.2652	4157.27	3.44103	2000.95
9.87053	2751.86	9.45951	2648.15
8.21257	2312.08	9.88421	2732.38
7.38249	2250.33	15.4036	5841.19
7.93566	2264.81	4.68518	1968.65
1.92551	2054.94	3.9589	2001.01
2.6833	2078.63	7.19952	2125.95
11.4996	3271.48	15.8771	6317.62
8.09594	2199.07	8.96216	2393.23
5.565	1984.25	14.0068	4743.59
7.79677	2209.25	10.2914	2843.17
10.7661	3032.66	6.09008	2026.48
13.9806	4634.34	10.2778	2882.33
3.33987	2101.53	15.3664	5728.24

or cubic. Both linear regression with the actual data and with the cubic transformation can be used to determine which form best fits the actual cost figures. The linear cost function is straightforward—the dependent variable is cost, and the independent variable is number of units produced. The resulting regression is

$$\text{cost} = 420.09 + 277.96X, \tag{6-9}$$

where X is the number of units produced.

As shown in Table 6-8, the computed F-test is 105.7, the t-test is 1.6, the $t\text{-test}_b$ is 10.28, and $r^2 = .757$.

The next step is to compare the results with the cubic fit to decide which

TABLE 6-8 LINEAR FIT TO THE CARLISLE COST DATA OF TABLE 6-7

Regression number 1 : Dependent variable is total cost.

Variable	Coefficients	Standard Error	t-Test
Constant	420.087	261.512	1.60638
Linear term : X_1	277.964	27.0341	10.282

$R^2 = 0.757$	$R = 0.870$	F-Test = 105.7

Actual	Predicted	Residuals	Percentage Error
2094.15	2473.27	−379.112	−.181034
3015.36	3513.34	−497.977	−.165146
2407.32	2966.75	−559.429	−.232386
2045.15	2148.25	−103.099	−5.04114E−02
4157.27	4107.32	49.9497	.012015
2751.86	3163.74	−411.872	−.14967
2312.08	2702.88	−390.804	−.169027
2250.33	2472.15	−221.817	−9.85707E−02
2264.81	2625.91	−361.105	−.159442
2054.94	955.308	1099.63	.535115
2078.63	1165.95	912.683	.439079
3271.48	3616.57	−345.092	−.105485
2199.07	2670.46	−471.396	−.214362
1984.25	1966.96	17.2888	8.71304E−03
2209.25	2587.31	−378.055	−.171123
3032.66	3412.67	−380.006	−.125304

TABLE 6-8 Continued

Actual	Predicted	Residuals	Percentage Error
4634.34	4306.19	328.151	7.08087E - 02
2101.53	1348.45	753.077	.358347
2295.63	2615.83	-320.205	-.139485
3201.87	3586.19	-384.319	-.120029
2789.65	3364.45	-574.802	-.206048
3488.61	3743.76	-255.146	-7.31367E - 02
2000.95	1376.57	624.378	.312041
2648.15	3049.49	-401.339	-.151555
2732.38	3167.54	-435.161	-.159261
5841.19	4701.73	1139.47	.195074
1968.65	1722.4	246.25	.125086
2001.01	1520.52	480.492	.240125
2125.95	2421.29	-295.341	-.138922
6317.62	4833.34	1484.28	.234943
2393.23	2911.24	-518.016	-.216451
4743.59	4313.47	430.125	.090675
2843.17	3280.72	-437.549	-.153895
2026.48	2112.91	-86.427	-4.26488E - 02
2882.33	3276.93	-394.599	-.136903
5728.24	4691.38	1036.87	.18101

Durbin-Watson Statistic = 2.00646.

of the two is better. This comparison can be made using the F-test or R^2 values for each regression model. The cubic form is

$$\text{cost} = a + b_1X_1 + b_2X_1^2 + b_3X_1^3. \qquad (6\text{-}10)$$

Letting $X_2 = X_1^2$

and $X_3 = X_1^3$,

equation (6-10) becomes

$$\text{cost} = a + b_1X_1 + b_2X_2 + b_3X_3. \qquad (6\text{-}11)$$

Equation (6-11) is a linear function based on the transformation of X_1^2 and X_1^3 of (6-10). These values are given in Table 6-9 and by defining as X_2

TABLE 6-9 CARLISLE COST DATA TRANSFORMED FOR CUBIC FIT

Total Cost	X_1	$X_2 = X_1^2$	$X_3 = X_1^3$
2094.15	7.3865	54.5604	403.01
3015.36	11.1283	123.838	1378.1
2407.32	9.16186	83.9396	769.043
2045.15	6.21721	38.6536	240.318
4157.27	13.2652	175.965	2334.2
2751.86	9.87053	97.4274	961.66
2312.08	8.21257	67.4462	553.907
2250.33	7.38249	54.5011	402.354
2264.81	7.93566	62.9747	499.746
2054.94	1.92551	3.70757	7.13895
2078.63	2.6833	7.20009	19.32
3271.48	11.4996	132.242	1520.73
2199.07	8.09594	65.5442	530.642
1984.25	5.565	30.9693	172.344
2209.25	7.79677	60.7896	473.962
3032.66	10.7661	115.908	1247.88
4634.34	13.9806	195.457	2732.61
2101.53	3.33987	11.1548	37.2555
2295.63	7.89939	62.4004	492.926
3201.87	11.3904	129.74	1477.79
2789.65	10.5926	112.204	1188.53
2488.61	11.9572	142.975	1709.58
2000.95	3.44103	11.8407	40.7442
2648.15	9.45951	89.4824	846.46
2732.38	9.88421	97.6977	965.665
5841.19	15.4036	237.271	3654.82
1968.65	4.68518	21.9509	102.844
2001.01	3.9589	15.6729	62.0475
2125.95	7.19952	51.833	373.173
6317.62	15.8771	252.082	4002.33
2393.23	8.96216	80.3203	719.843
4743.59	14.0068	196.19	2748
2843.17	10.2914	105.912	1089.99
2026.48	6.09008	37.0891	225.875
2882.33	10.2778	105.632	1085.66
5728.24	15.3664	236.125	3628.38

TABLE 6-10 LINEAR FIT TO THE CARLISLE COST DATA OF TABLE 6-9 (LINEAR, QUADRATIC, AND CUBIC TERMS INCLUDED)

Variable	Coefficients	Standard Error	t-Test
Constant	2156.56	98.2032	21.9602
Linear: X_1	−32.7729	41.3519	−.792537
Quadratic: X_2	−7.37235	5.07644	−1.45227
Cubic: X_3	1.61959	.185994	8.70778

$R^2 = 0.998$ $R = 0.999$ F-Test = 42382.4

Actual	Predicted	Residuals	Percentage Error
2094.15	2164.96	−70.8038	−3.38102E−02
3015.36	3110.84	−95.479	−3.16642E−02
2407.32	2483.	−75.6816	−3.14381E−02
2045.15	2057.05	−11.9075	−5.82231E−03
4157.27	4205.	−47.7295	−.011481
2751.86	2672.3	79.5615	2.89119E−02
2312.08	2287.28	24.8021	1.07272E−02
2250.33	2164.46	85.8704	3.81589E−02
2264.81	2241.6	23.2079	1.02472E−02
2054.94	2077.69	−22.7493	−1.10705E−02
2078.63	2046.83	31.7989	.015298
3271.48	3267.71	3.76123	1.14970E−03
2199.07	2267.44	−68.3743	−3.10924E−02
1984.25	2024.99	−40.7455	−2.05345E−02
2209.25	2220.5	−11.2509	−5.09261E−03
3032.66	2970.26	62.3972	2.05751E−02
4634.34	4683.11	−48.7725	−1.05241E−02
2101.53	2025.21	76.3227	3.63177E−02
2295.63	2235.98	59.6519	.025985
3201.87	3220.19	−18.314	−5.71976E−03
2789.65	2907.14	−117.49	−4.21165E−02
3488.61	3479.45	9.16309	2.62657E−03
2000.95	2022.48	−21.537	−1.07634E−02
2648.15	2557.77	90.3799	3.41294E−02
2732.38	2676.35	56.0312	2.05064E−02
5841.19	5821.82	19.3721	3.31646E−03

TABLE 6-10 Continued

Actual	Predicted	Residuals	Percentage Error
1968.65	2007.75	−39.1026	−1.98627E−02
2001.01	2011.76	−10.7516	−5.37307E−03
2125.95	2142.87	−16.9177	−7.95772E−03
6317.62	6259.92	57.6982	9.13290E−03
2393.23	2436.55	−43.3218	−1.81018E−02
4743.59	4701.77	41.8223	8.81658E−03
2843.17	2803.79	39.3779	.01385
2026.48	2049.36	−22.8828	−1.12919E−02
2882.33	2799.3	83.0303	2.88066E−02
5728.24	5788.66	−60.415	−1.05469E−02

and X_3 they can be linearly related to the total dollar cost. The regression results for this transformed data are shown in Table 6-10. The regression equation there is

$$\text{cost} = 2156.56 - 32.77X_1 - 7.37X_2 + 1.62X_3.$$

However, the t-test of X_1 is $-.793$, which is not significant, suggesting that X_1 has little influence on costs in this model and can therefore be dropped. After dropping X_1, the regression model can be reestimated as shown in Table 6-11, to obtain

$$\text{cost} = 2081.77 - 11.34X_2 + 1.76X_3.$$

This regression has an R^2 of .998, an F-test of 28598.2, and all the t-tests are significant. The R^2 of the cubic fit explains 99.8% of the total variations in cost, while that of the linear fit explains only 91.3%. This indicates that the cubic fit is better for forecasting future costs.

To estimate the cost of 10,000 units using the regression in Table 6-11 gives

$$\begin{aligned}\text{cost} &= 2081.77 - 11.34(10^2) + 1.76(10^3)\\ &= 2081.77 - 11.34(100) + 1.76(1000)\\ &= \$2,707,770.\end{aligned}$$

The choice of the appropriate transformation was aided in this case by economic theory. Oftentimes, higher order polynomials are also estimated by linear regression analysis. Several other nonlinear forms that are important in economic and management applications are described in the remainder of this section.

TABLE 6-11 LINEAR FIT TO THE CARLISLE COST DATA OF TABLE 6-9 EXCLUDING X_1 (QUADRATIC AND CUBIC TERMS ONLY)

Variable	Coefficients	Standard Error	t-Test
Constant	2081.77	27.0072	77.082
Quadratic: X_2	−11.3394	.839461	−13.5079
Cubic: X_3	1.761	5.21765E-02	33.7508

$R^2 = 0.998$ $R = 0.999$ F-Test = 28598.2

Actual	Predicted	Residuals	Percentage Error
2094.15	2172.79	−78.6301	−3.75474E−02
3015.36	3104.35	−88.9888	−2.95118E−02
2407.32	2484.22	−76.9019	−.031945
2045.15	2066.66	−21.5096	−1.05174E−02
4157.27	4196.95	−39.6807	−9.54488E−03
2751.86	2670.48	81.3857	2.95748E−02
2312.08	2292.39	19.6844	8.51374E−03
2250.33	2172.3	78.0334	3.46764E−02
2264.81	2247.72	17.085	7.54367E−03
2054.94	2052.3	2.64033	1.28487E−03
2078.63	2034.14	44.4859	2.14016E−02
3271.48	3260.23	11.2495	3.43867E−03
2199.07	2272.99	−73.9253	−3.36166E−02
1984.25	2034.09	−49.8455	−2.51206E−02
2209.25	2227.1	−17.8441	−.008077
3032.66	2964.95	67.7144	2.23284E−02
4634.34	4677.52	−43.1787	−9.31712E−03
2101.53	2020.88	80.644	.038374
2295.63	2242.22	53.4039	2.32633E−02
3201.87	3212.97	−11.0957	−3.46538E−03
2789.65	2902.44	−112.793	−4.04328E−02
3488.61	3471.08	17.5303	.005025
2000.95	2019.25	−18.3031	−9.14720E−03
2648.15	2557.7	90.4456	3.41543E−02
2732.38	2674.47	57.9128	.021195
5841.19	5827.39	13.8008	2.36266E−03
1968.65	2013.96	−45.3167	−2.30192E−02
2001.01	2013.31	−12.2997	−6.14674E−03

TABLE 6-11 Continued

Actual	Predicted	Residuals	Percentage Error
2125.95	2151.17	−25.2162	−1.18611E−02
6317.62	6271.39	46.2275	7.31724E−03
2393.23	2438.62	−45.3979	−1.89693E−02
4743.59	4696.31	47.2891	9.96904E−03
2843.17	2800.24	42.9246	1.50974E−02
2026.48	2058.96	−32.4832	−1.60293E−02
2882.33	2795.81	86.5229	3.00184E−02
5728.24	5793.82	−65.5732	−1.14474E−02

6/5/3 Logarithmic Transformation

If the function to be estimated is of the form $Z = AB^X$ (see Figure 6-3), taking logs of both sides gives

$$\log Z = \log A + X \log B. \tag{6-12}$$

Letting $Y = \log Z$,

$a = \log A$,

$b = \log B$,

equation (6-12) becomes

$$Y = a + bX. \tag{6-13}$$

As in the Lanard example of Table 6-4, the log of the dependent variable can be regressed against the values of X to obtain the values of a and b. The adjustments then needed to apply the results can be illustrated by assuming that the following regression coefficients have been obtained:

$$a = 0.301, \qquad b = -1.4332.$$

Thus, $Y = 0.301 - 1.4332X$,

since $a = \log A$

and $b = \log B$.

Thus $A = \text{antilog}\,(a) = \text{antilog}\,(0.301) = 2.0$,

$B = \text{antilog}\,(b) = \text{antilog}\,(-1.4332) = .038$, and $Y = 2.0 + .038X.$

So if $X = 10,$

$$Y = 2 + .038(10) = 2.38.$$

And since $Y = \log Z,$

$$Z = \text{antilog}\,(Y) = \text{antilog}\,(2.38) = 239.9.$$

Logarithmic transformations are particularly useful because the slope of the transformed function (6-13) can be used to approximate the growth rate and estimate elasticities (the percentage change in Y caused by a percentage change in X). Both are extremely useful in making policy decisions.

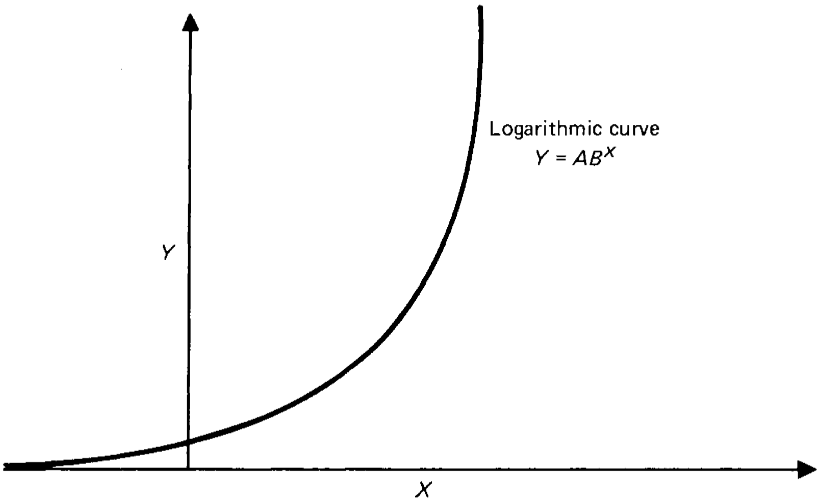

FIGURE 6-3 GRAPH OF LOGARITHMIC CURVE

6/5/4 Reciprocal Transformation

In order to estimate the total per unit cost, a reciprocal transformation (see Figure 6-4) can be used as shown in equation (6-14):

$$Y = a + \frac{b}{W}, \qquad (6\text{-}14)$$

where Y is the per unit total cost and W is the number of units products.

Letting $X = \frac{1}{W}$, equation (6-14) becomes

$$Y = a + bX.$$

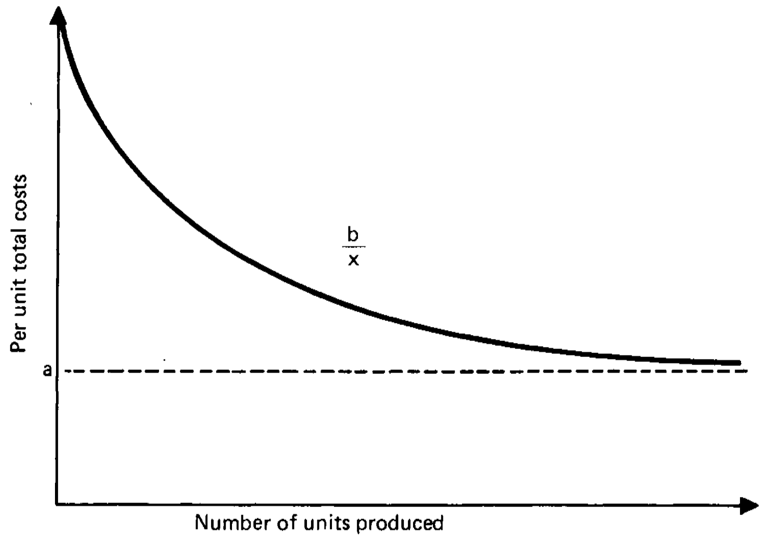

FIGURE 6-4 GRAPH OF A RECIPROCAL RELATIONSHIP

As an illustration of such a transformation, when

$a = 1.12$ and $b = 150,$

and the objective is to estimate the per unit total cost for 50 units, the following is obtained:

$$Y = 1.12 + 150\left(\frac{1}{50}\right) = 4.12.$$

In this instance, a is the per unit variable cost, while b is the fixed cost for a given level of production.

6/5/5 Double (Reciprocal and Logarithmic) Transformations (S-Curve)

The sales of some products follow an S-curve pattern. This pattern implies a slow start, a steep growth in sales, then a long period of saturation like that illustrated in Table 6-12 and Figure 6-5.

One functional form of an S-curve of this type is

$$Z = e^{a-(b/t)} \tag{6-15}$$

Taking the log of both sides, this becomes

$$\log_e Z = a - \frac{b}{t}$$

or $\log_e Z = a - bX,$

where $X = \frac{1}{t}.$

Finally $Y = a - bX,$ (6-16)

where $Y = \log_e Z.$

Equation (6-16) is of linear form so the values of a and b can be estimated using regression.

As an illustration of the S-curve transformation, Table 6-12 shows the data and the transformed values for Universal's TV sales. Table 6-13 shows the

TABLE 6-12 UNIVERSAL'S SALES OF COLOR TELEVISIONS

Time	Sales	1/Time	$\log_e$ (Sales)
1	.023	1	−3.77226
2	.157	.5	−1.85151
3	.329	.333333	−1.1117
4	.48	.25	−.733969
5	1.205	.2	.186479
6	1.748	.166667	.558472
7	1.996	.142857	.691145
8	2.509	.125	.919884
9	2.366	.111111	.861201
10	2.94	.1	1.07841
11	2.8714	9.09091E−02	1.0548
12	2.9346	8.33333E−02	1.07657
13	3.1346	7.69231E−02	1.1425
14	3.24	7.14286E−02	1.17557
15	3.148	6.66667E−02	1.14677
16	3.522	.0625	1.25903
17	3.54	5.88235E−02	1.26413
18	3.31	5.55556E−02	1.19695
19	3.547	5.26316E−02	1.2661
20	3.374	.05	1.2161
21	3.3745	.047619	1.21625
22	3.401	4.54545E−02	1.22407
23	3.6971	4.34783E−02	1.30755
24	3.493	4.16667E−02	1.25076

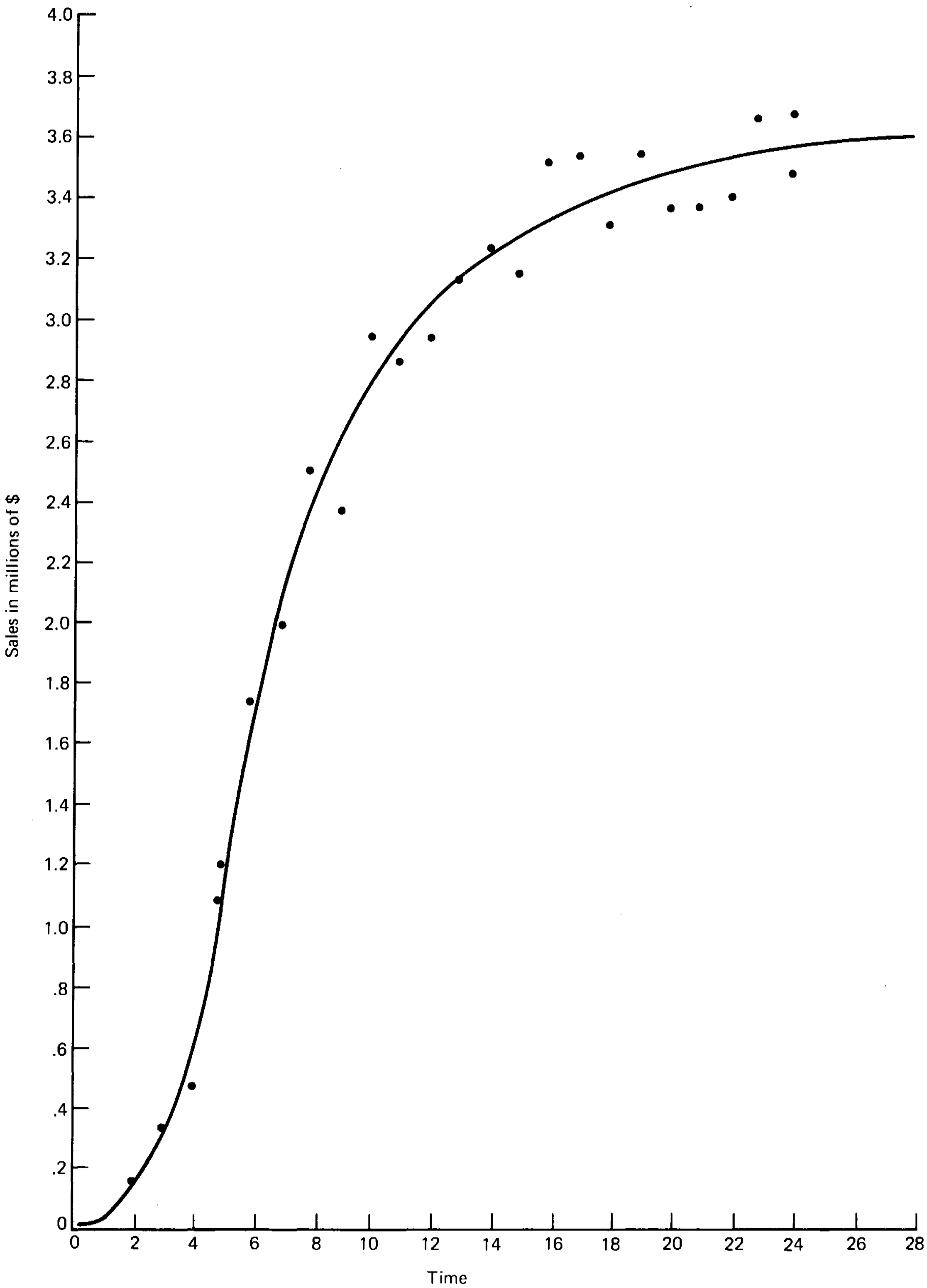

FIGURE 6-5 GRAPH OF UNIVERSAL'S COLOR TV SALES FROM TABLE 6-12—S-CURVE

regression results for the model (6-16), where Y is ($\log_e$ sales) and X is $1/t$. The R^2 for this regression is .976 and both the F- and t-tests are significant. Thus for forecasting purposes,

$$Y = e^{1.478-(5.786/t)} \qquad (6\text{-}17)$$

Equation (6-17) is a nonlinear equation that can be used for predicting the long-term behavior of products or technologies. In period 30, for example, Y(sales) will be

$$Y = e^{1.478-(5.786/30)}$$

$$= 3.613.$$

TABLE 6-13 REGRESSION MODEL FOR TRANSFORMED S-CURVE OF TABLE 6-12

Variable	Mean	Standard Deviation	Correlation Coefficients 3	4
3	.157332	.209559	1.000	−.976
4	.567637	1.24234	−.976	1.000

Variable	Coefficient	Standard Error	t-Test
Constant	1.478	7.11033E−02	20.7867
(3) 1/time	−5.78627	.275028	−21.0389

$R^2 = 0.953$ $\quad R = 0.976$ $\quad F$-Test = 442.6